

## 人工智慧治理雙元模型： 試論 ChatGPT 與 Conversational AI 在諮商上的可能應用及其倫理議題

王智弘\*

### 摘要

聊天機器人 ChatGPT 深受矚目，對校園與職場都帶來很大的衝擊，對學校教育與工作型態都可能會持續引發有待觀察的重要影響衝擊，其雖是科學與技術的重大進展，但是不可忽略其中所蘊含的文化與倫理議題。自從圖靈（Alan Mathison Turing, 1912-1954）在 1950 年提出模擬遊戲（Imitation Game）也就是圖靈測試（Turing's test）以回答機器是否能思考（Can machines think?）的問題，Weizenbaum 在 1966 發表了名為 ELIZA 的自然語言處理（Natural Language Processing, NLP）電腦程式，運用當事人中心學派的羅吉式（Rogerian）回應語法，以提供使用者類似與精神科醫師對話的經驗，可說是最早期的人工智慧（Artificial Intelligence, AI）運用於類似諮商情境的例子。1990 起以圖靈測試的標準為架構的 Loebner Prize 開始每年舉辦，AI 聊天機器人也在持續的發展。而其技術從運用模式匹配（Pattern matching）的演算概念，隨著電腦運算能力的逐漸提升，機器學習（Machine Learning）、深度學習（Deep Learning）、人工神經網絡（Artificial Neuro Network）、以及循環神經網絡（Recurrent Neuro Network, RNN）逐漸登場，以至 Google 公司運用注意力（Attention）機制提出了 Transformer 的程式概念，而 Open AI 公司據以發展出生成預先訓練轉換器（Generative Pre-trained Transformer, GPT），透過不斷加大數據庫參數量與運算能力，以及運用人工反饋的強化學習（Reinforcement Learning from Human Feedback, RLHF）之後，在 2022 年 11 月推出了眾所矚目的 ChatGPT，對於 ChatGPT 與 Conversational AI 在諮商（或者治療性溝通）上的可能應用值得探討，特別是其中可能引發的文化以及倫理的議題更是不能加以忽略，其中可能涉及文化與價值觀的偏見、資料的正確性與可用性、權力的集中與不當使用、通用人工智慧（artificial general intelligence, AGI）的隱憂，以及提供專業服務的資格能力、機械化或去人性化、專業服務的知情同意、專業服務的關係建立、診斷與技術使用、保密與預警（通報）、接受服務的公平性等議題都值得關切。最後，本文提出人工智慧治理的雙元模型以作為 AI 發展引發文化與倫理問題的對策。

**關鍵詞：**文化、倫理、ChatGPT、諮商、人工智慧治理雙元模型

王 智 弘\* 國立彰化師範大學輔導與諮商學系（ethicgm@gmail.com）

## 壹、人工智慧聊天機器人的發展

人工智慧 (Artificial Intelligence, AI) 聊天機器人 ChatGPT (Chat 是聊天, GPT 指的是 Generative Pre-trained Transformer, 生成式預先訓練轉換器) 深受矚目, 對校園與職場都帶來很大的衝擊, 對學校教育與工作型態都可能會持續引發有待觀察的重要影響衝擊, 其雖是科學與技術的重大進展, 但是不可忽略其中所蘊含與衍生的文化與倫理議題。自從圖靈 (Alan Mathison Turing, 1912-1954) (“Alan Turing,” 2023) 在 1950 年提出模擬遊戲 (Imitation Game) 也就是圖靈測試 (Turing’s test) (“Turing test,” 2023) 以回答機器是否能思考 (Can machines think?) (Turing, 1950) 的問題, 麻省理工學院 (Massachusetts Institute of Technology, MIT) 教授 Weizenbaum 在 1966 年發表了名為 ELIZA 的自然語言處理 (Natural Language Processing, NLP) 電腦程式, 運用當事人中心學派的羅吉斯式 (Rogerian) 回應語法, 透過模擬人類對話以提供使用者類似與精神科醫師對話的經驗, 可說是最早期的人工智慧運用於類似諮商情境的例子。ELIZA 也是被心理學界討論最多的人工智慧聊天程式 (王智弘, 2009; Fink, 1999; Grohol, 2004; Turkle, 1995; Whitty & Joinson, 2009)。其被歸於所謂的自動化 (automation) 的諮商服務 (Suler, 2002), 此等自動化諮商服務可能包括單機電腦軟體操作形式, 或者透過網路互動形式, 以及結合網路諮商專家系統或網路虛擬實境服務 (Virtual Reality) (王智弘, 2009)。ChatGPT 的出現及其可能的諮商應用, 其實早已在人們的構想與期待之中。

圖靈測試的提出受到廣泛的重視與回應, 並被視為 AI 發展的關鍵課題。以圖靈測試所定標準為架構之 Loebner Prize 比賽, 自 1990 起開始每年舉辦, 也吸引了各方 AI 好手投入研發, 歷年較為出名的得獎作品 (奪魁三次以上者), 比如: PC Therapist、A.L.I.C.E. (Artificial Linguistic Internet Computer Entity)、Jabberwacky, 以及 Mitsuku (後改稱為 Kuki, 曾奪魁五次, 是得獎最多次的作品) 等 (“Loebner Prize,” 2023), 這些作品反應了當時 AI 研發者的技術與構想, 也從而刺激了 AI 聊天機器人的不斷發展。

早期聊天程式的撰寫技術是運用模式匹配 (pattern matching) 的演算概念 (Wayback Machine, 2021), 此等演算法 (algorithm) (“Algorithm,” 2023) 的程式撰寫重點在於猜測與掌握使用者的溝通意圖, 透過演算規則以進行文字的取用與比對, 並藉以進行後續的對話與回應, 在自然語言處理領域的程式設計上, 相對而言是較為容易撰寫與使用的程式設計方法, 很多號稱是智慧型交談系統的程式, 其實都是運用此等方式產生 (陳鍾誠, 2023)。ELIZA 即是此等演算法的代表性產物, 雖然僅是文字介面, 但其成果確實可讓程式達到能與使用者持續溝通的功能, 甚至讓使用者誤以為是在與真人溝通。

而隨著 GPU (Graphics Processing Unit, 圖形處理器) 架構超級電腦的問世以及電腦運算能力 (HashRate) 的逐漸提升, 需要處理大量數據與強大運算能力的機器學習 (Machine Learning)、深度學習 (Deep Learning)、人工神經網絡 (Artificial Neuro Network)

以及循環神經網絡 (Recurrent Neuro Network, RNN) 等概念與技術乃逐漸登場，讓 AI 與聊天程式的功能越來越強大 (黃亦筠、陳良榕, 2023; “Recurrent neural network,” 2023)。在聊天程式的發展中也一直有尋求商業化的嚐試，比如 SmarterChild 就是其中一例，其曾被應用於美國線上 (America Online, AOL) 的 Instant Messenger 以及微軟的 Windows Live Messenger (前身為 MSN Messenger) 網路社群服務之上 (“SmarterChild,” 2023)。雖然聊天程式的發展如潮起潮落、來來去去，但是人們對發展能聊天的機器人始終充滿好奇與期待。

只是，雖然聊天程式一直有電腦玩家與資訊公司在持續加以研發，但是處理文字資料的技術則一直卡在輸入文字之處理數量受限的瓶頸，直到 Google 公司運用了注意力 (Attention) 機制的概念，在 2017 年的第三十一屆神經資訊處理系統會議 (31st Conference on Neural Information Processing Systems) 上所發表的 Attention Is All You Need 論文，正式提出了 Transformer 的程式概念 (Vaswani et al., 2017)，終於有了重大的突破。Transformer 是採用注意力機制的深度學習模型，透過將輸入的自然語言資料中各部分資料重要性的不同而賦予不同的權重分配，而能不受文字數量的限制，一次性處理所有的輸入資料，注意力機制可以為輸入序列中的任意位置提供上下文脈絡，而不必像 RNN 一樣一次只處理一個單詞，因此能允許更多的平行計算，提升了程式處理自然語言的效能 (“Transformer (machine learning model),” 2023)。這是自然語言演算法發展的關鍵性突破。

此一概念很快就被 OpenAI 發揚光大，這是一家自許要確保通用人工智慧 (artificial general intelligence, AGI) 能造福全人類的公司 (OpenAI, 2023a)，是由包含特斯拉 (Tesla) 執行長馬斯克 (Elon Musk，後宣稱擔心與 Tesla 發展自動駕駛 AI 的利益衝突而退出團隊)、前 Y Combinator 執行長阿特曼 (Sam Altman) 等，於 2015 年所共同創立的人工智慧研究機構—OpenAI 實驗室 (其由非營利組織 OpenAI Incorporated，及其營利組織子公司 OpenAI Limited Partnership 所組成)，Open AI 運用 Transformer 的概念，據以發展出 ChatGPT，並在微軟公司 (Microsoft) 持續注入資金與提供 Azure 線上超級運算平台的助益下，透過不斷加大數據庫參數量與運算能力以持續成長，比如 Open AI 公司在 2018 年 6 月推出 GPT-1，參數量是 1.2 億筆，2019 年 11 月推出 GPT-2，參數量是 15 億筆，在 2020 年 6 月推出 GPT-3，參數量是 1750 億，為改善 GPT 的回應品質，Open AI 公司開始運用人工反饋的強化學習 (Reinforcement Learning from Human Feedback, RLHF)，並於 2022 年 3 月推出 GPT-3.5，由於其聊天效果已相對成熟，因此，經持續改良後於才正式公開推出 ChatGPT (“ChatGPT,” 2023)。2023 年 2 月 7 日，微軟正式宣布將 ChatGPT 整合進 New Bing 之中 (“Microsoft Bing,” 2023)，並藉此更新資料以補充 ChatGPT 資料庫只收納到 2021 年 6 月的不足。Open AI 雖以 Open 為名，但是基於龐大的研發與運作費用，原為開源軟體 (open source software, OSS) (“Open-source software,” 2023) 的立

場，在 Microsoft 百分之百投資 OpenAI Limited Partnership 之後，也開啟了營利的商業運作模式。

ChatGPT 在 2022 年 11 月橫空出世、震驚世界，ChatGPT 的聊天效果令人驚艷，被認為應可通過圖靈測試，並開始對其應用有各式各樣的想像。經向 ChatGPT 詢問，其與 ELIZA 之差異何在，其回應略如下述（OpenAI, 2023b）：ChatGPT 和 ELIZA 都是基於自然語言處理技術的聊天程式，但是存在一些明顯的差異，ELIZA 的工作原理是透過將使用者輸入的文字與事先定義好的模式進行匹配，以生成回應，雖然 ELIZA 可以讓用戶感覺到被理解，但是程式本身缺乏真正的理解和學習能力，回應也缺乏靈活性和多樣性，而 ChatGPT 則是基於深度學習的程式，使用了「Transformer」的神經網絡架構，可以透過大量的數據訓練以學習自然語言，並生成流暢、自然的對話，因此更能理解和回應使用者的話語，在回應上也能更加靈活和多樣化，並且能夠持續學習和進化。上述 ChatGPT 的回答內容雖充滿自信但大體合宜，從 ELIZA 到 ChatGPT，AI 聊天程式的發展恍如世代交替，不可同日而語。有別於過去人臉辨識程式的單一功能分析式 AI，ChatGPT 屬於具廣泛功能之生成式 AI，在微軟的資金與網路運算平台的支持之下，其結合了演算法、海量資料與強大算力，而展現出強大的生成功能，使其廣受歡迎，創下了兩個月就達到上億使用人口的歷史紀錄（李光耀，2023）。然而，除了商業應用外，針對 ChatGPT 對人類社會所可能帶來的影響議題，也受到各界前所未有的矚目，包括，在各種領域上的可能相關應用，可能涉及的文化議題、社會影響議題、法律問題與倫理議題等（張良知，2023；詹婷怡，2023；Sallam, 2023; Zhuo et al., 2023），都持續受到學界與社會大眾的關切。

## 貳、AI 的終極目標與終極擔憂

AI 研發的終極目標也是人類對 AI 發展的終極擔憂，就是通用人工智慧 AGI 的發展，特別是「具身智能」（Embodied Intelligence）的發展，「史丹佛以人為本人工智慧研究院」（Stanford Institute for Human-Centered Artificial Intelligence, HAI）的共同院長李飛飛（Fei-Fei Li）表示：在 AI 越來越能具備對環境感知的能力之後，就可能發展出能與環境互動後產生自我學習、適應環境與能力演化的具身智能（楊孟軒，2023）。AI 機器人具備「具身智能」則能夠與人類一樣能夠在環境中自我學習，而不斷演化，AI 機器人是否會因此取代人類的主導地位，而不為人類所控制，甚至與人類為敵而攻擊人類，猶如科幻片所預言？此等議題聽來似乎較為超現實，但是卻是人類發展 AI 時存在於心中的隱憂。因此，生命未來研究所（Future of Life Institute）在 2023 年 3 月 23 日發表一封呼籲暫停大型 AI 實驗的公開信（Pause Giant AI Experiments: An Open Letter）。姑且先不就此等未來遠慮多加著墨，本文將就 AI 的應用與近憂層面，對 ChatGPT 與對話式人工

智慧 (Conversational AI, Co-AI) 在諮商上的可能應用，及其涉及的文化與倫理議題等即時課題進行探討。

### 參、ChatGPT 與 Co-AI 在諮商上的可能應用

由於 ChatGPT 的強大回應能力，引發各界的驚艷與關切，在各種應用 ChatGPT 的嚐試與想像之中，ChatGPT 與 Co-AI 在諮商（或者治療性溝通）上的可能應用，被認為是一個重要的研究議題與發展方向。目前觀之，其可能的應用包含下述：

1. 生涯諮商領域之應用：就如同生涯諮商也是最早投入網路諮商應用的理由一樣（王智弘等，2002；Boer, 2001），ChatGPT 與 Co-AI 因擁有強大的資料庫為後盾，此對於特別需要大量資訊提供的生涯諮商或職涯諮商領域而言，ChatGPT 與 Co-AI 無疑是一相當有用的生涯資訊來源與介面，確實可將其應用於職涯諮詢之上（曾荃鈺，2023）。

2. 專家諮詢系統之應用：由於 ChatGPT 與 Co-AI 在自然語言處理技術的重大改善，以及所具備強大資料庫與運算能力的基礎，能提供使用者表達流暢與資訊豐富的互動經驗，而可能作為類似專家諮詢系統之功能，確實已具備作為專家諮詢系統的良好條件。而結合電腦介面或手機 APP 在實務上亦可作為心理健康照顧系統之功能（Rucker, 2020）。

3. 線上評量系統之應用：ChatGPT 與 Co-AI 若能與心理評量工具適當的搭配使用，則可能應用於提供線上快速施測與解釋的心理評量服務。

4. 諮商輔助性資源之應用：在當事人尚未能獲得諮商師提供之諮商服務時，ChatGPT 與 Co-AI 的互動與溝通功能可能作為輔助性的晤談服務資源，在基於受到時間上或空間上的限制，當事人未能及時獲得諮商師協助時，提供當事人輔助性的溝通資源。

5. 諮商師教育與訓練之應用：ChatGPT 與 Co-AI 可作為模擬個案的練習之用，以提供受訓諮商師練習擔任諮商師（沈慶鴻，2019）或體驗當事人的經驗，其對話文本並可作為諮商師訓練教學與督導之用，進行個別或團體的指導或督導，以辨識或改進諮商晤談技巧。

6. 生成個案紀錄之應用：ChatGPT 與 Co-AI 可將諮商師與當事人諮商對話的錄音整理成個案紀錄。比如建立於 GPT-4 以及建立於其基礎之上的 Copilot 程式，即有將醫病對話整理成病歷的功能（李光耀，2023；Lee et al., 2023），因此，ChatGPT 與 Co-AI 可望應用於個案紀錄之整理工作。

惟就 ChatGPT 與 Co-AI 可能在諮商上的應用，其中所會引發的文化以及倫理的議題是不能加以忽略的，包括 ChatGPT 與 Co-AI 本身以及助人專業的特性，所可能引發的倫理議題都是要審慎加以思考的重點。就以「ChatGPT 可以應用在諮商上嗎？」詢問 ChatGPT，其回應略如下述（OpenAI, 2023c）：ChatGPT 模型雖然有強大的自然語言處理和生成能力，但要應用於諮商上，則有許多限制與挑戰，比如，其無法提供真人諮商師

所擁有的情感洞察力和人際交往能力，以便與當事人建立信任和共情的關係，其次，其亦無法擁有真人諮商師所需的專業知識和訓練，以便能夠正確地評估當事人問題，而提供有效的建議和治療，最後，其也無法保證當事人的隱私和保密性，因此，ChatGPT 不能完全取代真人諮商師的角色，但可在若干情況下作為輔助的工具，以提供基本的情緒支持和資訊，以及非常特定領域的知識與建議。ChatGPT 與 Co-AI 確實是有強大的資料庫與語言生成能力，有很大的潛力可應用於諮商服務上，惟其應用於諮商服務上會涉及當事人福祉與隱私等倫理議題，更有文化適切性的考慮，因此宜採取戒慎恐懼與謹慎的態度加以面對。

