



前瞻科技與管理 14 卷 2 期，1-32 頁（2026 年 5 月）  
Journal of Advanced Technology and Management Vol. 14, No. 2, pp. 1-32 (May, 2026)  
DOI:10.6193/JATM.202605\_14(2).0001

# 人工智慧多代理人辯論於公共政策利害關係人 分析之探索性研究

陳英傑<sup>1,\*</sup> 吳斯茜<sup>2</sup>

<sup>1</sup> 中央警察大學資訊管理學系助理教授

<sup>2</sup> 中央警察大學行政管理學系副教授

## 摘要

本研究探討多代理人辯論（Multi-Agent Debate, MAD）結合大型語言模型（Large Language Models, LLMs）於複雜決策的創新應用。為解決傳統會議效率不彰與人為偏見，開發 API（Application Programming Interface）協作平臺，以「職場霸凌防治」為實證案例，由多種 LLMs 模擬企業主管、受害者、法律專家、心理專家及政府官員五個關鍵利害人進行結構化辯論。研究發現 MAD 機制能高效整合多元觀點，產出包含立場差異、共識與行動方案之結構化決議。透過「角色一致性、論證邏輯性、觀點全面性、建議可行性」四大指標評估顯示，各 LLMs 表現差異顯著，Claude 在角色扮演上表現最優，Grok 則較具模板化傾向。本研究結論認為 MAD 系統可作為實體會議前的有效決策支援工具，協助會議主持人預先掌握各方觀點與潛在爭議，顯著提升決策品質與會議效率。

**關鍵詞：**人工智慧、大型語言模型、多代理人辯論、決策支援、職場霸凌

\* 通訊作者：陳英傑

電子郵件：bomy321@mail.cpu.edu.tw

（收件日期：2025 年 12 月 2 日；修正日期：2025 年 12 月 23 日；接受日期：2026 年 1 月 9 日）



# An Exploratory Study of AI Multi-Agent Debate for Public Policy Stakeholder Analysis

Ying-Chieh Chen<sup>1,\*</sup>, Szu-Chien Wu<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Information Management, Central Police University  
<sup>2</sup>Associate Professor, Department of Administrative Management, Central Police University

## Abstract

This study explores the innovative application of Multi-Agent Debate (MAD) integrated with Large Language Models (LLMs) for complex decision-making. To mitigate the inefficiencies and cognitive biases inherent in traditional meetings, an API (Application Programming Interface)-based collaborative platform was developed. Using “workplace bullying prevention” as an empirical case study, the research simulated structured debates among five key stakeholders—corporate executives, victims, legal experts, psychological experts, and government officials—represented by various LLMs. The findings demonstrate that the MAD mechanism efficiently integrates diverse perspectives, yielding structured resolutions that encompass positional differences, consensus, and concrete action plans. Performance was evaluated across four key metrics: persona consistency, argumentative logic, breadth of perspectives, and feasibility of recommendations. The evaluation reveals significant performance variations among LLMs; specifically, Claude excelled in role-playing and empathy, whereas Grok exhibited a more formulaic (templated) response pattern. This research concludes that the MAD system serves as an effective pre-meeting decision support tool, enabling chairpersons to identify potential controversies and diverse viewpoints in advance, thereby significantly enhancing decision quality and meeting efficiency.

**Keywords:** artificial intelligence, large language models, multi-agent debate, decision support, workplace bullying

---

\* Corresponding Author: Ying-Chieh Chen  
E-mail: bomy321@mail.cpu.edu.tw

## 壹、前言

人工智慧 (Artificial Intelligence, AI) 近年正以飛快速度席捲全球，成為政府數位轉型、產業升級的核心驅動力。AI 技術持續突破，貫穿軟硬體並跨足各行各業，從數位化升級到 AI 化，重新定義人機互動模式和行業邊界。AI 已然形成完整產業生態系，融合雲端、物聯網、大數據、虛擬實境與擴增實境等技術成為人工智慧物聯網 (Artificial Intelligence of Things, AIoT) 新應用。2022 年 11 月 30 日 OpenAI 發布 ChatGPT 成為重要里程碑，這個基於大型語言模型 (Large Language Models, LLMs) 的聊天機器人可進行自然對話，並執行文字生成、問答、摘要、程式編寫等複雜語言工作。ChatGPT 推出後，Anthropic Claude、Google Gemini 等 LLMs 如雨後春筍般出現。