#### 肆、ChatGPT 與 Co-AI 應用所涉及的文化議題

首先，就 ChatGPT 與 Co-AI 所涉及的文化議題而言，生成式 AI 是需要透過餵食資料庫的海量資料，而提供給 ChatGPT 與 Co-AI 作為訓練的資料庫參數可能主要來自於美國或西方的主流文化地區，而可能會有文化心理學所關注的文化取樣不完整或偏誤的問題產生，即如 Arnett (2008) 指出，在 2003-2007 年間，心理學六大領域的主要期刊發表的文章，有 68% 研究所使用的研究受試者是來自於美國，而即使累積研究受試者之比例至 96%，也僅包含比例佔 12% 世界人口的西方工業化國家。亦即過去西方心理學理論的發展會有建立於一個怪異樣本 (WEIRD sample) 的問題：是屬於西方的 (Western)，教育普及的 (Educated)，工業化的 (Industrialized)，富有的 (Rich)，以及民主化的 (Democratic) 文化人口樣本 (Henrich et al., 2010a, 2010b; Hwang, 2012)。也就是當學術研究的取樣在文化代表性有偏差或不完整的情況下，做出的研究結論就會有偏差與不完整的可能。由於 ChatGPT 與 Co-AI 的內容回應是建立在其以英語為主資料庫參數的深度學習之上，其參數訓練文本的取樣若不具有完整的文化代表性，即可能有此類似之文化抽樣偏差的議題值得關切，而據於此等資料庫運行的 ChatGPT 與 Co-AI，其回應使用者的內容亦可能有文化不全或文化偏差的問題。而針對 ChatGPT 與 Co-AI 自然語言模型的實徵研究都顯示出具有文化與價值上的偏好現象 (Arora et al., 2023; Cao et al., 2023)，可見 ChatGPT 與 Co-AI 的文化議題確實值得關切。在此等相關文化議題中，值得注意的包括：

1. 文化思維差距的議題：西方歐美的主流文化傳統偏重個人主義 (individualism) 的思維導向，而東方與華人文化傳統則偏重關係主義 (relationalism) 的思維導向 (黃光國, 2009)，在中西文化互動之中，不免產生文化的衝突或文明的衝突 (黃光國, 2019)，當以西方主流文化資料庫內容所建置之 ChatGPT 與 Co-AI 成為東方與華人文化社群成員的知識導師或常識顧問時，其所提供之知識與常識內容，其實是有個人主義思維導向的文化內容，可能會造成文化衝突與應用上的限制。

2. 文化話語權的議題：當 ChatGPT 與 Co-AI 成為人類文化內容的發言人，其透過英語文化資料庫內容進行深度學習所得的文化觀點，會成為其表達文化意見的價值立場，非英語文化或者東方與華人文化的文化話語權會相對無力或者無置喙之地。這也是在美國推出 ChatGPT 的同時，世界各國都在發展各自的大型 GPT 語言模型的原因（黃亦筠，2023），因為，其中有文化話語權與 AI 可信賴性的考量。

3. 文化殖民的議題：當 ChatGPT 與 Co-AI 成為強勢的文化輸出者，ChatGPT 與 Co-AI 的價值思維與表達內容是傳播西方（或美國）主流文化的傳聲筒，而西方（或美國）主流文化可能對全世界，特別是對非英語、東方與華人文化地區產生文化殖民的現象，值得政府與民間對此等文化議題的關切。由於臺灣的學術界本即有對敏察西方學術殖民與建立自主學術傳統的文化省思與呼籲（王智弘、朱建民，2019；王智弘等，2019；中華本土社會科學會等，2019；李瑞全，2019；黃光國，2015），在亞洲，或華人文化與東方文化地區宜對生成式 AI 所涉及之文化議題有所自覺與關切，並有相對應之因應策略。

我們宜對 ChatGPT 與 Co-AI 的發展對人類文明之可能貢獻，抱持正向積極的態度，但是對其所涉及的文化議題仍宜有所關切。觀察目前 Co-AI 的發展，不論是基於文化的議題、公益的目的或是商業化競爭的考慮，在美國之外的其他國家與地區也在積極的發展其他語言模型，比如，法國有開源語言模型「Bloom」，韓國有「HyperClova」（黃亦筠，2023），就華人文化地區而言，亦有發展華人語言模式聊天機器人的相應行動，包括中國大陸百度公司所發展出的「文心一言」（ERNIE Bot）（文心一言，2023）與北京智源 AI 研究院發起的「悟道」等簡體中文語言模型，以及臺灣結合聯發創新基地、中央研究院辭庫小組、國家教育研究院、法國的開源語言模型 Bloom 以及負責營運國家網路與計算中心超級電腦「台灣杉二號」的台智雲等，共同建置號稱「台版 GPT」的繁體中文語言模型（黃亦筠，2023）。此一名為「福爾摩沙大模型」（Formosa Foundation Model）的語言模型是透過 1760 億參數所建立，約與 GPT-3.5 同級，相對於 ChatGPT 模型的中文資料比例較低（繁體中文不及 0.1%；簡體中文不及 2%），「福爾摩沙大模型」則納入了 30% 的繁體中文資料進行訓練（吳家豪，2023），確實是具有本土化的特色。

可見 ChatGPT 與 Co-AI 的發展是西方科學文明（特別是美國）實力的展現，也是其文化力量的展現，基於商業的考慮、科學的考慮或文化的考慮，各個國家或地區也發起了 AI 語言模型的軍備競賽，猶如我們長期關心的社會科學本土化或諮商心理學本土化的議題，針對在地使用者的需求，納入自身的文化內涵應是政府與民間應該關切的議題，也是內容服務科技產業與學界應該關切的重要議題。

## 伍、ChatGPT 與 Co-AI 應用所涉及的倫理議題

其次，就 ChatGPT 與 Co-AI 涉及的倫理議題而言，OpenAI 公司本身就針對 AI 生

成語言模型可能衍生的問題與因應已進行相當程度的討論 (Hagendorf, 2020)，由於 2022 年被視為 AI 產製內容 (AI-Generated Content, AIGC) 元年，各界開始對於 AI 可能涉及的倫理問題已開始加以關注與討論 (台灣人工智慧學校, 2023; Bickley, & Torgler, 2022; Goldstein et al., 2023; Prem, 2023)，其中的重要議題包括：著作權歸屬與侵犯的議題，資料治理 (Data Governance)、學術倫理與深偽 (Deepfake) 造假新聞的議題等，對於 AI 在心理學上的應用也有相關的倫理議題討論 (Bartlett et al., 2023)，包括：非代表性資料樣本所造成的偏見，AI 運算過程 (演算程序) 的可解釋性，以及臨床應用上的考慮等等。ChatGPT 問世後其可能衍生的問題亦被加以探討 (張良知, 2023; Zhuo, 2023)，其中可能涉及個人隱私保護、文化與價值觀的偏見、資料的正確性與可用性、影響力的不當操作等議題。雖然 OpenAI 公司本身有針對其隱私權政策的相關聲明 (OpenAI, 2023d)，也有透過問答集對使用者做相關的說明 (OpenAI, 2023e)。但是相關的倫理問題與疑問可能還是有持續探討的必要。

對於 AI 或 ChatGPT 應用所涉及的倫理議題，謹簡要探討如下 (Bartlett et al., 2023; Sallam, 2023; Zhuo, 2023)：

1. 著作權歸屬與侵犯的議題：ChatGPT 是屬於生成式 AI 語言模型，需要大量內容的資料庫參數提供訓練，正如同其他具商業運用目的之生成式 AI 模型，諸如生成程式碼的 Codex 模型、生成向量的 Embedding 模型、生成圖片的 DALL-E 模型等 (李光耀, 2023)，都是需要有內容參數訓練的需求，由於 ChatGPT 與 Co-AI 所產生的內容是取材自大量的資料庫參數內容，而此等內容為加工自各界所創作或產製的內容而得，ChatGPT 與 Co-AI 所產生內容之著作權將如何歸屬呢？是否具有原創性？是否會有剽竊或著作權侵犯的問題呢？由此而產生的著作權訴訟案件也將是一個棘手的議題，無論是 OpenAI、微軟或是其他 AI 語言模型的營運組織都必須正視此等著作權的相關議題。

2. 資料的正確性與可用性的議題：ChatGPT 與 Co-AI 所產製的資料既然來自各界，資料的正確性與可用性如何被確保則是重要的課題。避免虛假、錯誤或傷害性 (toxic) 的資料進入模型訓練的資料庫，確保模型輸出資料的準確性與公正性，也要確保資訊的更新 (ChatGPT 資料庫有透明度的限制與只更新到 2021 年 6 月) 與模型的更新以免輸出資料的過時，並保持模型的穩定性與可靠性 (Zhuo et al., 2023)。由於基於 ChatGPT 與 Co-AI 所產製的不正確資料，可能有引註與參考文獻的錯誤，以及錯誤資料傳播的風險 (Sallam, 2023)，可能使用者據以做出錯誤決策而帶來錯誤與傷害，特別在涉及生理與心理健康議題時更可能對使用者造成傷害或風險，也都是 AI 語言模型必須關切的議題。

3. 文化與價值觀之偏見與強加的議題：AI 演算法會因資料搜集方式、設備、標註人員的偏見而受到影響，也就是演算法可能學習到偏見 (張良知, 2023)，而語言模型的建立與訓練若主要來自單一語言的資料庫，可能產生的偏見包括：社會刻板印象、不公平的歧視、排除其他語言的觀點與文化內容、以及抑制其他語言發展的單一語言主義

(Zhuo et al., 2023)。後二者正是學界擔心臺灣推動獨尊英語的國家雙語政策可能引發的危機與隱憂。ChatGPT 與 Co-AI 所產製的資料可能有文化偏見的議題已在前文加以討論，在此要進一步討論價值觀的偏見與強加，無論是性別平等的價值觀，家庭價值的價值觀，宗教的價值觀，種族的價值觀，政治的價值觀，經濟的價值觀，生命的價值觀，以至於生理、心理、精神、道德、甚至是性的價值觀，ChatGPT 與 Co-AI 所產製的資料會不會涉及價值觀的偏見與強加？

4. 社會影響力的不當操作與深偽造假的議題：當 ChatGPT 與 Co-AI 以及其他的生成式 AI 形成極大的社會影響力，此等影響力若為負面內容（暴力、色情、偏見）或被不當操作，其對大眾的風險就開始提高，經歷英國少女因社群媒體演算法所提供的不安全內容而輕生的悲劇（劉亭妤，2022），網路媒體可能引發的負面社會影響力受到大眾更多的關切，對於生成式 AI 的社會影響力被不當操作使用與被運用於深偽造假的憂慮也隨之提高。針對 ChatGPT 與 Co-AI 以及其他的生成式 AI 的社會影響議題，宜以健康照顧的實務觀點加以檢視，以期能避免被不當使用與造成傷害（Sallam, 2023）。因為，被使用的愈廣，對社會大眾的影響力也越大。

5. 個人隱私權與組織安全的保護問題：ChatGPT 目前提供的免費版本會將使用者的詢問資料收入其語言模型練習資料庫，因此，使用者上傳的詢問資料是不會被保密的。媒體報導韓國三星公司（Samsung）於 2023 年 3 月 11 日起導入 ChatGPT 不到 20 天，即傳出機密資料外洩事件（今周刊，2023）。為了避免有意圖不良的使用者從語言模型中取得資訊而傷害到個人的隱私或組織的安全，語言模型可以運用篩選程式技術並透過事前告知以降低此等風險（Zhuo et al., 2023）。而對使用者而言，個人自身的警覺性與組織的使用規範可能也是必要的因應策略。

6. 學術倫理的議題：ChatGPT 一開始開放使用，馬上引起關注的議題就是學術倫理，小至學生寫作業、交報告與考試的原創性與公平性，入學申請資料與學習歷程檔案的真實性，以至於學位論文、研究報告以及各項文藝競賽與學術成就評定等。當 AI 生成學術文本時，令人擔心有偏見、錯誤、引用來源錯誤與研究欺詐（比如，代筆、偽造或假研究）的風險（Sallam, 2023），以及關於版權的議題、著作權歸屬的議題、剽竊的議題以及作者身份確認的倫理議題層出不窮（Liebrenz et al., 2023）。學術界對此確實有所憂心，全球著名期刊 Nature（2023）即發表其關切的立場與對策，教育部臺灣學術倫理教育資源中心（2023）在 2023 年 3 月 30 日也辦理了「生成式 AI 對研究與學術倫理的影響」的演講。而以防學術抄襲著名的 Turnitin 公司也發展其 Turnitin AI 的功能而推出 Turnitin Originality 以做為檢查學生作業真實性的工具，可用來檢測 AI 輔助寫作或 AI 寫作，以辨識是否有作弊的現象（Caren, 2022; Chechitelli, 2023）。看起來，此等議題也是需要運用 AI 工具來加以因應。

7. 資料治理的議題：AI 或 ChatGPT 與 Co-AI 的應用最重要也最涉及公權力的議題

是資料治理的議題，此一議題也與上述的議題息息相關。資料治理被視為在保護使用者資料、保護組織資料、提升使用者體驗以及協助組織決策上具有積極的功能，而此等資料治理工作除了資料管理者本身的自我管控之外，也是需要公權力加以介入的，包括建立資料保護的制度架構（組織架構、隱私聲明、保護政策），以及操作流程（隱私保護流程、意識教育培訓、指標監視報告、風險管理流程、跨境轉移機制、國際監管註冊）（葉耿志，2020）。在 AI 資料治理上，各國與國際規範的建立與修訂，以及合作平台的設置都是必要的措施（詹婷怡，2023），比如，歐盟提出的「資料治理法」（Data Governance Act）（European Parliament, 2022）概念；經濟合作暨發展組織（Organisation for Economic Cooperation and Development, OECD）所提出的人工智慧準則（OECD AI Principles, 2019），以及為落實此等準則而在 2020 年成立的「全球人工智慧合作夥伴聯盟」（The Global Partnership on Artificial Intelligence, GPAI）平台（GPAI, 2021），其中並成立資料治理工作小組（Working Group on Data Governance），其目前的工作重點，包括：通過資料信任實現資料共享以實現社會利益，推動資料正義的研究與實踐，以及隱私提升技術計畫等（GPAI, 2022a），臺灣大學劉靜怡教授則是此一工作小組的唯一臺灣成員（GPAI, 2022b）。要與世界同步推動 AI 資料治理工作，臺灣除宜積極參與國際合作團隊以交流經驗之外，並宜著手推動「人工智慧發展基本法」的制定，使臺灣 AI 倫理準則邁向法制化（張麗卿，2021）。AI 的治理機制，對政府、對民間、對學界、對產業界都是不可迴避的課題。

## 陸、ChatGPT 與 Co-AI 應用於諮商可能涉及的倫理議題

再其次，就 ChatGPT 與 Co-AI 在諮商上可能應用涉及的倫理議題而言，ChatGPT 與 Co-AI 若用於諮商服務是否會涉及媒體中介諮商（media mediated counseling）（王智弘，2004）所可能引發的倫理議題，諸如（王智弘，2004，2009）：

1. 法定資格能力的議題：ChatGPT 與 Co-AI 諮商上可能應用可能有兩種溝通型態：一為「治療中的溝通」（communication in therapy），一為「治療性溝通」（therapeutic communication），其中最大的差別在「意圖」（intent）的不同（Barnlund, 1990），前者意圖在治療，是在建立治療關係為前提下的治療行為，提供治療者需有專業資格與執照；後者意圖在溝通，是在一般自然交談中發生的正向建設性改變（王智弘，2009）。ChatGPT 與 Co-AI 若是作為第一種服務，則其執行治療任務的法定資格與能力需要被確認，ChatGPT 與 Co-AI 作為第二種服務，則 ChatGPT 與 Co-AI 雖然並不需經歷法定資格的認證程序，但是，仍需有相對應的安全與倫理考量，以避免不適當的回應造成對使用者的傷害，特別是對具有危機風險與嚴重精神疾病的當事人而言，此等造成傷害的風險會大增，ChatGPT 與 Co-AI 的對話服務需有對此等風險進行評估與因應的機制。

2. 機械化或去人性化的議題：諮商服務若透過科技設備之中介媒體進行傳輸，則可

能會有引發當事人對諮商服務有機械化或去人性化的感受 (Tait, 1999)，因此可能影響諮商服務的效能，如何透過媒體的呈現方式與內容上的改善，以避免引發當事人此等感受相當重要 (王智弘，2009)。ChatGPT 與 Co-AI 若被運用於作為諮商服務的輔助功能，如何在介面上提升使用者的人性化感受經驗自是發展的重點，其中可能被考慮進一步運用聲音、影像或虛擬實境的技術以提供擬真的諮商服務經驗，其相對引發的倫理議題將更為複雜。

3. 知後同意與風險管控的議題：ChatGPT 與 Co-AI 若是執行諮商，則其執行治療任務的法定資格與能力需要被確認，並且要進行相關的知後同意程序，需要針對 ChatGPT 與 Co-AI 執行諮商服務的特性及其限制做清楚的說明，特別要針對保密與保密的限制做完整的描述，以盡量降低對使用者或社會大眾的可能風險，若 ChatGPT 與 Co-AI 不提供治療為目的，即使聊天過程出現了治療性的溝通，ChatGPT 與 Co-AI 亦不需經歷法定資格的認證程序，但是，在聊天過程中若有使用者揭露隱私資料時，應有被加以保密的機制；相對而言，當保密的例外發生時，比如有自我傷害或傷害他人的情緒與意圖時，則宜有相關辨識與轉介使用者尋求諮商的機制，以盡量降低風險的產生，無論 ChatGPT 與 Co-AI 是否有意圖或定位要提供諮商服務，從過去 ELIZA 以降聊天機器人的使用經驗來看，使用者有情緒支持與諮商需求時尋求聊天機器人的互動與諮詢意見似不可免，因此，ChatGPT 與 Co-AI 的研發與營運單位宜在開發與營運過程中，除應在程式設計上力求穩健、可信，排除偏見與傷害性，以及辨識不倫理與高風險的輸入提示 (Zhuo et al., 2023)，並宜納入心理諮商與治療專家的意見，以因應上述專業倫理與風險管控的實際需要。

4. 關係建立的議題：此等議題包含兩方面，一是關係的確認，一是關係的建立。運用非面對面的傳統諮商服務，諮商關係如何確認是需要面對的問題，特別是運用 ChatGPT 與 Co-AI 以提供輔助性諮商服務，其角色責任與權利義務關係如何確認，是一個重要的議題。其次是關係的建立，若諮商服務只透過文字或聲音等媒介，由於資訊的不足可能造成情感投射、移情與反移情的可能性增加 (Childress & Asamen, 1998)，其相對會產生關係異化的議題，除了精神分析治療取向可能會用以作為分析的課題之外，其他治療取向則會審視治療關係變化的影響。

5. 判斷與評估失誤的議題：由於受限於 ChatGPT 與 Co-AI 的媒體溝通方式，不若實體面對面溝通的完整資訊呈現，而可能有訊息的簡化與失真的問題，如藉以提供諮商或諮詢的服務，可能造成判斷與評估上的誤差，因而造成解釋與回應的錯誤，若加上資料傳輸過程中有生成程式上的錯誤或資料的遺失，更會增加此等問題的風險，而可能無法滿足當事人的需要，甚至造成傷害 (王智弘，2009；Childress & Asamen, 1998; Ford, 1993; Morrissey, 1997; Shapiro & Schulman, 1996)，ChatGPT 與 Co-AI 的研發與營運者亦應將等風險納入考量。

6. 保密與預警(通報)的議題：如上述倫理議題對風險管控的考量，ChatGPT 與 Co-AI 的研發與營運者宜對使用者意圖尋求諮商或諮詢服務時，要特別注意隱私保密的問題，並透過程式設計與營運風險管控的規劃，盡量降低使用者因洩密而產生的個人與組織風險。此外，當面臨使用者可能自我傷害或傷害他人的意圖時，則為保密的例外，ChatGPT 與 Co-AI 的研發與營運者宜考量在程式設計與營運風險管控中加入預警與通報的機制(王智弘，2009)。從知情同意，判斷評估到保密議題的考量，都需納入風險管控的機制，以盡量降低對使用者或社會大眾的可能風險。

7. 使用者接受服務公平性的議題：由於民眾接近媒體與使用媒體的機會不一，如何力求所有民眾均能公平的得到 ChatGPT 與 Co-AI 的媒體溝通機會是需加考慮的議題，在程式設計與使用者設備條件上考慮近用性(accessibility)的議題，以增加當事人的公平使用機會(王智弘，2009)，政府與民間的資通訊部門在通訊基礎設施上的投資也是必要。

8. 文化與價值觀之偏見與強加的議題：ChatGPT 與 Co-AI 在諮商上之可能應用除會涉及上述諮商服務的倫理議題之外，其提供之服務內容亦會涉及本文前述之文化適切性議題(Arora et al., 2023; Cao et al., 2023)，以及文化與價值觀之偏見與強加的議題(Zhuo et al., 2023)，亦即諮商服務宜有文化適切性的考量，也就是本土化的考量，此等議題早為本土社會科學(黃光國，2009，2015，2018)與本土諮商心理學(王智弘，2009，2013，2014，2016，2017，2020，2023；王智弘等，2017；林幸台、王智弘，2018；陳秉華，2017)所關切，在 AI 時代更值得諮商專業人員加以注意。

## 柒、ChatGPT 與 Co-AI 倫理議題的因應：人工智慧治理的雙元模型

對於人類發展 AI 所可能涉及的倫理議題的回應與思考，可能還是要回到人們如何思考倫理議題的立場，根據 Kitchener (1984) 所提出的倫理辨明模式(model of ethical justification)，主張提供諮商服務的專業人員在面臨倫理問題而必須做出判斷的時候，會同時進行(1)直覺判斷(intuition)和(2)臨界評估(critical evaluation)兩個層面的思考，直覺判斷的思考是以個人的直覺進行反應，並就個人的倫理意識即時給出倫理問題之因應，但是，如此因應恐有失周全原，因此，需再進行臨界評估的思考，經由進一步透過參考相關專業倫理守則與法規以及倫理原則以進行倫理判斷，並提出專業人員應盡力維護接受諮商服務當事人的五大權益：(1)自主決定權；(2)諮商受益權；(3)免受傷害權；(4)公平待遇權；(5)要求忠誠權。而相對於這五大權益，專業助人人員則自我要求要負起(1)專業責任；(2)倫理責任；(3)法律責任等三大責任(牛格正、王智弘，2008)，以保障當事人的權益。也就是西方文化的倫理觀強調：外在客觀的理性規範，以保護服務對象的五大權益為出發點，然後要求助人者善盡三大責任，從遵守外部規範以保護當事人權益開始，並持續努力以履行責任並提高道德表現(王智弘，

2013, 2018, 2022)。此亦 Kohlberg (1964) 所提倡的道德發展過程, Van Hoose 和 Paradise (1979) 則進一步修正 Kohlberg 的觀點, 並提出了助人者的道德發展層次。

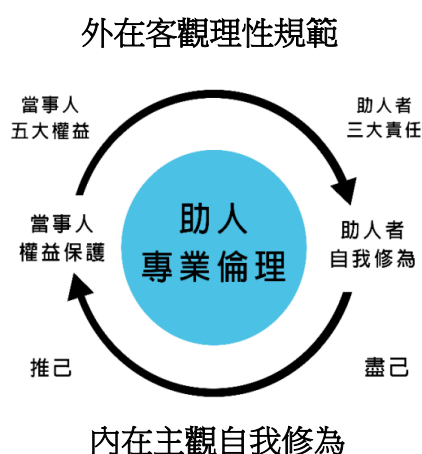
Van Hoose 和 Paradise 認為 (1979) 助人專業人員在進行倫理判斷的歷程會經歷五級倫理導向的思考: (1) 獎懲導向: 主要考量在行為結果是否會為自己帶來外界的賞罰; (2) 機構導向: 主要考量在行為表現是否符合自己的角色與機構的規定; (3) 社會導向: 主要考量行為表現是否符合社會的期待與法律的規範; (4) 個人導向: 主要考量行為表現是否合乎當事人的福祉與專業倫理守則的自律要求; (5) 良心導向: 主要考量行為表現是否符合個人道德良知的自我抉擇。

相較於西方倫理觀是由外而內的學習過程, 東方和華人的倫理觀則有一個由內而外的實踐過程, 我們可以從儒家的思想中觀察到這一點 (Wang, 2013, 2018, 2022), 《論語·里仁》有一段孔子與曾子的對話: 子曰:「參乎, 吾道一以貫之。」曾子曰:「唯」。子出, 門人問曰:「何謂也?」曾子曰:「夫子之道, 忠恕而已矣。」對此, 宋朝朱熹加以註解為:「盡己之為忠, 推己之為恕」《論語集注·里仁第四》, 也就是儒家所強調的「忠」, 也就是道德, 其核心的修為工作在「盡己」, 也就是自我要求; 儒家所強調的「恕」也就是倫理, 其核心的修為工作在「推己」, 也就是善待他人。對助人專業工作而言, 從助人者自我修為的「盡己」功夫開始, 透過專業助人的實踐過程, 要進一步做到善待當事人以保護其權益的「推己」功夫之目標。這是華人儒家文化對專業倫理意涵的探討重點。

將上述中西文化的倫理觀加以整合, 則可以發現倫理思考其實具有「外在客觀理性規範」與「內在主觀自我修為」的雙元性質, 也就是, 透過將上述的倫理判斷思維之雙元性質加以統整, 可以得到「助人專業倫理雙元模型」(如圖 1) (王智弘, 2013, 2018; Wang, 2022)。

圖 1

助人專業倫理雙元模型



助人專業倫理雙元模型

註: 引自王智弘著, 2018, 「諮商倫理議題的過去、現在與未來」, 載於蕭文、田秀蘭 (主編), 臺灣輔導一甲子 (356 頁), 心理出版社。

雙元模型以當事人權益保護與助人者自我修為作為兩端。Van Hoose 與 Paradise (1979) 五級倫理思考模式之前三級導向主要關注「外在客觀理性規範」，是屬於倫理的外元性質，而其第四、第五級導向則主要關注「內在主觀自我修為」，是屬於倫理的內元性質，外元層次強調要符合外在的理性規範，以保護「當事人的五大權益」為出發點，而以達到善盡「助人者的三大責任」為目標，而內元層次強調要符合內在的自我修為，以良心自我修為的「盡己」功夫為出發點，而以善待當事人的「推己」功夫為目標。也就是倫理的考量與判斷要同時兼顧外在客觀的理性規範與內在主觀的自我修養之內外元倫理思考，外元倫理思考以保護當事人權益的立場出發，透過符合法規要求的行為，以達到善盡助人者責任的目標，內元倫理思考以盡己的良心出發，透過符合良善動機與良善行動的自我修為，以達成符合良善結果以善待當事人的推己目標。因此，要因應發展 AI 所涉及的倫理議題，可能要從啟動公權力以制定外部管理規範，以及啟動 AI 專業人員（研發與營運組織）自律以確立內部自我要求，同時採取雙元倫理思維的策略才能因應目前的發展局勢。因為公權力立法程序較長但強制有效，專業（組織）自律強制性較低但可即時調整，政府立法與企業自律宜雙管齊下。

由 ChatGPT 的出現並不是一個 AI 研究成果的終點，而是新一波對話式人工智慧（Co-AI）研發的起點，但是，不同於過去的 AI 研究成果僅在科技圈獲得注意與討論，由於 ChatGPT 是一般大眾可以普遍使用的自然語言模型，ChatGPT 的出現因此造成了很大的「公眾覺醒」效應（楊孟軒，2023），也就是引起了全世界民眾對 AI 的巨大關注。除了 Google 公司在 Google Cloud 推出了 Dialogflow CX 與其他一系列 AI 產品之外（Google, 2023），國際商業機器公司（International Business Machines Corporation, IBM）推出 Watson Discovery（IBM, 2023），亞馬遜書店（Amazon）的 AWS（Amazon Web Services）平台也推出了 Accenture（Amazon, 2023），特斯拉公司的馬斯克也成立了一家 AI 新創公司 X.AI，要挑戰 OpenAI（陳冠榮，2023），Co-AI 以及所有 AI 模型的發展應該是前途未可限量的，而 AI 模型在諮商與所有領域上的可能應用，及其引發的文化與倫理的相關問題，也是值得持續加以探討的議題。在 AI 技術持續發展的過程中，公權力要投入對 AI 資料治理的監管（European Parliament, 2022; GPAI, 2021; OECD, 2019），對 AI 的科學發展計畫要建立倫理審查制度（Bartlett et al., 2023），AI 專業人員以及研發與營運組織要加強自律以確立內部自我要求與規範，本文提出人工智慧治理雙元模型（The Duality model of AI Governance）作為人工智慧涉及倫理問題的因應策略，如圖 2 所示。

AI 專業人員與企業盡量能維護 AI 服務使用者的五大權益：（1）自主使用權：尊重使用者能自主決定使用權與隱私權 AI；（2）使用受益權：讓使用者能從使用 AI 中受益；（3）免受傷害權：避免使用者因使用 AI 而受害；（4）公平待遇權：讓所有使用者都能有公平使用 AI 的機會；與（5）要求忠誠權：負責、穩定、可信與可靠的 AI 服務給使用者。而相對於這五大權益，AI 專業人員與企業則自我要求要負起：（1）專業責任：提供專業的服務；（2）倫理責任：提供合乎倫理的服務；（3）法律責任：提供符合法規要求的服務，並擔負必要的法律責任，以保障當事人的權益。

圖 2

人工智慧治理雙元模型



人工智慧治理雙元模型主張 AI 倫理的考量與判斷要同時兼顧外在客觀的理性規範與內在主觀的自我修為之內外元倫理思考，外元倫理思考以保護使用者權益的立場出發，透過符合法規要求的行為，以達到 AI 專業人員與企業責任的目標，內元倫理思考以盡己的良心出發，透過符合良善動機與良善行動的自我修為，來提供 AI 服務，以達成符合良善結果以善待使用者的推己目標。其中所謂的內在自我修為，也就是佛教所提倡的戒、定、慧的修行過程，從自我規範開始，自我規範、問心無愧則可得內心清淨，內心清淨則可得心神安定，心神安定則可得觀照分明而生智慧，以及要達到自我良知與人生智慧的提升，自我規範與自律是實踐過程的開始。

總而言之，人工智慧的治理，要同時啟動政府的公權力以制定外部的 AI 管理規範，比如經濟學人雜誌就建議各國政府可成立類似歐洲核子研究組織（CERN）的單位以研究 AI 的安全與倫理問題（李立心，2023），而 openAI 執行長奧特曼（Altman）在出席美國國會聽證會時也建議政府應考慮：設立新的部門以管理大型 AI 模型，並建立 AI 模型的安全標準以評估其風險，以及對 AI 模型進行關鍵指標的專家獨立審核制度等（王珮羽，2023），以作為制定相關規範的基礎。同時要啟動民間 AI 專業人員與企業的自律以確立內部自我要求，比如 OpenAI 公司本身對 AI 生成語言模型的倫理討論（Hagendorf, 2020）或 IBM 內部設立的倫理委員會等（王珮羽，2023），透過同時採取雙元倫理思維的策略，政府立法與企業自律雙管齊下，以因應 AI 快速發展所引發的倫理議題。

由於科技的不斷發展，將科技運用於民眾醫療服務，包括心理諮商與心理治療服務的趨勢自不可擋。對於運用通訊科技而非面對面進行諮商與心理治療服務，其實一開始在心理學界或政府部門都有反對的聲浪，理情行為療法（Rational Emotive Behavior Therapy, REBT）的創始人 Albert Ellis（1913-2007）是公開跳出來力挺的大師，主張當面對面不可得，運用不同通訊科技來進行心理治療完全是可行的（Ellis, 2005），由於臺

灣將網路科技應用於諮商服務已有二十年以上的發展歷史，雖然民眾有網路諮商的需求，國外政府也有將網路諮商納入常態服務的先例，但臺灣政府一直受限於「執行業務以一處為限」的醫療慣例，持反對立場，直到 2019 年 11 月 29 日才公告《心理師執行通訊心理諮商業務核準作業參考原則》加以認可，而此參考原則及其後續的修訂皆有不利民眾求助與規定不合時宜的爭議（林韋萱，2021），在政府的開放性或準備度不足的情況下，臺灣要發展科技在諮商上應用之路，前景可能仍不開朗。但是在 AI 科技快速發展的浪潮下，將 AI 運用於諮商服務工作之上，包括將目前發展之 ChatGPT 等 Co-AI 應用在輔助諮商工作上，或者未來進一步結合聲音與影像而發展出虛擬實境諮商（virtual reality counseling）或者元宇宙諮商（Metaverse Counseling），以及目前已經在開展中之精準諮商（precision counseling）等實務發展方向並不會停下腳步，也會帶來相對的倫理議題與挑戰（王智弘，2019，2022；王智弘、羅家玲，2021；沈慶鴻，2019；林淑君等，2021；Abrams, 2021; Fulmer, 2018, 2019; Jee & Heaven, 2021），因此，如何應用人工智慧以輔助心理諮商與心理治療工作，是諮商專業要持續投入的跨學科研究重點。而人工智慧在各專業領域的應用研究勢必將繼續火熱進行，相關的文化、倫理、法律等問題也將成為熱點話題。有關 AI 治理的討論，應該是產、官、學各界共同關心的頭等大事。而「人工智慧治理的雙元模型」則可能是處理人工智慧治理中涉及倫理問題的可行策略。