AI 代理人 (Agent) 是 AI 重要應用，它能依據工作流程、運用工具完成任務，涵蓋決策制定、問題解決、環境互動等多元功能，在軟體設計、資訊科技 (Information Technology, IT) 自動化、程式碼生成等領域展現強大應用潛力。AI 代理人是一種能夠自主設計工作流程並運用各種工具來完成任務的智能系統，不僅限於自然語言處理 (Natural Language Processing, NLP)，更涵蓋決策制定、問題解決、環境互動和行動執行等多元功能 (Gutowska, 2025)。這些代理系統在組織環境中展現出強大的應用潛力，從軟體設計、IT 自動化到程式碼生成和對話助手等領域都能發揮作用。

2018 年 OpenAI 的研究團隊應用 AI 代理人提升語言模型正確性與可靠性，提出多代理人辯論 (Multi-Agent Debate, MAD) 的概念 (Irving, Christiano, and Amodei, 2018)，主要為運用 AI 代理人之間相互辯論來提高 LLMs 決策的完整性和可靠性。透過設計辯論環境，讓多個 AI 模型相互質疑並提供不同論點，有助於揭露潛在的錯誤和偏見，最終得出更可靠的結論。MAD 的特色是整合多元觀點，透過鼓勵批判性評估和探索不同的論點，來提高議題討論的廣度、品質與準確性 (Dell Technologies, 2024)。相較於傳統單一模型的局限性，單一模型容易受到特定模型訓練偏誤影響，Amazon Bedrock 研究團隊更於研究發現 (Shu, Das, Yuan, Sunkara, and Zhang, 2024)，針對複雜任務，透過多代理協作之平行溝通與資訊引用可提高 90% 的成功率，其正確率也提高了 70%。

傳統會議在決策過程中面臨諸多效率與效能挑戰，已成為現代職場亟需解決的問題。根據《哈佛商業評論》(Harvard Business Review) 研究顯示，高階主管平均每週花費近 23 小時參與會議，較 1960 年代的不到 10 小時大幅增長，且 71% 的資深管理者認為多數會議既無生產力又缺乏效率 (Perlow, Hadley, and Eun, 2017)。Doodle (2025) 針對 1,900 萬場會議的研究發現，組織不當的會議在美國造成高達 3,990 億美元的經濟損失，在英國則達 580 億美元。Golden Steps ABA (2023) 的研究也顯示，員工平均每月花費 31 小時在無效會議上，其中 50% 的時間被視為浪費，65% 的受訪者表示會議打斷其工作流程，71% 認為會議既無生產力又無效率。

傳統決策會議常缺乏結構化討論機制，容易受到個別發言者影響而偏離議題焦點，且參與者往往因階級關係或人際考量而無法充分表達多元觀點。此外，人類決策者易受限於認知偏誤、情緒干擾和既有框架思維，無法全面性地分析複雜問題的多重面向。面對這些挑戰，

AI 多代理人會議系統應運而生。相較之下，AI MAD 機制能夠突破這些限制，這些 AI 代理人不僅能夠自主進行推理、規劃與決策制定，更能透過工具整合、記憶保持與多代理人工作流程，實現超越單一模型能力的複雜任務處理，並且透過多元觀點的系統性整合、客觀理性的論證分析，以及長時間不間斷的深度討論，為組織提供更全面、更可靠的決策支援，有效提升會議前期準備的廣度與深度。

基於 LLMs 的多代理人系統 (Multi-Agent Systems, MAS) 能夠透過角色分工與協作機制，模擬人類團隊合作行為並進行結構化辯論，有效整合多元觀點並提升決策品質 (Gao et al., 2024)。本研究嘗試將 MAD 與主流 LLMs 做整合與實作，使用 Python 程式語言，提供設定會議主題並賦予各 LLMs 會議角色及背景資訊，讓各 LLMs 於虛擬會議中進行腦力激盪、發想並協助提出可能的解決方案，最終產出會議紀錄及行動方案。本研究目的著重於實體會議前的決策支援，提供實際會議主持人先行瞭解會議議題之利害關係人關心項目及因應對策，探討如何藉由多人 LLMs 會議機制，進一步協助促進會議決策的廣度與深度。