此外，本期本土諮商心理學刊收錄有四篇刊登文章，皆是本土化研究的重要研究結果，值得本刊讀者加以閱讀欣賞，首先是李佳澤、劉彥君與夏允中（2023）所撰寫的「儒家關係主義婚姻衝突歷程與關係轉化模式」，透過會通中西文化的思路以探討含攝中西文化背景脈絡下的婚姻衝突歷程內涵，以及婚姻關係的轉化是如何發生；其次是李仁豪（2023）所撰寫的「如何建構普世暨含攝文化理論」，則透過「一種心智，多重心態；普世主義，考量分疏」的文化心理學理路，提出先驗普世性和實徵普世性之「雙重普世進路」以做為建構本土社會科學普世暨含攝文化理論之參考；再其次是宋珮綺與許維素（2023）所撰寫的「焦點解決反思團隊督導成效之初探研究：以學校輔導與諮商人員為例」，則在探討焦點解決反思團隊督導對中小學校輔導與諮商人員的督導成效，可視為焦點解決督導模式的本土化研究；最後是鄧志平與李雅婷（2023）所撰寫的「工作意義是什麼？以中學教師為例」，則在試圖拓展及區辨工作意義之內涵，同時採用本位（emic）取向的研究視角，進行分析國內教師完成工作任務後之歷程式工作意義，作為建構工作意義本土模式之實務基礎。本土化研究是臺灣社會科學與諮商心理學研究的活路，值得大家一起來投入。

### 參考文獻

中華本土社會科學會、東方人文學術研究基金會、鵝湖月刊社（2019）。反思五四百年，重構文化中國。鵝湖月刊，527，3-4。

- 今周刊 (2023)。才導入 **ChatGPT** 不到 20 天...三星爆 3 起機密資料外洩！公司回應：加強內部監管。https://www.businessday.com.tw/article/category/183015/post/202304020018/?utm\_source=likr&utm\_medium=pubguess
- 文心一言 (2023年4月26日)。在**維基百科**。https://zh.wikipedia.org/zh-tw/文心一言
- 牛格正、王智弘 (2008)。助人專業倫理。心靈工坊。
- 王珮羽 (2023)。ChatGPT 之父：AI 很危險！他建議用「三道紅線」，防堵失控的人工智慧。數位時代，2023 年 5 月 18 日。https://www.bnext.com.tw/article/75294/ai-congressional-hearing-chatgpt-sam-altman
- 王智弘 (2004)。後心理師法時代社區諮商輔導機構發展的重要議題：服務轉型、倫理建立與媒體運用。九十三年全國生命線年會大會手冊，58-64。
- 王智弘 (2009)。網路諮商、網路成癮與網路心理健康。學富文化。
- 王智弘 (2013)。積極投入含攝文化理論建構：以助人專業倫理雙元模型為例。台灣心理諮商季刊，5 (3)，vi-xi。
- 王智弘 (2014)。含攝華人文化的心理諮商與心理治療觀：一次單元諮商模式的建構。台灣心理諮商季刊，6 (3)，vi-xv。
- 王智弘 (2016)。含攝儒家功夫論的本土專業倫理觀：從助人倫理雙元模型談儒家的倫理自我修為之道。台灣心理諮商季刊，8 (2)，vi-xii。
- 王智弘 (2017)。探索本土諮商心理學的發展理路。本土諮商心理學學刊，9 (1)，vii-x。
- 王智弘 (2018)。諮商倫理議題的過去、現在與未來。載於蕭文、田秀蘭 (主編)，台灣輔導一甲子 (335-368頁)。心理。
- 王智弘 (2019)。國際與本土、變遷與因應：網路諮商的本土發展。本土諮商心理學學刊，10 (3)，vi-xx。
- 王智弘 (2020)。含攝文化諮商的進路：從跨文化諮商、多元文化諮商到本土化諮商。本土諮商心理學學刊，11 (1)，vi-xii。
- 王智弘 (2022)。科技在諮商上的運用：網路諮商、精準諮商與元宇宙諮商。輔導季刊，58 (2)，1-12。
- 王智弘 (2023)。如何善用東西文化以開展本土化研究：以建構含攝文化理論的知識論策略與本土社會科學創造詮釋學為工具。本土諮商心理學學刊，14 (1)，vi-xx。
- 王智弘、朱建民 (2018)。新五四新挑戰：重構文化中國。(中國時報 2019/05/03 時論廣場A14版)
- 王智弘、夏允中、張蘭石、陳復、黃光國 (2019)。總結五四，再創未來：華人本土社會科學宣言，鵝湖月刊，527，5-8。
- 王智弘、楊淳斐、張勻銘 (2002)。生涯輔導服務網路化的做法與方向。生涯快遞，5，

32-36。

王智弘、劉淑慧、孫頌賢、夏允中（2017）。文化脈絡中的危機、轉機與復原力：本土諮商心理學研究的方向、目標與策略。**中華輔導與諮商學報**，**50**，1-28。

王智弘、羅家玲（2021）。台灣本土心理諮商的探索：從本土諮商到精準諮商。**本土諮商心理學學刊**，**12**（3），vi-xiv。

台灣人工智慧學校（2023）。**2023生成式AI衝擊工作坊**。<https://aigc2023.aiacademy.tw/>

宋珮綺、許維素（2023）。焦點解決反思團隊督導成效之初探研究：以學校輔導與諮商人員為例。**本土諮商心理學學刊**，**14**（2），77-109。

李仁豪（2023）。如何建構普世暨含攝文化理論。**本土諮商心理學學刊**，**14**（2），42-76。

李立心（2023）。人工智慧高速發展，威脅人類存亡？現在暫停AI，是過度反應。**天下雜誌**，**772**，129-130。

李光耀（2023，4月17日）。**ChatGPT與AI產業應用大未來**〔會議場次報告〕。「ChatGPT：如何應用AI提升你的競爭力」論壇，彰化，台灣。

李佳澤、劉彥君、夏允中（2023）。儒家關係主義婚姻衝突歷程與關係轉化模式。**本土諮商心理學學刊**，**14**（2），1-41。

李瑞全（2019）。當代新儒家對五四運動之評價與回應，**鵝湖月刊**，**527**，0-2。

吳家豪（2023）。台智雲發表企業級大型語言模型，避免中文答非所問。**中央社**。2023年5月17日。<https://www.cna.com.tw/news/ait/202305170113.aspx>

宋珮綺、許維素（2023）。焦點解決反思團隊督導成效之初探研究：以學校輔導與諮商人員為例。**本土諮商心理學學刊**，**14**（2），77-109。

沈慶鴻（2019）。諮商教育與服務的未來：人工智慧發展下的影響與回應。**本土諮商心理學學刊**，**10**（4），34-54。

林幸台、王智弘（2018）。台灣輔導與諮商專業的繼承與開展：本土輔導與諮商專業發展的回顧與探索。**中華輔導與諮商學報**，**53**，1-22。

林韋萱（2021）。通訊心理諮商開放緩不濟急：大疫當頭，上萬名個案誰來承接？**報導者**。<https://www.twreporter.org/a/covid-19-telecom-psychological-counseling>

林淑君、鄧志平、黃宗堅、王智弘、羅家玲、鄭鈴諭、陳嫻任（2021，10月）。**大學校園精準心理諮商與輔導系統之建置：以彰化師大為例**〔口頭論文報告〕。第四屆社會科學本土化學術研討會暨第五屆本土諮商心理學學術研討會。國立彰化師範大學，彰化縣。

張良知（2023）。**AI掀科技倫理議題、演算法偏見恐衝擊言論自由**。2023/1/2中央社（2023/1/22更新）。<https://www.cna.com.tw/news/ait/202301200011.aspx>

張麗卿（2021）。**AI倫理準則及其對臺灣法制的影響**。台灣人工智慧行動網。<https://ai.iias.sinica.edu.tw/ai-ethics-guidelines-in-taiwan/>

- 陳秉華 (2017)。臺灣是多元文化的社會。載於陳秉華 (主編)，*多元文化諮商在台灣* (3-26頁)。心理。
- 陳冠榮 (2023)。挑戰 OpenAI，馬斯克成立 AI 新創 X.AI。科技新報，2023年4月15日。  
<https://technews.tw/2023/04/15/elon-musk-plans-ai-startup-to-rival-openai/>
- 陳鍾誠 (2016)。人工智慧：採用JavaScript實作。https://ccckmit.github.io/aibook/html/eliza.html
- 曾荃鈺 (2023)。用ChatGPT超速學習！3指令3能力，成為善用AI協作的人才。獨立評論@天下，2023年3月7日。https://opinion.cw.com.tw/blog/profile/523/article/13367
- 黃亦筠 (2023)。全球首個繁體中文模型：國家隊打造「台版GPT」。天下雜誌，770，88-91。
- 黃亦筠、陳良榕 (2023)。AI夢幻軍團出列。天下雜誌，770，74-83。
- 黃光國 (2009)。儒家關係主義：哲學反思、理論建構與實徵研究。心理。
- 黃光國 (2015)。盡已與天良：破解韋伯的迷陣。心理。
- 黃光國 (2018)。內聖與外王：儒家思想的完成與開展。心理。
- 黃光國 (2019)。中西文明的夾縫：改變台灣命運的起手式。時報文化。
- 楊孟軒 (2023)。機器人會變成殺手嗎？AI大神「養孩子」哲學。天下雜誌，770，96-98。
- 葉耿志 (2020)。資料治理的數據風險管理與推動建議。國土及公共治理季刊，8 (3)，48-53。
- 詹婷怡 (2023，4月17日)。生成式AI發展與法制因應分析〔會議場次報告〕。「ChatGPT：如何應用AI提升你的競爭力」論壇，彰化，臺灣。
- 臺灣學術倫理教育資源中心 (2023)。3/30 歡迎報名線上講座：生成式AI對研究與學術倫理的影響。https://ethics.moe.edu.tw/news/detail/289/
- 劉亭妤 (2022)。英國14歲少女莫莉輕生，調查認定社群媒體提供不安全內容釀成悲劇。2022/10/04關鍵評論。https://www.thenewslens.com/article/174289
- 鄧志平、李雅婷 (2023)。工作意義是什麼？以中學教師為例。本土諮商心理學學刊，14 (2)，110-133。
- Abrams, Z. (2021). The promise and challenges of AI. *Monitor on psychology*, November/December, 62-69.
- Alan Turing. (2023, April 28). In *Wikipedia*. https://en.wikipedia.org/wiki/Alan\_Turing
- Amazon. (2023). *Conversational AI platform*. https://aws.amazon.com/solutions/consulting-offers/accenture-conversational-ai-platform/?nc1=h\_ls.
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602-614.
- Arora, A., Kaffee, L. A., & Augenstein, I. (2023). *Probing pre-trained language models for*

- cross-cultural differences in values*. <https://doi.org/10.48550/arXiv.2203.13722>
- Barnlund, D. C. (1990). Therapeutic communication. In G. Gumpert & S. L. Fish (Eds.), *Talking to strangers: Mediated therapeutic communication* (pp. 11-28). Ablex.
- Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2023). Computational scientific discovery in psychology. *Perspectives on Psychological Science*, 18(1), 178-189.
- Bickley, S. J., & Torgler, B. (2022). Cognitive architectures for artificial intelligence ethics, *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01452-9>
- Boer, P. M. (2001). *Career counseling over the internet: An emerging model for trusting and responding to online clients*. Lawrence Erlbaum Associates.
- Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). *Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study*. <https://doi.org/10.48550/arXiv.2303.17466>
- Caren, C. (2022). *AI writing: The challenge and opportunity in front of education now*. <https://www.turnitin.com/blog/ai-writing-the-challenge-and-opportunity-in-front-of-education-now>
- ChatGPT. (2023, April 30). In *Wikipedia*. <https://en.wikipedia.org/wiki/ChatGPT>
- Chechitelli, A. (2023). *Sneak preview of Turnitin's AI writing and ChatGPT detection capability*. <https://www.turnitin.com/blog/sneak-preview-of-turnitins-ai-writing-and-chat-gpt-detection-capability>
- Childress, C. A., & Asamen, J. K. (1998). The emerging relationship of psychology and the internet: Proposed guideline for conducting internet intervention research. *Ethics & Behavior*, 8, 19-35.
- Ellis, A. (2005). Foreword. In K. Derrig-Palumbo & F. Zeine, *Online therapy: A therapist's guide to expanding your practice*. W. W. Norton & Company.
- European Parliament. (2022). *Regulation of the European Parliament and of the Council on European data governance and amending regulation (EU) 2018/1724 (Data Governance Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868>
- Fink, J. (1999). *How to use computers and cyberspace in the clinical practice of psychotherapy*. Jason Arosen.
- Ford, B. D. (1993). Ethical and professional issues in computer-assisted therapy. *Computers in Human Behavior*, 9, 387-400.
- Fulmer, R. (2018). Counseling with artificial intelligence. *Counseling Today*, January 16, 2018. <https://ct.counseling.org/2018/01/counseling-artificial-intelligence/>
- Fulmer, R. (2019). Artificial intelligence and counseling: Four levels of implementation. *Theory*

- & *Psychology*, 22(1), 1-13.
- Future of Life Institute. (2023, March 22). *Pause giant AI experiments: An open letter*.  
<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Global Partnership on Artificial Intelligence. (2021 updated). *Frequently asked questions*.  
<https://gpai.ai/about/gpai-faq.pdf>
- Global Partnership on Artificial Intelligence. (2022a). *Working group on data governance*.  
<https://gpai.ai/projects/data-governance/>
- Global Partnership on Artificial Intelligence. (2022b). *Data governance: A framework paper for GPAI's work on data governance 2.0*. <https://gpai.ai/projects/data-governance/Data%20Governance%20-%20A%20Framework%20Paper%20for%20GPAI%E2%80%99s%20Work%20on%20Data%20Governance%202.0%20.pdf>
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative language models and automated influence operations: Emerging threats and potential mitigations*. <https://doi.org/10.48550/arXiv.2301.04246>
- Google. (2023). *Conversational AI for richer, more intuitive experiences*. <https://cloud.google.com/conversational-ai>
- Grohol, J. M. (2004). Online counseling: A historical perspective. In R. Kraus, J. S. Zack, & G. Stricker (Eds.), *Online counseling: A handbook for mental health professionals* (pp. 51-68). Elsevier Academic Press.
- Hagendorf, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30, 99-120. <https://doi.org/10.1007/s11023-020-09517-8>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, 466(1), 29.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, 33(2-3), 61-83.
- Hwang, K. K. (2012). *Foundation of Chinese psychology: Confucian social relations*. Springer.
- International Business Machines Corporation. (2023). *What is conversational AI?*  
<https://www.ibm.com/topics/conversational-ai>
- Jee, C. & Heaven, W. D. (2021). The therapists using AI to make therapy better. *MIT Technology Review*, December 6, 2021. <https://www.technologyreview.com/2021/12/06/1041345/ai-nlp-mental-health-better-therapists-psychology-cbt/>
- Kitchener, K. S. (1984). Intuition, critical evaluation and ethical principles: The foundation for ethical decision in counseling psychology. *The Counseling Psychologist*, 12(3), 43-55.
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman

- & L. W. Hoffman (Eds.), *Review of child development research*, Vol. 1 (pp. 381-431). Russel Sage Foundation.
- Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England Journal of Medicine*, 388(13), 1233-1239. <https://doi.org/10.1056/nejmsr2214184>
- Liebrezn, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, 5(3), E105-E106.
- Loebner Prize. (2023, March 22). In *Wikipedia* [https://en.wikipedia.org/wiki/Loebner\\_Prize](https://en.wikipedia.org/wiki/Loebner_Prize)
- Microsoft Bing. (2023, April 29). In *Wikipedia*. [https://en.wikipedia.org/wiki/Microsoft\\_Bing](https://en.wikipedia.org/wiki/Microsoft_Bing)
- Morrissey, M. (1997, November). NBCC WebCounseling Standards unleash intense debate. *Counseling Today*, 6, 8, 12.
- Nature. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613, 612. <https://www.nature.com/articles/d41586-023-00191-1>
- OpenAI. (2023a). *About*. <https://openai.com/about>
- OpenAI. (2023b). ChatGPT 對「ChatGPT 與 ELIZA 相較有何異同?」的問題回應。  
<https://chat.openai.com/>
- OpenAI. (2023c). ChatGPT 對「ChatGPT 可以應用在諮商上嗎?」的問題回應。  
<https://chat.openai.com/>
- OpenAI. (2023d). *Privacy policy*. <https://openai.com/policies/privacy-policy>
- OpenAI. (2023e). *What is ChatGPT?-Commonly asked questions about ChatGPT*.  
<https://help.openai.com/en/articles/6783457-what-is-chatgpt>
- Open-source software. (2023, April 19). In *Wikipedia*. [https://en.wikipedia.org/wiki/Open-source\\_software](https://en.wikipedia.org/wiki/Open-source_software)
- Organisation for Economic Cooperation and Development. (2019). *OECD AI Principles overview*. <https://oecd.ai/en/ai-principles>
- Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*.  
<https://doi.org/10.1007/s43681-023-00258-9>
- Recurrent neural network. (2023, April 7). In *Wikipedia*. [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network)
- Rucker, M. (2020). *Using artificial intelligence for mental health*.  
<https://www.verywellhealth.com/using-artificial-intelligence-for-mental-health-4144239>
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11, 887.

- Shapiro, D. E., & Schulman, C. E. (1996). Ethical and legal issues in E-Mail therapy. *Ethics & Behavior*, 6, 107-124.
- SmarterChild. (2023, April 26). In *Wikipedia*. <https://en.wikipedia.org/wiki/SmarterChild>
- Suler, J. (2002). The future of online clinical work. *Journal of Applied Psychoanalytic Studies*, 4(2), 265-270.
- Tait, A. (1999). Face-to-face and at a distance: The mediation of guidance and counseling through the new technologies. *British Journal of Guidance & Counseling*, 27(1), 113-122.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Turing test. (2023, April 17). In *Wikipedia*. [https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)
- Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. Simon & Schuster.
- Transformer (machine learning model). (2023, April 30). In *Wikipedia*. [https://en.wikipedia.org/wiki/Transformer\\_\(machine\\_learning\\_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))
- Van Hoose, W. H., & Paradise, L. V. (1979). *Ethics in counseling and psychotherapy: Perspectives in issues and decision making*. Carroll Press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, C. H. (2022). An inclusive theory of ethics based on Chinese culture: The Duality Model of Professional Ethics for Helpers. In A. K. Giri & S. C. Wu (Eds.) *Cross-currents of social theorizing of contemporary Taiwan* (pp. 187-206). Palgrave Macmillan.
- Wayback Machine. (2021). *Pattern matching overview*. <https://web.archive.org/web/20210513182221/https://docs.microsoft.com/en-us/dotnet/csharp/pattern-matching>
- Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- Whitty, M.T., & Joinson, A. N. (2009). *Truth, lies and trust on the Internet*. Routledge.
- Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). *Exploring AI ethics of ChatGPT: A diagnostic analysis*. <https://doi.org/10.48550/arXiv.2301.12867>

## **Duality Model of Artificial Intelligence Governance: On the Possible Application of ChatGPT and Conversational AI in Counseling and Its Ethical Issues**

Chih-Hung Wang\*

### **Abstract**

The chat robot ChatGPT has attracted much attention, and it has had a significant impact on both campuses and workplaces. It may continue to impact school education and work patterns that need to be observed. Although it is a major advancement in science and technology, it cannot ignore the cultural and ethical issues it contains. Since Turing (Alan Mathison Turing, 1912-1954) proposed the Imitation Game (Turing's test) in 1950 to answer the question of whether machines can think (Can machines think?), Weizenbaum published ELIZA in 1966, a computer program for natural language processing (NLP). ELIZA uses the Rogerian response grammar of the client-centered school to provide users with an experience similar to a conversation with a psychiatrist. It is one of the early examples of artificial intelligence (AI) being used in counseling situations. Since 1990, the Loebner Prize, based on the Turing test standard, has been held annually, and AI chatbots have also been developed. And its technology from the use of pattern matching calculus concept, with the gradual improvement of computer computing power, machine learning, deep learning, artificial neural network, and recurrent neuro network (RNN) have gradually appeared. Google used the attention mechanism to propose the program concept of Transformer, and Open AI developed the Generative Pre-trained Transformer (GPT) based on it. After continuously increasing the number of database parameters and computing power and using reinforcement learning from human feedback (RLHF) with artificial feedback, the much-anticipated ChatGPT was launched in November 2022. ChatGPT is being considered for consultation (or possible application in therapeutic communication) and is worth discussing, especially the cultural and ethical issues that may arise from it. These issues may involve cultural and value biases, correctness and availability of data, concentration and improper use of power, concerns about artificial general intelligence (AGI), the competency to provide professional services, mechanization or dehumanization,

informed consent in professional services, relationship building in professional services, diagnosis and technology use, confidentiality and duty to warn (reporting), and fairness of receiving services. These issues are worthy of concern. This paper proposes a Duality Model of Artificial Intelligence Governance as a countermeasure to the cultural and ethical issues caused by the development of AI.

*Keywords: culture, ethics, ChatGPT, Counseling, Duality Model of Artificial Intelligence Governance*

---

Chih-Hung Wang\*      Department of Guidance and Counseling, National Changhua  
University of Education ( ethicgm@gmail.com )

## **The Development of Artificial Intelligence Chatbots**

Artificial Intelligence (A.I.) chat robot ChatGPT (G.P.T. refers to Generative Pre-trained Transformer) has attracted much attention and has dramatically impacted both campus and workplace. Both school education and work patterns may continue to trigger significant impacts that need to be observed. Although it is a considerable advancement in science and technology, the cultural and ethical issues contained and derived from it cannot be ignored. Since Turing (Alan Mathison Turing, 1912-1954) (“Alan Turing,” 2023) proposed the imitation game in 1950, that is, Turing's test (“Turing test,” 2023) to answer whether the machine can think (Can machines think?) (Turing, 1950), the Massachusetts Institute of Technology (M.I.T.) professor Weizenbaum published a natural language processing (N.L.P.) computer program called ELIZA in 1966, using the client-centered Rogerian response grammar, which simulates human conversations to provide users with an experience similar to talking to a psychiatrist, can be said to be the earliest example of artificial intelligence being used in a similar counseling situation. ELIZA is psychology's most discussed artificial intelligence chat program (Fink, 1999; Grohol, 2004; Turkle, 1995; Wang, 2009; Whitty & Joinson, 2009). It is classified as the so-called automation consultation service (Suler, 2002), which may include the form of stand-alone computer software operation, the state of interaction through the Internet, the combination of Internet consultation expert system, and online virtual-reality counseling, etc. (Wang, 2009). The emergence of ChatGPT and its possible counseling application have already been conceived and expected by people.

The proposal of the Turing test has received extensive attention and response and is regarded as a critical issue in the development of AI. The Loebner Prize competition, based on the standards established by the Turing Test, has been held every year since 1990 and has also attracted A.I. experts worldwide to invest in research and development. The more famous award-winning works over the years (those who have won the championship three times or more), such as P.C. Therapist, A.L.I.C.E. (Artificial Linguistic Internet Computer Entity), Jabberwacky, and Mitsuku (later renamed Kuki, won five times, is the most award-winning work), etc. (“Loebner Prize,” 2023), these works reflect the technology and technology of A.I. developers at that time conception, thus stimulating the continuous development of A.I. chatbots.

In the programming design in the field of natural language processing, it is relatively easy to write and use programming methods. The writing technology of early chat programs is to use the calculation concept of pattern matching (Wayback Machine, 2021), and the

programming of these algorithms (“Algorithm,” 2023) focuses on guessing and grasping the user's communication intentions to use and comparing text through calculus rules, and to carry out subsequent dialogues and responses. Many so-called intelligent chat system programs are produced using these methods (Chen, 2023). ELIZA is a representative product of these algorithms. Although it is only a text interface, its results can allow the program to communicate continuously with the user and even make the user mistakenly believe that it is communicating with a natural person.

With the advent of GPU (Graphics Processing Unit)-based supercomputers and the gradual improvement of computer computing power (HashRate), machine learning, and deep learning (that need to process a large amount of data and mighty computing power), artificial neural network, and recurrent neural network (R.N.N.) and other concepts and technologies are gradually appearing, making A.I. and chat programs more and more powerful (Huang & Chen, 2023; “Recurrent neural network,” 2023). In the development of chat programs, there have always been attempts to seek commercialization. For example, SmarterChild is one of them. It was once applied to Instant Messenger of America Online (AOL) and Windows Live Messenger (formerly M.S.N. Messenger) of Microsoft Internet community service (“SmarterChild,” 2023). Although the development of chat programs ebbs and flows, people come and go, but people are always full of curiosity and expectations for the development of chat robots.

However, although computer players and information companies have continued to develop chat programs, the technology for processing text data has been limited in the number of input texts until Google proposed the concept of the attention mechanism. “The Attention Is All You Need” paper published at the 31st Conference on Neural Information Processing Systems in 2017 formally proposed the concept of “Transformer” (“Transformer (machine learning model),” 2023; Vaswani et al., 2017). they have made a significant breakthrough.

OpenAI quickly carried forward this concept. This company promised to ensure that artificial general intelligence (AGI) can benefit all humanity (OpenAI, 2023a); co-founded by Sam Altman (former Y Combinator CEO) and Elon Musk (Tesla CEO, who later claimed to be concerned about the conflict of interest with Tesla in developing self-driving AI and quitting the team), etc. in 2015. OpenAI Laboratory (which consists of the non-profit organization OpenAI Incorporated and its for-profit subsidiary OpenAI Limited Partnership), Open AI uses the concept of Transformer to develop ChatGPT and continues to be injected funds and provided Azure for computing by Microsoft Corporation (Microsoft). With the help of the online supercomputing platform, OpenAI can continue to grow by continuously increasing the number of database parameters and computing power. For example, Open AI launched GPT-1

in June 2018, with a parameter volume of 120 million, and GPT in November 2019. -2, the number of parameters is 1.5 billion, and GPT-3 was launched in June 2020 with parameters of 175 billion. To improve the response quality of GPT, Open AI company began to use the reinforcement learning of human feedback (Reinforcement Learning from Human Feedback, RLHF) and launched GPT-3.5 in March 2022. Since its chat effect is relatively mature, ChatGPT was officially launched after continuous improvement (“ChatGPT,” 2023). On February 7, 2023, Microsoft officially announced the integration of ChatGPT into New Bing (“Microsoft Bing,” 2023) and updated the data to supplement the ChatGPT database until June 2021. Although Open AI is named Open, based on the vast R&D and operating costs, it was initially an open source software (open source software, OSS) (“Open-source software,” 2023) position. After Microsoft invested 100% in OpenAI Limited Partnership, it opened a part-profit business operation model.

ChatGPT was born in November 2022 and shocked the world. ChatGPT' 's chat effect is impressive, and it is considered that it should pass the Turing test, and it has begun to have various imaginations about its application. After the author asked about the difference between ChatGPT and ELIZA, the response was as follows (OpenAI, 2023b): ChatGPT and ELIZA are chat programs based on natural language processing technology, but there are some apparent differences. The working principle of ELIZA is to match the text entered by the user with a pre-defined pattern to generate a response; However, although ELIZA can make the user feel understood, the program lacks accurate understanding and learning capabilities, and the response lacks flexibility and diversity. ChatGPT is a program based on deep learning, using the neural network architecture of “Transformer,” which can learn natural language through a large amount of data training, and generate smooth and natural dialogues, so it can better understand and respond to the user's words, It can also be more flexible and diverse in response and can continue to learn and evolve. Although ChatGPT' 's answer is full of confidence, it is generally appropriate. From ELIZA to ChatGPT, the development of AI chat programs seems to be different from generation to generation. Different from the single-function analytical AI of the past face recognition program, ChatGPT is a generative AI with a wide range of functions. With the support of Microsoft's funds and network computing platform, it combines algorithms, massive data, and powerful computing. Power shows a power generation function, making it popular and setting a historical record of reaching hundreds of millions of users within two months (Lee, 2023). However, in addition to commercial applications, the possible impact of ChatGPT on human society has also attracted unprecedented attention from all walks of life, including potential related applications in various fields, cultural issues, social impact issues,

legal issues, and ethical issues (Chang, 2023; Sallam, 2023; Zhan, 2023; Zhuo et al., 2023) continue to attract the attention of academics and the general public.

### **Ultimate Goals of AI and Ultimate Concerns about AI**

The ultimate goal of AI research and development is also the ultimate concern of human beings about the development of AI, which is the development of general artificial intelligence (AGI), especially the development of “embodied intelligence.” Fei-Fei Li, co-director of Human-Centered Artificial Intelligence (HAI), said: after AI is more and more able to perceive the environment, it may develop the ability to interact with the environment and produce self-learning, adapting to the environment and develop embodied intelligence (Yang, 2023). AI robots with “embodied intelligence” can learn by themselves in the environment like humans and continue to evolve. Will AI robots replace humans in a dominant position, not be controlled by humans, and even attack humans as enemies? as predicted in a science fiction film? These issues may sound surreal, but they are hidden worries in the hearts of human beings when they develop AI. Without dwelling on such future concerns, this article will focus on AI's benefits and immediate risks, including the possible application of ChatGPT and Conversational AI (Co-AI) in counseling and the related cultural and ethical issues.

### **Possible Application of ChatGPT and Co-AI in Counseling**

Due to the solid answering ability of ChatGPT, it has aroused surprise and concern from all walks of life. Among the various attempts and imaginations of the application of ChatGPT, the possible application of ChatGPT and Co-AI in counseling (or therapeutic communication) is considered necessary in research topics and development directions. So far, its potential applications include the following:

1. Application in the field of career counseling: Just as career counseling was the first reason for the application of online counseling (Boer, 2001; Wang et al., 2002), powerful databases back ChatGPT and Co-AI, Therefore, for career counseling or career consultation fields that require a large amount of information, ChatGPT and Co-AI are undoubtedly a beneficial career information source and interface, and they can indeed be applied to career counseling (Zeng, 2023).

2. Application of expert consultation system: Due to the significant improvement of ChatGPT and Co-AI in natural language processing technology, as well as the foundation of

robust database and computing capabilities, it can provide users with smooth expression and interactive experience rich in information, and it is possible as a function similar to an expert advisory system, it does have good conditions as an expert advisory system. In practice, the combination of computer interface or mobile APP can already be used as a function of the mental health care system (Rucker, 2020).

3. Application of online assessment system: If ChatGPT and Co-AI can be appropriately used with psychological assessment tools, they may be applied to provide online rapid assessment and interpretation of psychological assessment services.

4. Application of counseling auxiliary resources: When the clients have not been able to obtain counseling services provided by counselors, the interaction and communication functions of ChatGPT and Co-AI may be used as additional counseling service resources. Because of space and time constraints, when the clients fail to obtain the counselor's assistance, ChatGPT and Co-AI can provide other communication resources for the clients.

5. It can be used for the education and training of counselors: ChatGPT and Co-AI can be used as exercises for simulated cases to provide trainee counselors to practice as counselors (Shen, 2019) or to experience the experience of clients, the dialogue text can also be used as counselor training, teaching, and supervision for individual or group guidance to identify or improve counseling skills.

6. The application of generating case records: ChatGPT and Co-AI can organize the recordings of the consultation conversations between counselors and parties into case records. For example, GPT-4 and the Copilot program, based on it, organize doctor-patient discussions into medical records (Lee, 2023; Lee et al., 2023). Therefore, ChatGPT and Co-AI are expected to be applied to manage case records.

However, regarding the possible application of ChatGPT and Co-AI in counseling, the cultural and ethical issues that may arise cannot be ignored, including the ethical issues that may arise from the characteristics of ChatGPT and Co-AI itself and the helping profession; these are all points to be considered carefully. Just ask ChatGPT, “Can ChatGPT be applied to counseling?” The response is as follows (OpenAI, 2023c): The ChatGPT model has powerful natural language processing and generation capabilities. Suppose it is applied to counseling. Many limitations and challenges will be a concern; for example, it cannot provide the emotional insight and interpersonal skills that an actual counselor possesses to build a relationship of trust and empathy with the client. Secondly, it cannot have the professionalism required by an essential counselor knowledge and training to correctly assess the client's problems and provide practical advice and treatment. Finally, it cannot guarantee the privacy and confidentiality of

the client. Therefore, ChatGPT cannot completely replace the role of a natural counselor, but it can be used in some. It can be used as an additional tool to provide essential emotional support, information, and particular domain knowledge and advice. ChatGPT and Co-AI have robust database and language generation capabilities, but when they are applied to counseling services that also involve the issues of well-being and privacy of the clients, their cultural and ethical considerations must be more cautious.

### **Cultural issues involved in the application of ChatGPT and Co-AI**

First, regarding the cultural issues involved in ChatGPT and Co-AI, generative AI needs to feed massive data from the database. The database parameters provided to ChatGPT and Co-AI as training may mainly come from the United States or the English cultural areas. So, there will be problems of incomplete or biased cultural sampling that cultural psychology pays attention to. That is, as Arnett (2008) pointed out, between 2003 and 2007, among the articles published in major journals in the six major fields of psychology, 68% of the research subjects used in the research were from the United States, and even if the accumulated research subjects to proportion reaches 96%, and only includes western industrialized countries that account for 12% of the world's population. That is to say, the development of Western psychological theories in the past will be based on a WEIRD sample: it belongs to the Western, Educated, Industrialized, Rich, and Democratic cultural population samples (Henrich et al., 2010a, 2010b; Hwang, 2012). That is to say; when the sampling of academic research is biased or incomplete regarding cultural representation, the research conclusions may be limited and incomplete. Since the content response of ChatGPT and Co-AI is based on the deep learning of its English-based database parameters, if the sampling of its parameter training text is not entirely culturally representative, there may be a similar issue of cultural sampling bias that is worthy of concern. According to ChatGPT and Co-AI operating on these databases, the content of the responses to users may also have cultural incompleteness or cultural deviation. The empirical studies on ChatGPT and Co-AI, natural language models, have shown cultural and value preferences (Arora et al., 2023; Cao et al., 2023), which shows the cultural issues of ChatGPT and Co-AI. It is indeed cause for concern. Among these related cultural issues, noteworthy ones include:

1. The issue of the gap in cultural thinking: Western mainstream cultural traditions in Europe and the United States emphasize individualism thinking orientation. In contrast, Eastern and Chinese cultural traditions emphasize relationalism thinking orientation (Hwang, 2009). There will inevitably be cultural or civilizational conflicts among them (Hwang, 2019). When

ChatGPT and Co-AI built with the content of Western mainstream cultural databases become knowledge mentors or common-sense consultants for members of Eastern and Chinese cultural communities, the content of knowledge and common sense it provides is cultural content oriented by individualistic thinking, which will cause cultural conflicts and limitations in application.