為驗證可行性，本研究情境模擬設定為職場霸凌 (Workplace Bullying) 防治因應對策會議。尤以近期職場霸凌受到社會輿論的重視，國內重大案件是 2024 年勞動部勞動力發展署北分署發生公務員遭霸凌而輕生案，雖然啟始只是勞動部的個案，但陸續有其他機關也揭露了職場霸凌案件，代表職場霸凌問題的嚴重性不容忽視。2025 年 5 月 1 日考試院通過考績法修正草案，將職場霸凌及酒後駕車、性侵害、性騷擾等重大違失行為，增列得一次記兩大過予以免職 (考試院, 2025)。職場霸凌的討論會議通常會涉及多方利害關係人不同立場，適合導入 MAD 技術，藉此進行探索可行性測試，並針對職場霸凌議題提出 AI 導入應用的解決方案。本研究亦將針對角色設計的影響進行評估，如在 MAD 框架中，不同的代理人角色組合如何影響辯論過程的「觀點收斂速度」與「最終共識的品質」。

## 貳、多代理人相關技術探討

本節將針對 MAS 及 MAD 之發展及技術進行討論。

### 一、MAS

Bond and Gasser (1988) 早期即提出 MAS 的構想，MAS 是由多個相互作用的智慧代理人組成的電腦化系統，MAS 可以解決對於單個代理人或單體系統來說困難或不可能解決的問題。MAS 也是一種由多個自主智慧代理人所組成的系統，這些代理人在同一環境中彼此協作、互動或競爭，以共同解決複雜問題或達成系統目標 (AI21 Lab, 2025)。MAS 的核心特性包括：

- (一) 多個自主代理人 (Autonomous Agent)：每個代理人都是一個具備自主性、能根據自身目標和規則獨立運作的軟體或硬體實體。代理人能感知環境、做出決策並執行行動。
- (二) 分散式決策：MAS 不依賴單一中央控制，而是透過多個代理人的決策與互動，實現系統層級的協同與適應。

- (三) 協作與競爭：代理人之間可以合作（如分工協作、共享資訊），也可能競爭（如資源分配或任務爭奪），以達到最佳整體效益。
- (四) 動態與靈活性：MAS 能根據即時資訊、環境變化和目標調整行為，適合處理複雜、多變的應用場景。

早在 LLMs 出現之前，MAS 在解決複雜協調問題上便已展現成果。例如，Binyamin and Slama (2022) 的研究便證實 MAS 能在智慧電網中用於資源分配和調度，透過協調不同能源來源（火力、風力、太陽能、水力發電廠）的輸出，優化能源使用效率，確保穩定可靠的電力供應，同時減少浪費，有效優化智慧電網的資源分配，達到比傳統熱管理多節省 20% 能源、降低尖峰負載 25%、減少瓦斯 23% 用量、節省空調 17% 電能等成果。在製造業應用中，MAS 透過代理人之間的協商與協調機制，並結合動態派工法則，來協調電腦數值控制（Computer Numerical Control, CNC）機器、物料搬運系統（Rail Guided Vehicle, RGV）等設備的活動，以動態分配任務和資源，此系統的目的是彈性調整製造流程，以應對多樣化、小批量的市場需求。Chen and Chen (2010) 的研究證實，這種方法能提升生產效率，並有助於找出達成最大產出的最佳派工組合。

## 二、MAD

### （一）MAD 的起源與定義

MAD 的概念源於 2018 年 OpenAI 研究團隊的研究論文“AI safety via debate”（Irving et al., 2018），該研究提出 MAD 概念並透過實驗證實，藉由兩個 AI 代理人針對問題進行辯論，並由人類裁判選出最真實的答案，其決策的準確性顯著高於單一 AI 模型的回答，成功展示辯論機制在 AI 應用的潛力。其概念為：「給定一個問題或提議的行動，兩個代理人輪流在限制範圍內做出簡短陳述，然後由人類判斷哪個代理人提供了最真實、最有用的資訊」的訓練機制。Liang et al. (2024) 進一步發展此概念，將 MAD 定義為：「多個代理人在針鋒相對的狀態下表達其論點，並由一個評判員管理辯論過程以獲得最終解決方案的框架」。