2. The issue of cultural discourse power: When ChatGPT and Co-AI become the spokespersons of human cultural content, the cultural views obtained through an in-depth study of the range of English databases will become their value standpoints for expressing cultural opinions. The cultural discourse power of non-English, Eastern, and Chinese cultures will be relatively weak or have no place to speak. This is also why countries worldwide are developing their own large-scale GPT language models because ChatGPT was launched in the United States (Huang, 2023) because there are cultural discourse rights and AI trustworthiness considerations.

3. The issue of cultural colonization: when ChatGPT and Co-AI become strong cultural exporters, the value thinking and expression content of ChatGPT and Co-AI are the mouthpieces for spreading Western (or American) mainstream culture, while Western (or American) mainstream culture may cause cultural colonization to the whole world, especially to non-English, Eastern and Chinese cultural areas, and it is worthy of the government and the people's concern on such cultural issues. Because Taiwan's academic circles already have cultural reflections and appeals for sensitive Western academic colonization and the establishment of independent academic traditions (Chinese Indigenous Social Sciences Association et al., 2019; Li, 2019; Hwang, 2015; Wang et al., 2019; Wang & Zhu, 2019), in non-English, Asia, or regions with Eastern culture and Chinese culture, it is advisable to be conscious and concerned about the cultural issues involved in generative AI, and have corresponding coping.

We should hold a positive attitude towards the possible contribution of the development of ChatGPT and Co-AI to human civilization. However, we should still be concerned about the cultural issues involved. Observing the current development of Co-AI, whether based on cultural, public welfare purposes or commercial competition considerations, other countries and regions outside the United States are also actively developing other language models. For example, France has an open-source language model, “Bloom,” and South Korea has “HyperClova” (Huang, 2023). As far as the Chinese cultural area is concerned, there are also corresponding actions to develop chat robots with Chinese language models, including the “Wen Xin Yi Yan” created by Baidu in mainland China (ERNIE Bot) (“Wen Xin Yi Yan,” 2023) and Simplified Chinese language models such as “Enlightenment” initiated by Beijing Zhiyuan AI Research Institute, as well as Taiwan’s United MediaTek Innovation Base, Academia Sinica

Lexicon Group, National Institute of Education, France's open-source language model Bloom and Taiwan Web Service(TWS), which is in charge of operating the supercomputer "Taiwan Computing Cloud, TWCC" of the Taiwan Advanced Research and Education Network(TWAREN), jointly built a traditional Chinese language model called "Taiwan version of GPT" (Huang, 2023). This language model, called "Formosa Foundation Model," is established through 176 billion parameters, which is about the same level as GPT-3.5. Compared with the ChatGPT model, the proportion of Chinese data is relatively low (traditional Chinese is less than 0.1 %; Simplified Chinese is less than 2%), and the "Formosa Foundation Model" incorporates 30% of traditional Chinese materials for training (Wu, 2023), which is indeed indigenous.

It can be seen that the development of ChatGPT and Co-AI is a demonstration of the strength of Western scientific civilization (especially the United States) and a demonstration of its cultural power. Based on commercial, scientific, or cultural considerations, various countries or regions have also initiated an arms race of AI language models is like the indigenization of social science or counseling psychology that we have been concerned about for a long time. According to the needs of local users, incorporating their cultural connotations should be an issue that the government and the people should care about. They are also important issues that the technology industry and academia should be concerned about.

### **Ethical issues involved in the application of ChatGPT and Co-AI**

Secondly, regarding the ethical issues involved in ChatGPT and Co-AI, OpenAI has conducted considerable discussions on the possible problems and responses to AI-generated language models (Hagendorf, 2020). From 2022, the first year of AI-Generated Content (AIGC), all walks of life have begun to pay attention to and discuss the ethical issues that AI may involve (Bickley & Torgler, 2022; Goldstein et al., 2023; Prem, 2023; Taiwan AI Academy, 2023), where important issues include: copyright ownership and infringement issues, data governance, academic ethics, deep fake and fake news issues, etc. There are also ethical issues related to AI application in psychology (Bartlett et al., 2023), including bias caused by non-representative data samples, interpretability of AI calculation process (calculation program), clinical application considerations, etc. Although the OpenAI company has a statement on its privacy policy (OpenAI, 2023d), and also provides relevant instructions to users through a question and answer set (OpenAI, 2023e). Issues that may arise after the launch of ChatGPT have also been discussed (Chang, 2023; Zhuo et al., 2023), which may involve topics such as personal privacy protection, cultural and value bias, correctness and availability of data, and improper operation of influence. However, related ethical issues and doubts may still need to be continuously

discussed. For the ethical issues involved in the application of AI or ChatGPT, I would like to briefly discuss the following (Bartlett et al., 2023; Sallam, 2023; Zhuo et al., 2023):

1. Issues of copyright ownership and infringement: ChatGPT is a generative AI language model that requires database parameter training with a large amount of content, and other generative AI models with commercial application value, such as the Codex model for generating code and the one for generating vectors Embedding model, DALL-E model for generating pictures, etc. (Lee, 2023) all require content parameter training. Since the content generated by ChatGPT and Co-AI is drawn from many database parameter contents, how will the copyright of the content produced by ChatGPT and Co-AI be attributed? Is it original? Are there issues of plagiarism or copyright infringement? The resulting copyright lawsuits will also be a thorny issue. Whether it is OpenAI, Microsoft, or other AI language model operating organizations, they must face up to these copyright-related issues.

2. The issue of correctness and usability of data: Since the data produced by ChatGPT and Co-AI come from all walks of life, how to ensure the correctness and usability of the data is an important issue. Avoid false, wrong, or harmful (toxic) data from entering the model training database, ensure the accuracy and fairness of the model output data, and also ensure that the information is updated (ChatGPT database has limitations of transparency and only be updated until June 2021) and upgrade the model to avoid outdated output data and maintain the stability and reliability of the model (Zhuo et al., 2023). Due to the incorrect data produced based on ChatGPT and Co-AI, there may be errors in citations and references and the risk of disseminating inaccurate data (Sallam, 2023), which may cause users to make wrong decisions, mistakes, and harm, especially regarding physical and mental health issues that may cause damage or risk to users, are also issues that AI language models must be concerned about.

3. Issues of bias and imposition of culture and values: AI algorithms will be affected by the bias of data collection methods, equipment, and annotators; that is, the algorithm may learn bias (Chang, 2023), while the language model potential biases in building and training primarily from monolingual databases include: social stereotypes, unfair discrimination, exclusion of perspectives and cultural content in other languages, and monolingualism that inhibits the development of other languages (Zhuo et al., 2023). The issue of possible cultural bias in the data produced by ChatGPT and Co-AI has been discussed above. The latter two are exactly the crisis and hidden worries that academic circles worry about Taiwan's promotion of the national bilingual policy that only respects English. Here we will further discuss the bias and imposition of values, whether it is the values of gender equality, family values, religious values, race values, political values, economic values, life values, and even physical, psychological, spiritual, moral,

and even sexual values, will the data produced by ChatGPT and Co-AI involve the prejudice and imposition of values?

4. Issues of improper operation of social influence and deep fakes: When ChatGPT and Co-AI, and other generative AIs have significant social influence, if such influence is harmful content (violence, pornography, prejudice) or If it is improperly manipulated, its risk to the public begins to increase. Because the more widely it is used, the greater its influence on the people. After experiencing the tragedy of a British girl committing suicide due to unsafe content provided by social media algorithms (Liu, 2022), the negative social influence that online media may cause. With more public concern, concerns about the social influence of generative AI being misused and used in deep fakes have also increased. Regarding the social impact issues of ChatGPT, Co-AI, and other generative AIs, it is advisable to examine them from the practical point of view of health care to avoid misuse and harm (Sallam, 2023).

5. Protection of personal privacy and organizational security: The free version currently provided by ChatGPT will include user query data into its language model practice database, so the query data uploaded by users will not be kept confidential. The media reported that South Korea's Samsung used ChatGPT on March 11, 2023, less than 20 days after, and the leaked confidential information was reported (Business Today, 2023). To prevent malicious users from obtaining information from the language model and harming the privacy of individuals or the organization's security, the language model can use screening program technology and notify in advance to reduce such risks (Zhuo et al., 2023). Personal vigilance and organizational usage norms may also be necessary coping strategies for users.

6. Issues of academic ethics: As soon as ChatGPT was opened for use, the issues that immediately attracted attention were academic ethics, ranging from the originality and fairness of students' homework, reports, and exams to the authenticity of admission application materials and learning history files, Even dissertations, research reports, various literary competitions and academic achievement evaluations. When AI generates academic texts, there are concerns about the risk of bias, errors, citation errors, and research fraud (e.g., ghostwriting, falsified or fake research) (Sallam, 2023), as well as issues about copyright, copyright attribution, the issue of plagiarism and the ethical issues of authorship affirmation abound (Liebrenz et al., 2023). The academic community is indeed worried about this. The world-renowned journal Nature (2023) published its position and countermeasures for its concerns—implications for research and academic ethics. And Turnitin, which is famous for preventing academic plagiarism, has also developed its Turnitin AI function and launched Turnitin Originality as a tool for checking the authenticity of students' homework, which can be used to

detect AI-assisted writing or AI writing to identify whether there is cheating (Caren, 2022; Chechitelli, 2023). These issues also require using AI tools to deal with them.

7. The issue of data governance: The most critical issue of the application of AI or ChatGPT and Co-AI is the issue of data governance, which is also closely related to the issues mentioned above. Data governance is considered to have positive functions in protecting user data, protecting organizational data, improving user experience, and assisting corporate decision-making. In addition to the self-control of data managers, such data governance also requires public power Intervened, including the establishment of the institutional framework for data protection (organizational structure, privacy statement, protection policy) and operational processes (privacy protection process, awareness education, and training, indicator monitoring report, risk management process, cross-border transfer mechanism, international regulatory registration) (Ye, 2020). Regarding AI data governance, establishing and revising national and international norms and establishing cooperation platforms are necessary (Zhan, 2023). For example, the “Data Governance Act” (European Parliament, 2022) concept; the Artificial Intelligence Principles (OECD AI Principles, 2019) proposed by the Organization for Economic Cooperation and Development (OECD), and the “Global Artificial Intelligence Cooperation” established in 2020 to implement these principles The Global Partnership on Artificial Intelligence (GPAI) platform (GPAI, 2021), in which the Working Group on Data Governance was established. Its current focus includes: realizing data sharing through data trust and Realizing social interests, promoting data justice research and practice, and privacy enhancement technology projects (GPAI, 2022a). Professor Jing-Yi Liu of National Taiwan University is the only Taiwanese member of this working group (GPAI, 2022b).

To promote AI data governance simultaneously with the world, Taiwan should not only actively participate in international cooperation teams to exchange experiences but also start to promote the formulation of the “Basic Law for the Development of Artificial Intelligence” to make Taiwan's AI ethics guidelines legalized (Chang, 2021). The governance mechanism of AI is an unavoidable topic for the government, the people, the academic circle, and the industry.

### **Ethical issues that may be involved in the application of ChatGPT and Co-AI in counseling**

Secondly, in terms of ethical issues that may be involved in the application of ChatGPT and Co-AI in counseling, ChatGPT and Co-AI used in counseling services as media-mediated counseling may lead to ethical issues, including such (Wang, 2004, 2009):

1. The issue of legal qualifications: ChatGPT and Co-AI may be used in counseling communication. There may be two types of communication: one is “communication in therapy,” and the other is “therapeutic communication;” the most significant difference lies in the difference in “intent” (Barnlund, 1990). The former intends to treat, which is a therapeutic behavior to establish a therapeutic relationship, the treatment provider must have professional qualifications and licenses. And the latter intends to communicate, which is a positive and constructive change that occurs in general natural conversation (Wang, 2009). If ChatGPT and Co-AI are used as the first service, their legal qualification and ability to perform therapeutic tasks must be confirmed. If ChatGPT and Co-AI are used as the second service, ChatGPT and Co-AI do not need to go through legal qualifications. However, corresponding safety and ethical considerations are still required to avoid inappropriate responses causing harm to users, especially for clients with crisis risks and severe mental illnesses. The risks will increase significantly, and the dialogue service of ChatGPT and Co-AI needs to have a mechanism to assess and respond to these risks.

2. The issue of mechanization or dehumanization: If the counseling service is transmitted through the intermediary media of technological equipment, it may cause the clients to feel the mechanization or dehumanization of the counseling service (Tait, 1999), which may affect how to improve the effectiveness of counseling services through media presentation and content is very important to avoid causing such feelings to the clients (Wang, 2009). If ChatGPT and Co-AI are used as auxiliary functions for counseling services, how to improve the user's humanized experience on the interface is the focus of development, and further use of audio, video, or virtual reality technologies may be considered. To provide a simulated counseling service experience, the ethical issues raised by it will be more complicated.

3. Issues of informed consent and risk control: If ChatGPT and Co-AI perform counseling, their legal qualifications and abilities to perform treatment tasks need to be confirmed, and to carry out relevant, informed consent procedures, it is necessary to target ChatGPT and Co-AI for making a clear description of the characteristics and limitations of AI executive counseling services, especially the complete description of confidentiality and confidentiality limitations, to minimize possible risks to users or the general public. If ChatGPT and Co-AI do not provide treatment for the purpose, even if there is therapeutic communication during the chat, ChatGPT and Co-AI do not need to go through the legal qualification certification process. However, if any user discloses private information during the conversation, how can it be protected by the confidentiality rule? And the exception of privacy occurs as the emotions and intentions of self-harm or harm to others; there should be a mechanism for identification and referral for

counseling to minimize the risk. No matter whether ChatGPT and Co-AI intend to provide counseling services or not, from the experience of using chatbots from ELIZA and below, users seek to chat when they have emotional support and counseling needs. The interaction and counseling of robots seem inevitable. Therefore, the R&D and operation units of ChatGPT and Co-AI should not only strive for stability and credibility in program design but also eliminate bias and harm and identify unethical and high-risk input prompts (Zhuo et al., 2023) and consult the opinions of counseling and psychotherapy experts to meet the actual needs of professional ethics and risk control mentioned above.

4. Issues of relationship building: These issues include two aspects one is relationship confirmation, and the other is relationship establishment. Using non-face-to-face traditional counseling services, how to confirm the counseling relationship is a problem that needs to be faced. In particular, using ChatGPT and Co-AI to provide auxiliary counseling services, how to confirm the relationship between role responsibilities and rights and obligations is an important issue. The second is the establishment of relationships. If counseling services are only provided through media such as text or voice, the possibility of emotional projection, transference, and counter-transference may increase due to insufficient information (Childress & Asamen, 1998), which will relatively create relationships. The issue of alienation, in addition to the psychoanalytic treatment approach which may be used as a topic of analysis, examines the effects of changes in the therapeutic relationship.

5. Issues of misjudgment and misevaluation: Due to the limitation of the media communication methods of ChatGPT and Co-AI, it is not as complete as the presentation of face-to-face communication, and there may be problems of simplification and distortion of information, such as providing counseling or consultation services may cause errors in judgment and evaluation, resulting in errors in interpretation and response. Suppose there are program errors or data loss during data transmission. In that case, the risk of such problems will increase, and may not meet the needs of the clients and even cause harm (Childress & Asamen, 1998; Ford, 1993; Morrissey, 1997; Shapiro & Schulman, 1996; Wang, 2009), the developers and operators of ChatGPT and Co-AI should also take such risks Take into consideration.

6. Confidentiality and duty to warn (reporting) issues: Considering the above-mentioned ethical issues on risk management and control, the developers and operators of ChatGPT and Co-AI should pay special attention to privacy and confidentiality when users intend to seek counseling or consultation services, and through program design and operational risk control planning, minimize the personal and organizational risks of users due to leaks. In addition, when faced with the user's intention to harm himself or others, it is an exception for confidentiality.

From informed consent, judgment, and assessment to consideration of confidentiality issues, all must be included in the risk management and control mechanism to minimize possible risks to users or the general public. Adding the duty to warn and reporting mechanisms to program design and operational risk control (Wang, 2009) is the policy that the developers and operators of ChatGPT and Co-AI should consider.