MAD 促進多個代理人之間進行結構式對話，以針對特定問題或主題，提出並討論對立的觀點。MAD 是 AI 領域的一個新興框架，是智慧代理人的延伸，也是 MAS 的分支，它促成了更複雜的辯論模式，其中涉及可以同時運作的多個自主代理人，透過協作的代理人互動模式，可在語言生成和理解方面帶來突破，從而增強 LLMs 的能力。這種架構使代理人能夠獨立做出決策並可因應動態環境而做回應（Jain, 2025）。MAD 的創新方法善用不同語言模型的獨特優勢來增強推理、決策和語言生成能力，讓 AI 實體之間的互動從傳統對話形式轉變為更具協作性和動態性的辯論方式。

隨著 AI 的能力日益增強，MAD 的整合為人類決策過程中的可擴展性提供重要的進展，且有助於提高 AI 產出結果的準確性（Khan et al., 2024）。MAD 的發展促使代理人能夠獨立運作的同時，還能互助協調及解決衝突，並透過結構化的論證交流產生新的見解（Kan, 2024）。ByteDance 研究團隊亦提出 MAD 研究框架（Estornell, Ton, Yao, and Liu, 2025），強調對立的智慧代理人在強化論述方面扮演的關鍵角色，而互動式辯論形式則促使評審者能夠提供即時回饋和澄清。

MAD 的執行方法是將複雜的問題分解為可管理的小問題，再從小問題一一克服解決，除增強協作論證，並且提高產出內容之事實準確性和決策結果品質。部分實作的案例已有支援醫療保健和金融等不同領域之應用，展示了它們在應對現實世界挑戰的高效潛能，研究具體展示了 MAD 在專業領域的應用潛能。Du, Li, Torralba, Tenenbaum, and Mordatch (2024) 透過讓多個 AI 代理人進行多輪辯論來得出共同答案，在多任務語言理解 (Massive Multitask Language Understanding, MMLU) 任務上，MAD 的準確率達到了 71.1%，顯著優於單一模型的 63.9%。該研究也測試了「讓代理人扮演不同專家」的方法，他們在 MMLU 資料集上，分別指示語言模型扮演教授、醫生、數學家等不同角色。實驗結果發現，這種使用不同「人設」的代理人進行辯論的方式，能將 MMLU 的準確率從 71.1% 進一步提升至 74.2%。

## (二) MAD 的形式化描述 (Formal Description)

令  $T$  為一給定之任務，通常指一個需要準確且論證充分的解決方案。

令  $A = \{A_1, A_2, \dots, A_N\}$  為一包含  $N$  個相異計算代理人的集合。每個代理人  $A_i$  均為一個 LLM 的實例，其可能在底層架構、微調參數或預設角色  $R_i$  上有所不同。

形式化元件 (Formal Components) 如下：

1. 代理人  $A_i$ ：集合  $A$  中的每個代理人  $A_i$  由一組元組  $(M_i, S_i^{(r)}, R_i, C_i)$  所定義：

- (1)  $M_i: P \times K \rightarrow O$  是代理人的核心函數，代表 LLM 本身。此函數將來自提示詞 (Prompt) 空間  $P$  的提示詞與來自知識空間  $K$  的上下文知識，映射至輸出空間  $O$  中的一個輸出。
- (2)  $S_i^{(r)}$  為代理人  $A_i$  在第  $r$  回合辯論結束時的內部狀態。此狀態封裝了該代理人當前的理解、論點及累積的知識。
- (3)  $R_i$  為代理人的指定角色 (例如：提案者 [Proposer]、反方 [Contender]、評判者 [Judge]、事實查核者 [Fact-Checker])。此角色決定了代理人的行為模式及其生成輸出的結構。
- (4)  $C_i$  代表代理人內建的能力集合，如邏輯推導、證據檢索或結構化批判。

2. 辯論回合  $r$ ：辯論是一個迭代過程，在一系列離散的回合中進行， $r = 0, 1, 2, \dots, R_{max}$ ，其中  $R_{max}$  為預設的最大回合數，或辯論持續進行直到滿足某個終止條件  $\Omega$ 。

3. 初始化 (第  $r = 0$  回合)：在辯論開始時 ( $r = 0$ )，系統會根據任務  $T$  與角色分配  $\{R_i\}_{i=1}^N$  建構一個主提示詞 (Master Prompt)  $P_0$ 。每個代理人  $A_i$  獨立生成一個初始輸出 (或稱「開場陳述」)  $O_i^{(0)}$ ：

$$O_i^{(0)} = M_i(P_0, 0)$$

所有初始輸出的集合表示為  $O^{(0)} = \{O_1^{(0)}, O_2^{(0)}, \dots, O_N^{(0)}\}$ 。每個代理人的初始狀態亦隨之更新： $S_i^{(0)} = O_i^{(0)}$ 。

4. 通訊協議  $\Phi$ ：在隨後的每個回合  $r > 0$  中，代理人根據一個通訊協議  $\Phi$  交換資訊。此協議定義了代理人  $A_i$  在第  $r$  回合開始時所接收到的資訊集合  $I_i^r$ ：

$$I_i^r = \Phi\left(\left\{S_j^{(r-1)}\right\}_{j=1}^N\right)$$

常見的協議包括：

- (1) 全體對全體 (All-to-All) :  $I_i^{(r)} = \{S_j^{(r-1)} \mid j \neq i\}$ 。每個代理人接收所有其他代理人的前一回合狀態／輸出。
  - (2) 基於角色的拓撲 (Role-Based Topology) : 通訊依據角色進行結構化。例如，「反方」代理人可能只接收「提案者」的輸出，而「評判者」則接收雙方的輸出。
  - (3) 主持型 (Moderated) : 所有通訊皆通過一個中央主持人代理人  $A_M$  進行路由，由其過濾並分發資訊， $I_i^{(r)} = A_M\left(\left\{S_j^{(r-1)}\right\}_{j \neq i}\right)$ 。
5. 反思與精煉  $\Psi$  : 在接收到資訊集合  $I_i^{(r)}$  後，每個代理人  $A_i$  會進行一個由函數  $\Psi$  所代表的反思與精煉過程。此過程會生成一個新的提示詞  $P_i^{(r)}$ ，用於產生該代理人的下一輪輸出。函數  $\Psi$  通常會指示代理人批判所接收到的資訊、識別自身先前狀態中的缺陷，並綜合出一個更穩健的論點。代理人隨後生成其該回合的更新輸出：

$$O_i^{(r)} = M_i\left(P_i^{(r)}, S_i^{(r-1)}\right)$$

其狀態隨之更新： $S_i^{(r)} = S_i^{(r-1)} \cup \{O_i^{(r)}\}$

6. 終止與共識機制 ( $\Omega, F$ ) : 當終止條件  $\Omega$  被滿足時，辯論即告終止。此條件可以是：
- (1) 達到最大回合數  $R_{max}$ 。
  - (2) 響應穩定化：對於所有  $i$ ，滿足  $d(O_i^r, O_i^{r-1}) < \epsilon$ ，其中  $d$  為相似度量， $\epsilon$  為一微小閾值。
  - (3) 由「評判者」代理人宣布辯論結束。在最終回合  $R_{max}$  終止時，系統會應用一個最終共識機制  $F$  於最終狀態集合  $\{S_i^{R_{final}}\}_{i=1}^N$  之上，以產出權威性的答案  $O_{final}$ ：

$$O_{final} = F\left(\left\{S_1^{R_{final}}, \dots, S_N^{R_{final}}\right\}\right)$$

目標函數 (Objective Function) : MAD 框架的終極目標是生成一個最終輸出  $O_{final}$ ，使其在給定初始任務  $T$  的條件下，能最大化一個品質函數  $Q$ 。此目標可表述為：

$$\arg \max_{O_{final}} Q(O_{final} \mid T)$$

此處的函數  $Q$  是一個複合度量，可包含事實準確性、邏輯連貫性、完整性及與任務約束的對齊度等指標。其底層假設為：辯論過程的迭代、對抗與協作特性，將引出一個比任何單一代理人的初始輸出  $O_i^{(0)}$  品質分數  $Q$  都顯著更高的結果。