7. Issues of fairness of service acceptance by users: Since people have different access to media and opportunities to use media, how to ensure that all people can get the media communication opportunities of ChatGPT and Co-AI somewhat is an issue that needs to be considered. The issue of accessibility must be considered in program design and user equipment conditions to increase the fair use opportunities of the clients (Wang, 2009). Investment in communication infrastructure by the government and private information and communication departments is also a critical approach.

8. Issues of bias and imposition of culture and values: In addition to the possible application of ChatGPT and Co-AI in counseling, it will involve the above-mentioned ethical issues of counseling services, and the content of the services it provides will also include the cultural appropriateness mentioned above in this article (Arora et al., 2023; Cao et al., 2023), as well as issues of prejudice and imposition of culture and values (Zhuo et al., 2023), that is, counseling services should have culturally appropriate considerations, That is, the consideration of indigenization, such issues have long been concerned by indigenous social sciences (Hwang, 2009, 2015, 2018) and indigenous counseling psychology (Chen, 2017; Lin & Wang, 2018; Wang, 2009, 2013, 2014, 2016, 2017, 2020, 2023; Wang et al., 2017) are more worthy of the attention of counseling professionals in the AI era.

### **The response of ChatGPT and Co-AI ethical issues: Duality Model of Artificial Intelligence Governance**

To respond to and think about the ethical issues that may be involved in the development of AI for human beings, it may still be necessary to return to the standpoint of how people think about ethical issues. According to the model of ethical decision proposed by Kitchener (1984), it is advocated professionals providing counseling services face ethical issues and must make judgments; they will simultaneously think at two levels: (1) intuitive judgment and (2) critical evaluation. The thinking of intuitive judgment is based on personal intuition, responding and responding immediately to ethical issues based on individual ethical awareness. However, such reactions may be incomplete. Therefore, it is necessary to conduct critical evaluation thinking

and further refer to relevant professional ethics codes, regulations, and ethical principles. To make ethical judgments, it is proposed that professionals should try their best to protect the five major welfare of the clients receiving counseling services: (1) the welfare of autonomy, (2) the welfare of beneficence, (3) the welfare of nonmaleficence, (4) the welfare of justice and (5) the welfare of fidelity. Compared with these welfare of the clients, professional helpers must take on three major duties: (1) professional duty, (2) ethical duty, and (3) legal duty (Niu & Wang, 2008) protects clients' welfare.

Wang (2013, 2018, 2022a) concluded that the ethical view of Western culture emphasizes the following: external, objective, and rational regulations, protecting the five significant types of client welfare as the starting point, and then asking the helpers to fulfill their three significant duties. This starts with obeying external norms to protect the rights and interests of the clients concerned and continues to strive to fulfill their responsibilities and improve their moral performance. The moral development process advocated by Kohlberg (1964) and Van Hoose and Paradise (1979) further revised and proposed the ethical development level of helpers. Van Hoose and Paradise (1979) believed that helping professionals would experience five levels of ethical orientation in making ethical judgments: (1) reward and punishment orientation: mainly consider whether the behavior results will bring external rewards and punishments to themselves; (2) Institutional exposure: The primary consideration is whether the behavior conforms to the regulations of one's role and the organization; (3) societal orientation: The primary concern is whether the behavior conforms to the expectations of the society and the legal norms; (4) individual orientation: self-discipline requirements of the professional code of ethics; (5) conscience orientation: The primary consideration is whether the behavior conforms to the self-decision of personal moral conscience.

Compared with the learning process of Western ethics from the outside to the inside, the ethics of the East and the Chinese have a practice process from the inside to the outside, which we can observe from Confucianism (Wang, 2013, 2018, 2022a). There is a brilliant dialogue between Confucius and the philosopher Tsang in *The Analects*. Master Confucius said, "Tsang! My way (Tao) is bound together with one continuous strand." The philosopher Tsang said: "Yes, I see." Then Confucius walked out, and others asked, "What's that?" The philosopher Tsang said: "The way of the Master is 'Zhong' and 'Shù,' nothing more" (Le Jin, *The Analects*).

The Song Dynasty Confucian Zhu Xi commented on this allusion: "Do what we can as Zhong (loyalty, self-requirement, and self-cultivation) and treat others well as Shù (kindness, goodness to others)" (Li Ren IV, Annotations of *Analects*). It is thus known that Confucianism emphasizes loyalty, that is, morality. Its core cultivation is doing oneself good, self-requirement,

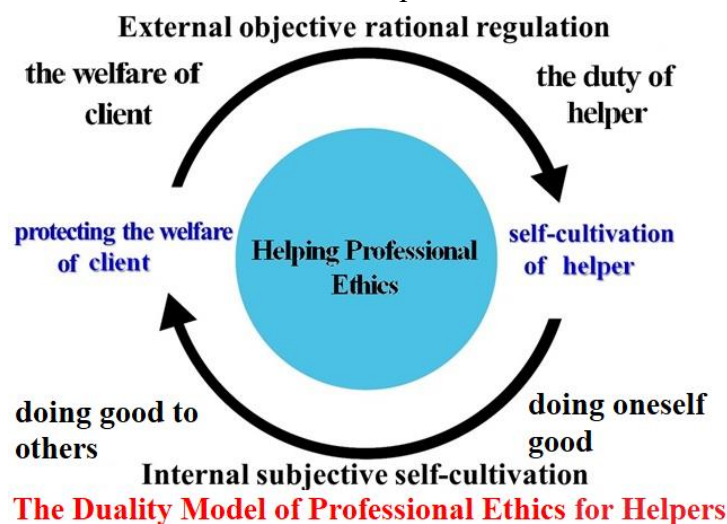
and self-cultivation; Confucianism emphasizes kindness and ethics. Its core cultivation is being considered to others, that is, doing good to others.

Wang (2013, 2018, 2022a) concluded that the ethical view of Chinese Confucian culture emphasizes internal subjective self-cultivation and helps the helper's self-cultivation (doing oneself good) as the starting point; it further asks the helper to be kind to the client (doing good to others).

And, Wang integrated the ethical thinking of Chinese and Western cultures and found that ethical thinking has a dual nature of “external objective rational norms” and “internal subjective self-cultivation,” so he proposed the “The Duality Model of Professional Ethics for Helpers” (Figure 1) (Wang, 2018, 2022a), to integrate the above-mentioned ethical judgment thinking in terms of its dual nature.

**Figure 1**

The Duality Model of Professional Ethics for Helpers



Note: Quoted from Wang, C. H. (2022a). An inclusive theory of ethics based on Chinese culture: The Duality Model of Professional Ethics for Helpers. In A. K. Giri & S. C. Wu (Eds.) *Cross-currents of social theorizing of contemporary Taiwan* (p. 192). Palgrave Macmillan.

The Duality Model of Professional Ethics for Helper stakes the protection of the client’s welfare and the helper’s self-cultivation as the two ends. The first three levels of the five-level ethical thinking model of Van Hoose and Paradise (1979) mainly focus on “external objective rational norms,” which belong to the external nature of ethics, while other the fourth and fifth levels of ethical thinking mainly focus on “internal subjective self-cultivation,” which belongs to the internal core of ethics. The external level emphasizes compliance with external rational

norms, with the starting point of protecting the “five welfare of the clients,” while the goal is to fulfill the “three duties of helpers.” In contrast, the internal level emphasizes conforming to internal self-cultivation, starting from the “doing oneself good” self-cultivation of conscience and “doing good to others” - being kind to the clients as the goal.

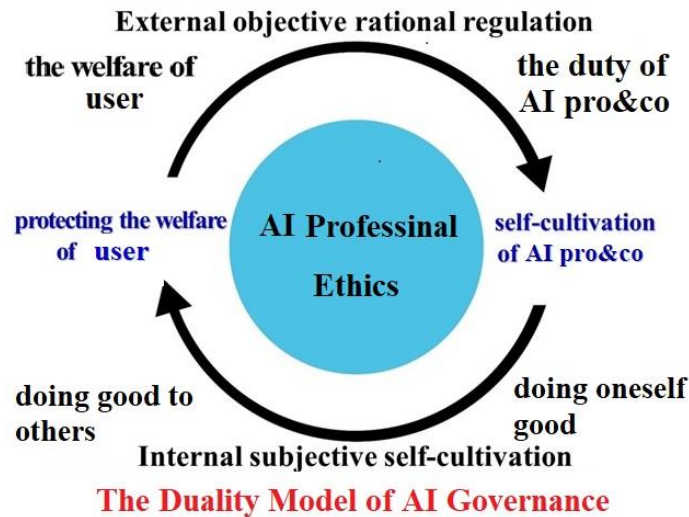
That is to say, ethical considerations and judgments should consider both external objective rational norms and internal subjective self-cultivation. External ethical thinking starts from the standpoint of protecting the welfare of the clients and achieves goodness through behaviors that meet the requirements of laws and regulations. The goal of helping others is to do oneself good. Internal ethics thinking starts with doing one’s conscience and through self-cultivation that conforms to good motives and good actions to achieve the goal of doing good to others-good results for clients. Therefore, to deal with the ethical issues involved in the development of AI, it may be necessary to activate public power to formulate external management norms and to initiate self-discipline of R&D and operation organizations to establish internal self-requirements. At the same time, a strategy of dual ethical thinking can be adopted to cope with the current development situation. Because the legislative process of public power is long but mandatory, and organizational self-discipline is less mandatory but can be adjusted immediately, government legislation and corporate self-discipline should go hand in hand.

The emergence of ChatGPT is not the end of AI research results but the starting point of a new wave of conversational artificial intelligence (Co-AI) research and development. However, since ChatGPT is a natural language model that the general public can commonly use, the emergence of ChatGPT has a significant “public awakening” effect (Yang, 2023). It has aroused great attention to AI from all over the world. In addition to Google launching Dialogflow CX and a series of other AI products on Google Cloud (Google, 2023), International Business Machines Corporation (IBM) launched Watson Discovery (IBM, 2023), Amazon Bookstore AWS platform also launched Accenture (Amazon, 2023), and Musk(Tesla) established an AI start-up company X.AI. to challenge OpenAI (Chen, 2023). The development of Co-AI and all AI models should be the future that cannot be limited; the possible application of AI models in counseling and all fields and the cultural and ethical issues it raises are also topics worthy of continuous discussion. In the process of continuous development of AI technology, public powers should invest in the supervision of AI data governance (European Parliament, 2022; GPAI, 2021; OECD, 2019) and establish an ethical review system for AI scientific development plans (Bartlett et al., 2023), AI R&D and operation organizations should strengthen self-discipline to establish internal self-requirements and norms. We can propose The Duality model

of AI Governance as the strategy to deal with ethical issues involved in AI, as shown in Figure 2.

**Figure 2**

The Duality model of AI Governance



From the Duality Model, AI professionals and corporations (AI pro&co) try their best to safeguard the five primary welfare of AI service users: (1) autonomy: respect for users to independently determine the use rights and privacy rights of AI; (2) beneficence: allow users to benefit from the use of AI ; (3) nonmaleficence: to prevent users from being harmed by using AI; (4) justice: to allow all users to have fair opportunities to use AI; and (5) fidelity: responsible, stable, trustworthy, and reliable AI services to users. Compared with this five welfare. AI professionals and enterprises require themselves to take (1) professional duty: provide professional services; (2) ethical duty: provide ethical services; (3) legal duty: provide services that meet legal requirements and bear the necessary legal responsibilities to protect the welfare of the users.

A two-pronged approach is adopted to deal with the ethical issues raised by the rapid development of AI. The ambidextrous model of artificial intelligence governance advocates that the consideration and judgment of AI ethics should consider both external objective, rational norms, and internal subjective self-cultivation. Internal and external meta-ethical thinking starts from the standpoint of protecting the welfare of users through compliance with laws and regulations. The required behaviors are to achieve the goals of AI professionals and corporate duty. Intrinsic ethical thinking starts by doing one's conscience and provides AI services through self-cultivation that conforms to good motives and good actions to achieve the goal of

conforming to conscience. The good result is the self-promotion goal of treating the user kindly.

This kind of internal self-cultivation can refer to the cultivation process of discipline, meditation, and wisdom advocated by Buddhism, starting from self-regulation and improving self-conscience. Self-discipline is the beginning. AI services through self-cultivation that conform to good motives and reasonable actions to achieve goodness. The result is to be kind to the user's self-promotion goal. Therefore, to govern the issue of AI ethics, it is necessary to activate public power to formulate external AI management norms simultaneously and, at the same time, start the self-discipline of AI professionals and corporate (AI pro & co) to establish internal self-requirements.

All in all, the governance of artificial intelligence must simultaneously activate the government's public power to formulate external AI management norms. For example, the Economist magazine suggested that governments of various countries can establish units similar to European Organization for Nuclear Research (CERN) to study AI safety and ethics. (Li, 2023), and openAI CEO Altman also suggested that the government should consider: setting up a new department to manage large-scale AI models and establishing safety standards for AI models to evaluate their Risks, as well as an expert independent review system for crucial indicators of AI models (Wang, 2023), as the basis for formulating relevant norms. At the same time, it is necessary to activate the self-discipline of private AI professionals and enterprises to establish internal self-requirements, such as OpenAI's ethical discussion on AI-generated language models (Hagendorf, 2020) or the ethics committee established within IBM (Wang, 2023). Adopt the strategy of dual ethical thinking, government legislation, and enterprise self-discipline to deal with the ethical issues caused by the rapid development of AI.

Due to the continuous development of science and technology, the trend of applying technology to public medical services, including counseling and psychotherapy, is unstoppable. There was some opposition in the psychology community or government departments at the beginning to using communication technology instead of face-to-face counseling and psychotherapy services. The founder of Rational Emotive Behavior Therapy (REBT), Albert Ellis (1913-2007), is a master who jumped out publicly and advocated that if face-to-face is not available, It is entirely feasible to use different communication technologies for psychotherapy (Ellis, 2005). Since Taiwan has used Internet technology for counseling services for more than 20 years, although the public has a demand for online counseling, foreign governments have precedents for incorporating online counseling into regular services already. Taiwan's government has been restricted by the medical practice of “limiting the execution of business to one place” and held an opposing position before. It was not until November 29, 2019, that

the “Reference Principles for the Approval of Psychologists Performing Telecommunication Counseling Business” was announced for approval. This reference principle, its original vision, and subsequent revisions are somewhat considered unfavorable to the public seeking help, and the regulations are too strict (Lin, 2021). In the case of the government's lack of openness or preparation, Taiwan may still have an uncertain future in developing the application of technology in counseling. However, under the wave of rapid development of AI technology, the application of AI to counseling services includes applying the currently developed Co-AI, such as ChatGPT, to assist counseling work or further combining audio and video in the future to create virtual reality counseling or metaverse counseling, as well as the practice development direction of precision counseling that is currently underway, the tendency will not stop, but will also bring relative ethical issues and challenges (Abrams, 2021; Fulmer, 2018, 2019; Jee & Heaven, 2021; Lin et al., 2021; Shen, 2019; Wang, 2019, 2022b; Wang & Luo, 2021).

Therefore, how to apply AI to assist psychotherapy and counseling work is the focus of continuous interdisciplinary research in the counseling profession. The research on AI applications in various professions and industries will hot continue, and the related issue of culture, ethics, and laws will be the topic irritable too. Discussion on AI governance should be a top issue of common concern to all walks of life in industry, government, and academia. And “The Duality model of AI Governance” may be the accountable strategy to deal with ethical issues involved in AI governance.