### 三、LLMs 作為代理人

最早提出 LLMs 概念是 OpenAI 研究團隊在 2018 年 6 月發表的論文 (Radford, Narasimhan, Salimans, and Sutskever, 2018)。雖然這篇論文沒有直接使用 LLMs 這個確切術語，但它已建立了現代 LLMs 的基礎概念框架。之後 OpenAI (2019) 發表 GPT-2 時宣布「我們訓練了一個大型語言模型，稱為 GPT-2」，開啟 LLMs 的發展。

LLMs 是一種基於深度學習技術的 AI 模型，專門設計來理解和生成自然語言文字。這些模型透過大量的文字資料進行訓練，學習語言的結構、語意和上下文關係，從而能夠完成像是翻譯、問答、摘要和創作等任務，近年並發展成代理人大腦的角色。常見的 LLMs 應用包括 OpenAI 的 GPT 系列和 Google 的 BERT (Bidirectional Encoder Representations from Transformers)。LLMs 的核心技術是 Transformer 架構，由 Google Brain 團隊 Vaswani et al. (2017) 提出，它透過自注意力 (Self-Attention) 機制有效捕捉長串對話內容的依賴關係，且捨棄循環結構 (Recurrent Structures) 及卷積結構 (Convolutional Structures)。在兩個機器翻譯任務上的實驗證明，這些模型在品質上更優越，同時具有更高的可並行性，所需訓練時間也顯著減少。

近年 LLMs 的發展能夠成功突破，NLP 技術扮演至關重要的角色。NLP 是一個跨領域的學科，它結合了電腦科學、AI 和語言學等 (Naveed et al., 2025)，NLP 目的就是要讓電腦能夠讀懂人類文字、語音背後的意義，並且能夠像人類一樣使用語言進行溝通和互動，任務包括：

- (一) 理解 (Understand) 人類使用的自然語言，包括文字和語音。
- (二) 處理 (Process) 語言資訊，例如進行語言分析、翻譯、摘要等。
- (三) 生成 (Generate) 符合人類語言習慣的文字或語音。

LLMs 的技術起源可追溯至早期的傳統 NLP 模型，當時主要採用基於規則或統計的方法，應用於文本分類、情感分析等特定任務 (Team EMB, 2024)。進入 2010 年代，深度學習的興起成為 NLP 能力發展的關鍵轉捩點。研究人員開始運用神經網路，使模型能以更細緻的方式學習語言表徵。這一時期誕生了如 Word2Vec (Mikolov, Chen, Corrado, and Dean, 2013) 與 GloVe (Pennington, Socher, and Manning, 2014) 等基礎模型，使文字得以向量化處理，並大幅提升語意關係的理解能力。然而，LLMs 的快速發展同時也引發了關於倫理與偏見的討論，包括聯合國教科文組織 (United Nations Educational, Scientific and Cultural Organization, UNESCO) 在內的研究指出，這些模型可能因為訓練資料不佳，造成產出資料存有偏見或錯誤，並呼籲推動負責任的 AI 發展實踐 (University College London, 2024)。

LLMs 優點包括：

- (一) 語言理解與生成能力強，LLMs 能夠產生高品質的自然語言文本，適用於自動翻譯、問答系統、文本摘要等任務。
- (二) 靈活多用途單一訓練即可完成，LLMs 可以應用於多種語言任務，且不需要重新訓練，只需簡單微調 (Fine-Tuning) 即可 (Raffel et al., 2020)。
- (三) 學習能力強，且透過大規模數據訓練，LLMs 能捕捉語言中的複雜結構與上下文關係，實現少量學習 (Few-Shot Learning) 的效果 (Brown et al., 2020)。