## References

- Abrams, Z. (2021). The promise and challenges of AI. *Monitor on psychology*, November/December, 62-69.
- Alan Turing. (2023, April 28). In *Wikipedia*. [https://en.wikipedia.org/wiki/Alan\\_Turing](https://en.wikipedia.org/wiki/Alan_Turing)
- Algorithm. (2023, April 19). In *Wikipedia*. <https://en.wikipedia.org/wiki/Algorithm>
- Amazon. (2023). *Conversational AI platform*. [https://aws.amazon.com/solutions/consulting-offers/accenture-conversational-ai-platform/?nc1=h\\_ls](https://aws.amazon.com/solutions/consulting-offers/accenture-conversational-ai-platform/?nc1=h_ls).
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602-614.
- Arora, A., Kaffee, L. A., & Augenstein, I. (2023). *Probing pre-trained language models for cross-cultural differences in values*. <https://doi.org/10.48550/arXiv.2203.13722>
- Barnlund, D. C. (1990). Therapeutic communication. In G. Gumpert & S. L. Fish (Eds.), *Talking to strangers: Mediated therapeutic communication* (pp. 11-28). Ablex.
- Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2023). Computational scientific discovery

- in psychology. *Perspectives on Psychological Science*, 18(1), 178-189.
- Bickley, S. J., & Torgler, B. (2022). Cognitive architectures for artificial intelligence ethics, *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01452-9>
- Boer, P. M. (2001). *Career counseling over the internet: An emerging model for trusting and responding to online clients*. Lawrence Erlbaum Associates.
- Business Today. (2023). *It's been less than 20 days since ChatGPT was introduced...Samsung exploded 3 leaks of confidential data! Company Response: Strengthen internal supervision*. [https://www.businesstoday.com.tw/article/category/183015/post/202304020018/?utm\\_source=likr&utm\\_medium=pubguess](https://www.businesstoday.com.tw/article/category/183015/post/202304020018/?utm_source=likr&utm_medium=pubguess)
- Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). *Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study*. <https://doi.org/10.48550/arXiv.2303.17466>
- Caren, C. (2022). *AI writing: The challenge and opportunity in front of education now*. <https://www.turnitin.com/blog/ai-writing-the-challenge-and-opportunity-in-front-of-education-now>
- Chang, L. C. (2021). *AI ethics guidelines and its implications for Taiwan's legal system*. Taiwan AI Wise Agent Network. <https://ai.iias.sinica.edu.tw/ai-ethics-guidelines-in-taiwan/>
- Chang, L. C. (2023). *AI raises technological ethical issues, and algorithmic bias may impact freedom of speech*. 2023/1/2 Central News Agency (2023/1/22 update). <https://www.cna.com.tw/news/ait/202301200011.aspx>
- ChatGPT. (2023, April 30). In *Wikipedia*. <https://en.wikipedia.org/wiki/ChatGPT>
- Chechitelli, A. (2023). *Sneak preview of Turnitin's AI writing and ChatGPT detection capability*. <https://www.turnitin.com/blog/sneak-preview-of-turnitins-ai-writing-and-chat-gpt-detection-capability>
- Chen, P. H. (2017). Taiwan is a multicultural society. In P. H. Chen (Ed.), *Multicultural Counseling in Taiwan* (pp. 3-26). Psychological Publishing.
- Chen, G. R. (2023). To challenge OpenAI, Musk established AI startup X.AI. *Science and Technology News*, April 15, 2023. <https://technews.tw/2023/04/15/elon-musk-plans-ai-startup-to-rival-openai/>
- Chen, Z. C. (2016). *Artificial Intelligence: Implemented in JavaScript*. <https://ccckmit.github.io/aibook/htm/eliza.html>
- Childress, C. A., & Asamen, J. K. (1998). The emerging relationship of psychology and the internet: Proposed guideline for conducting internet intervention research. *Ethics & Behavior*, 8, 19-35.

- Chinese Indigenous Social Sciences Association, Eastern Humanities Foundation, & Legein Society. (2019). Reflecting on the 100th Anniversary of the May 4th Movement, Reconstructing Cultural China. *Legein Monthly*, 527, 3-4.
- Ellis, A. (2005). Foreword. In K. Derrig-Palumbo & F. Zeine, *Online therapy: A therapist's guide to expanding your practice*. W. W. Norton & Company.
- European Parliament. (2022). *Regulation of the European Parliament and of the Council on European data governance and amending regulation (EU) 2018/1724 (Data Governance Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868>
- Fink, J. (1999). *How to use computers and cyberspace in the clinical practice of psychotherapy*. Jason Arosen.
- Ford, B. D. (1993). Ethical and professional issues in computer-assisted therapy. *Computers in Human Behavior*, 9, 387-400.
- Fulmer, R. (2018). Counseling with artificial intelligence. *Counseling Today*, January 16, 2018. <https://ct.counseling.org/2018/01/counseling-artificial-intelligence/>
- Fulmer, R. (2019). Artificial intelligence and counseling: Four levels of implementation. *Theory & Psychology*, 22(1), 1-13.
- Global Partnership on Artificial Intelligence. (2021 updated). *Frequently asked questions*. <https://gpai.ai/about/gpai-faq.pdf>
- Global Partnership on Artificial Intelligence. (2022a). *Working group on data governance*. <https://gpai.ai/projects/data-governance/>
- Global Partnership on Artificial Intelligence. (2022b). *Data governance: A framework paper for GPAI's work on data governance 2.0*. <https://gpai.ai/projects/data-governance/Data%20Governance%20-%20A%20Framework%20Paper%20for%20GPAI%E2%80%99s%20Work%20on%20Data%20Governance%202.0%20.pdf>
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative language models and automated influence operations: Emerging threats and potential mitigations*. <https://doi.org/10.48550/arXiv.2301.04246>
- Google (2023). *Conversational AI for richer, more intuitive experiences*. <https://cloud.google.com/conversational-ai>
- Grohol, J. M. (2004). Online counseling: A historical perspective. In R. Kraus, J. S. Zack, & G. Stricker (Eds.), *Online counseling: A handbook for mental health professionals* (pp. 51-68). Elsevier Academic Press.
- Hagendorf, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30, 99-120. <https://doi.org/10.1007/s11023-020-09517-8>

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, 466(1), 29.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, 33(2-3), 61-83.
- Huang, Y. Y. & Chen, L. R. (2023). AI fantasy army came out. *Commonwealth Magazine*, 770, 74-83.
- Huang, Y. Y. (2023). The world's first traditional Chinese model: the national team creates "Taiwan version of GPT". *Commonwealth Magazine*, 770, 88-91.
- Hwang K. K. (2009). *Confucian relationalism: Philosophical reflection, theoretical construction, and empirical research*. Psychological Publishing.
- Hwang, K. K. (2012). *Foundation of Chinese psychology: Confucian social relations*. Springer.
- Hwang, K. K. (2015). *Cultivation and conscience: Solving Webe's maze*. Psychological Publishing.
- Hwang, K. K. (2018). *Inner sageliness and outer kingliness: The accomplishment and unfolding of Confucianism*. Psychological Publishing.
- Hwang, K. K. (2019). *The gap between Chinese and western civilizations: The way to change Taiwan's destiny*. China Times Publishing.
- International Business Machines Corporation. (2023). *What is conversational AI?* <https://www.ibm.com/topics/conversational-ai>
- Jee, C. & Heaven, W. D. (2021). The therapists using AI to make therapy better. *MIT Technology Review*, December 6, 2021. <https://www.technologyreview.com/2021/12/06/1041345/ai-nlp-mental-health-better-therapists-psychology-cbt/>
- Kitchener, K. S. (1984). Intuition, critical evaluation and ethical principles: The foundation for ethical decision in counseling psychology. *The Counseling Psychologist*, 12(3), 43-55.
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research*, Vol. 1 (pp. 381-431). Russel Sage Foundation.
- Lee, K. Y. (2023, April 17). *ChatGPT and the future of AI industry applications* [Conference Session Report]. "ChatGPT: How to use AI to enhance your competitiveness" forum, Changhua, Taiwan.
- Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England Journal of Medicine*, 388(13), 1233-1239. <https://doi.org/10.1056/nejmsr2214184>
- Li, L. X. (2023). The rapid development of artificial intelligence threatens the survival of human

- beings? Suspending the AI now is an overreaction. *Commonwealth Magazine*, 772, 129-130.
- Li, R. Q. (2019). Contemporary New Confucianism's evaluation and response to the May 4th Movement, *Legein Monthly*, 527, 0-2.
- Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, 5(3), E105-E106.
- Lin, H. T., & Wang, C. H. (2018). Inheritance and development of Taiwan's guidance and counseling profession: Review and exploration of the development of local guidance and counseling profession. *Chinese Journal of Guidance and Counseling*, 53, 1-22.
- Lin, S. J., Deng, Z. P., Huang, Z. J., Wang, C. H., Luo, J. L., Zheng, L. Y., & Chen, Y. R. (2021, October). *Establishment of precision guidance and counseling system on university campus: Taking National Changhua University of Education as an example*. The 4th International Indigenous Social Science Conference & The 5th International Indigenous Counseling Psychology Conference. National Changhua University of Education, Changhua County.
- Lin, W. X. (2021). The slowdown of communication psychological counseling is not urgent: Who will undertake the tens of thousands of cases in the face of the epidemic? *Reporter*. <https://www.twreporter.org/a/covid-19-telecom-psychological-counseling>.
- Liu, T. Y. (2022). Molly, a 14-year-old girl in the United Kingdom, committed suicide. The investigation found that social media provided unsafe content and caused tragedy. 2022/10/04 *Critical review*. <https://www.thenewslens.com/article/174289>
- Loebner Prize. (2023, March 22). In *Wikipedia*. [https://en.wikipedia.org/wiki/Loebner\\_Prize](https://en.wikipedia.org/wiki/Loebner_Prize)
- Microsoft Bing. (2023, April 29). In *Wikipedia*. [https://en.wikipedia.org/wiki/Microsoft\\_Bing](https://en.wikipedia.org/wiki/Microsoft_Bing)
- Morrissey, M. (1997, November). NBCC WebCounseling Standards unleash intense debate. *Counseling Today*, 6, 8, 12.
- Nature. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613, 612. <https://www.nature.com/articles/d41586-023-00191-1>
- Niu, G. C. & Wang, C. H. (2008). *Helping Professional Ethics*. Psygarden.
- OpenAI. (2023a). *About*. <https://openai.com/about>
- OpenAI. (2023b). *ChatGPT's response to the question "What are the similarities and differences between ChatGPT and ELIZA?"*. <https://chat.openai.com/>
- OpenAI. (2023c). *ChatGPT's response to the question "Can ChatGPT be applied to counseling?"* <https://chat.openai.com/>

- OpenAI. (2023d). *Privacy policy*. <https://openai.com/policies/privacy-policy>
- OpenAI. (2023e). *What is ChatGPT?-Commonly asked questions about ChatGPT*. <https://help.openai.com/en/articles/6783457-what-is-chatgpt>
- Organisation for Economic Cooperation and Development. (2019). *OECD AI Principles overview*. <https://oecd.ai/en/ai-principles>
- Open-source software. (2023, April 19). In *Wikipedia*. [https://en.wikipedia.org/wiki/Open-source\\_software](https://en.wikipedia.org/wiki/Open-source_software)
- Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics* <https://doi.org/10.1007/s43681-023-00258-9>
- Recurrent neural network. (2023, April 7). In *Wikipedia*. [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network)
- Rucker, M. (2020). *Using artificial intelligence for mental health*. <https://www.verywellhealth.com/using-artificial-intelligence-for-mental-health-4144239>
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11, 887.
- Shapiro, D. E., & Schulman, C. E. (1996). Ethical and legal issues in E-Mail therapy. *Ethics & Behavior*, 6, 107-124.
- Shen, C. H. (2019). Future of counseling eand services: The impact and response of artificial Intelligence development. *Journal of Indigenous Counseling Psychology*, 10(4), 34-54.
- SmarterChild. (2023, April 26). In *Wikipedia*. <https://en.wikipedia.org/wiki/SmarterChild>
- Suler, J. (2002). The future of online clinical work. *Journal of Applied Psychoanalytic Studies*, 4(2), 265-270.
- Tait, A. (1999). Face-to-face and at a distance: The mediation of guidance and counseling through the new technologies. *British Journal of Guidance & Counseling*, 27(1), 113-122.
- Taiwan Academic Ethics Education Resource Center. (2023). *3/30 Welcome to sign up for the online lecture: The Impact of Generative AI on Research and Academic Ethics*. <https://ethics.moe.edu.tw/news/detail/289/>
- Taiwan AI Academy. (2023). *2023 Generative AI impact workshop*. <https://aigc2023.iaacademy.tw/>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Turing test. (2023, April 17). *Wikipedia* [https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)
- Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. Simon & Schuster.
- Transformer (machine learning model) (2023, April 30). In *Wikipedia*. [https://en.wikipedia.org/wiki/Transformer\\_\(machine\\_learning\\_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))

- Van Hoose, W. H., & Paradise, L. V. (1979). *Ethics in counseling and psychotherapy: Perspectives in issues and decision making*. Carroll Press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, C. H. (2004). *Important issues in the development of community counseling and counseling institutions in the age of post-Psychologists Act: Service transformation, ethical establishment, and media use*. Ninety-Three Annual National Lifeline Conference General Assembly Manual, 58-64.
- Wang, C. H. (2009). *Internet counseling, Internet addiction and Internet mental health*. Proed Publishing.
- Wang, C. H. (2013). Active involving with the construction of culture-inclusive theory: The Duality Model of Professional Ethics for Helpers as an example. *Taiwan Counseling Quarterly*, 5(3), vi-xi.
- Wang, C. H. (2014). Chinese culture-inclusive view of counseling and psychotherapy: The construction of single session counseling model. *Taiwan Counseling Quarterly*, 6(3), vi-xv.
- Wang, C. H. (2016). Indigenous view of Confucius culture-inclusive profession ethics: From the Duality Model of Helping Professional Ethics to Confucius way of ethical self-cultivation.. *Taiwan Counseling Quarterly*, 8(2), vi-xii.
- Wang, C. H. (2017). Exploring the development of indigenous counseling psychology. *Journal of Indigenous Counseling Psychology*, 9(1), vii-x.
- Wang, C. H. (2018). Discuss the past, present and future of ethical issues. In W. Xiao & X. L. Tian (Eds.), *Taiwan counseling sixty years* (pp. 335-368). Psychological Publishing.
- Wang, C. H. (2019). International and indigenous, change and response: The indigenous development of online counseling. *Journal of Indigenous Counseling Psychology*, 10(4), vi-xvi.
- Wang, C. H. (2020). Approaches with cultural-inclusive counseling: From cross-cultural counseling, multicultural counseling to indigenous counseling. *Journal of Indigenous Counseling Psychology*, 11(1), vi-xii.
- Wang, C. H. (2022a). An inclusive theory of ethics based on Chinese culture: The Duality Model of Professional Ethics for Helpers. In A. K. Giri & S. C. Wu (Eds.) *Cross-currents of social theorizing of contemporary Taiwan* (pp. 187-206). Palgrave Macmillan.
- Wang, C. H. (2022b). Application of technology in counseling: Internet counseling, precision

- counseling and metaverse counseling. *Counseling Quarterly*, 58(2), 1-12.
- Wang, C. H. (2023). How to make good use of Eastern and Western cultures to carry out indigenization research: Using “Epistemological Strategy for Constructing Culture-inclusive Theories” and “Indigenous Social Science Creative-Hermeneutics” as tools. *Journal of Indigenous Counseling Psychology*, 14(1), vi-xx.
- Wang, C. H., & Jue, J. M. (2018). New May Fourth new challenge: Reconstructing cultural China. (*China Times* 2019/05/03, Times Square, p. A14)
- Wang, C. H., Liu, S. H., Sun, S. H., & Shiah, Y. J. (2017). Crisis, chance, resilience and culture context: Probing into the direction, goal and strategy of indigenous counseling psycholog. *Chinese Journal of Guidance and Counseling*, 50, 1-28.
- Wang, C. H., & Luo, J. L. (2021). Exploration of local psychological counseling in Taiwan: From indigenous counseling to precision counseling. *Journal of Indigenous Counseling Psychology*, 12(3), vi-xiv.
- Wang, C. H., Xia, Y. Z., Chang, L. S., Chen, F., & Hwang, K. K. (2019). Summarizing the May Fourth Movement and recreating the future: A manifesto of Chinese indigenous social sciences. *Legein Monthly*, 527, 5-8.
- Wang, C. H., Yang, C. F., & Chang, Y. M. (2002). The practice and direction of Internet-based career counseling services. *Career Express*, 5, 32-36.
- Wang, P. Y. (2023). The father of ChatGPT: AI is dangerous! he suggested using “three red lines” to prevent out-of-control artificial intelligence. *Business Next*, May 18, 2023. <https://www.bnext.com.tw/article/75294/ai-congressional-hearing-chatgpt-sam-altman>
- Wayback Machine. (2021). *Pattern matching overview*. <https://web.archive.org/web/20210513182221/https://docs.microsoft.com/en-us/dotnet/csharp/pattern-matching>
- Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- Wen Xin Yi Yan. (2023, April 26). In *Wikipedia*. <https://zh.wikipedia.org/zh-tw/%E6%96%87%E5%BF%83%E4%B8%80%E8%A8%80>
- Whitty, M.T., & Joinson, A. N. (2009). *Truth, lies and trust on the Internet*. Routledge.
- Wu, J. H. (2023). Taiwan Web Service published a large-scale enterprise-level language model to Avoid giving irrelevant answers to Chinese questions. *CNA*. May 17, 2023. <https://www.cna.com.tw/news/ait/202305170113.aspx>
- Yang, M. X. (2023). Will robots become killers? AI master’s philosophy of “raising children”. *Commonwealth Magazine*, 770, 96-98.
- Ye, G. Z. (2020). Data risk management and promotion recommendations for data governance.

*Public Governance Quarterly*, 8(3), 48-53.

Zeng, Q. Y. (2023). Learn superfast with ChatGPT! 3 prompts and 3 abilities, become a talent who makes good use of AI collaboration. *Independent review @ Common Wealth*, March 7, 2023. <https://opinion.cw.com.tw/blog/profile/523/article/13367>

Zhan, T. Y. (2023, April 17). *Generative AI development and legal response analysis* [Conference Session Report]. “ChatGPT: How to use AI to enhance your competitiveness” forum, Changhua, Taiwan.

Zhuo, T. Y., Huang, Y. J, Chen, C. Y, & Xing, Z. C. (2023). *Exploring AI ethics of ChatGPT: A diagnostic analysis*. <https://doi.org/10.48550/arXiv.2301.12867>