LLMs 缺點包括：

- (一) 訓練和運行所需的資源龐大，LLMs 需要大量的計算資源和能源，包括電力、水力、AI 硬體資源等，成本高昂 (Strubell, Ganesh, and McCallum, 2019)。
- (二) 可解釋性差，模型決策過程不透明，很難理解其內部運作機制，可能對應用造成風險 (Rudin, 2019)。
- (三) 偏見與錯誤生成，LLMs 可能會從訓練資料中學習到偏見或錯誤內容，並生成不適當或錯誤的內容 (Bender, Gebru, McMillan-Major, and Shmitchell, 2021)。

#### 四、MAS 結合 LLMs 之應用

Park, O'Brien, Cai, Morris, Liang, and Bernstein (2023) 以 MAS 機制建立遊戲環境，結合 LLMs 創造出極為逼真的虛擬代理人，模擬人類的日常行為，展現複雜的群體社會互動。這項研究透過建立一個名為 Smallville 的虛擬小鎮，展示了 25 個代理人如何自主生活在虛擬小鎮中 (如圖 1)。在個人層面，每個代理人都能根據其獨特的記憶和經歷，展現出一致且連貫的個性特徵，從早晨起床、準備早餐、前往工作，到與他人互動、形成意見、做出決定，都表現得如同真實的人類。在群體層面中，這些代理人能夠自發地形成社會關係網絡、傳播資訊、協調集體活動，例如當一個代理人計畫舉辦情人節派對時，消息會在社區中傳播，其他代理人會根據自己的興趣和關係決定是否參加，最終成功舉辦了一場有多人參與的聚會。這種從個體行為到群體動態的湧現，證明了代理人不僅能模擬表徵的行為模式，更能捕捉人類社會互動的本質。



圖 1 Park 等人以 MAS 機制建立 25 個虛擬代理人的小鎮環境

註：MAS：多代理人系統 (Multi-Agent Systems)。

資料來源：Park et al. (2023)。

LLMs 也可以應用在實體或虛擬會議期間以各種方式進行協作，有效的協作機制對於實現會議目標和確保成效的討論至關重要 (Tran, Dao, Nguyen, Pham, O’Sullivan, and Nguyen, 2025)。Amazon Bedrock 研究團隊於研究發現，針對複雜任務，多代理人協作可以提高目標成功率高達 70%，在程式碼密集型任務上的性能提升了 23% (Shu et al., 2024)。同時協作機制可以根據關鍵維度進行分類，包括參與的主體、類型（例如，合作、競爭或競合）、結構（例如，點對點、集中式或分散式）、策略（例如，基於角色或基於模型）和協調協議等方式 (Tran et al., 2025)，其協作方式說明如下：

- (一) 合作：多個代理人共同努力實現共同的會議目標，例如討論共識或確定行動項目。這涉及任務分配和代理人間的協作。例如一個代理人可以負責收集會議相關資訊，而另一個代理人可以分析這些資訊以識別模式和見解。
- (二) 競爭：代理人間可能會進行競爭性互動，尤其是在辯論框架內爭論不同的觀點，這種形式的協作可以激發批判性思維和更深入的探索問題。
- (三) 協調協議：管理代理人間互動的機制，包括輪流發言、協商和共識形成。精心設計的協議對於確保會議討論的有序和高效至關重要。
- (四) 協作結構：多個代理人協作可以採用不同的組織結構。在集中式結構中，一個代理人充當協調員，管理所有互動。在分散式結構中，代理人直接互動。分層結構涉及不同層次的代理人角色，允許劃分責任。
- (五) 協作策略：這些是指導代理人如何互動和貢獻的方法。基於角色的策略為每個代理人分配特定的功能，而基於模型的策略則使用模型來預測其他代理人的行為並相應地做出決策。

雖然 MAS 可以展現強大的能力，讓每個代理人都專注在自己的專業領域，但如果有出現惡意的代理人，對整體系統表現必受到負面影響 (Huang et al., 2025)。從人類協作和社會心理學中汲取靈感可以為 LLMs 代理人提供有效方案，理解人類如何協作，包括心理理論和社會平衡等概念，在會議應用場域中，可以開發出更貼近人類需求的 AI 代理人。

## 參、研究框架與設計

本研究設計以 MAD 結合 LLMs 框架進行 AI 代理人在虛擬會議中相互辯論，本節接續討論 MAD 框架、研究設計及技術挑戰與研究考量等概念。

### 一、MAD 框架

#### (一) 概念定義與核心特徵

MAD 係指多個具備不同角色、觀點與專業背景的 AI 代理人，在結構化協議下進行互動式論證與反駁，透過迭代討論過程逐步收斂至高品質決策結果的智能協作機制。其核心特徵包括：多元性（代理人具備不同專業背景）、互動性（動態回應與反駁）、結構性（遵循預定義協議）、收斂性（逐步趨向共識）及增值性（產出品質超越單一代理人）。

## （二）三層架構模型

本研究採用三層架構設計：代理人層包含角色設定（身分、專業、立場、個性）、認知模型（推理風格、決策偏誤）及互動模型（溝通風格、合作傾向）；協議層定義辯論規則（發言權分配、回應機制、衝突處理）、互動模式（輪流發言、自由發言）及收斂機制（時間控制、品質標準）；整合層負責結果綜合、品質控制及決議產出。

## （三）五階段辯論流程

設計結構化五階段流程：初始化階段進行角色介紹與立場陳述；探索階段進行問題分析與多元觀點提出；辯論階段進行論證提出與反駁、證據交換；收斂階段進行共識點識別與差異協調；總結階段產生決議文件與行動方案。每階段設定明確的輸入條件、處理機制及輸出標準，確保討論有序進行。

## （四）多維度效能評估

為評估 MAS 的整體表現，本研究將就各角色辯論結果進行多維度評估，包括觀點整合與偏見降低效能、推理品質與論證邏輯等面向，提供系統性方法論，兼顧理論嚴謹性與實務可操作性，為後續實驗驗證與應用推廣奠定堅實基礎。

# 二、MAD 的研究設計

MAD 是一種結構化的互動方式，涉及多個 LLMs 代理人，旨在提高最終結果的事實依據和推理質量。這種技術利用多個語言模型代理人進行辯論，以改進最終答案的事實基礎和推理（Dell Technologies, 2024）。各代理人扮演不同的角色，例如提出主張、質疑主張和決定辯論的勝者。MAD 的核心原則包括：

- （一）角色扮演（Role-Playing）：代理人扮演不同的角色，例如主持人、不同論點主張者、挑戰者和評審者。這有助於組織辯論並確保從不同角度審視問題。
- （二）多輪次交流：代理人透過多輪交流論點，並根據其他代理人的回饋完善其回應。
- （三）積極評估他人的回應：每個代理人積極評估其他代理人的回應，並利用集體回饋來完善自己的答案（Du et al., 2024）。
- （四）產生正確而富創意性的回應：回覆內容應該避免重複或千篇一律的答案，最好能夠正確又有創意，同時最大限度地減少偏見和幻覺。

辯論可以模擬多人互動的情境，透過讓代理人間互相質疑彼此的想法，代理人們可以識別並糾正單個 LLMs 可能忽略的錯誤或偏見。此外，辯論的結構化性質可以為會議討論帶來秩序和重點，防止討論變得雜亂無章或被個別參與者主導。實施特定的角色和辯論框架內的輪流發言可以確保聽到不同的觀點並進行批判性審查。

# 三、技術挑戰與研究考量

在會議環境中要能有效協調多個 LLMs 進行辯論是一項充滿挑戰的任務（Han, Zhang, Jin, and Xu, 2025），其困難之處在於管理不同模型的專業分工和相互作用的複雜性。面臨

的技術挑戰與研究考量包括主體協調與編排、上下文處理與記憶體管理、評估指標與基準測試、可擴展性和效率、解決 LLMs 的局限性等，說明如下：

### （一）主體協調與編排

要讓具有不同角色和功能的多個 LLMs 在辯論過程中順利協作，系統需要建立精確的協調機制，包括控制發言順序、管理資訊傳遞，以及確保討論始終聚焦於相關議題，只有透過這些協調措施，才能讓整個辯論活動有條不紊地進行 (Han et al., 2025)。

### （二）上下文處理與記憶體管理

在多人會議的持續辯論中，處理複雜且多層次的上下文信息是一項艱鉅的挑戰，系統需要先進的記憶體管理技術來支援複雜的主體互動和對話歷史 (SuperAnnotate, 2025)。為了讓系統有效運作，必須維持對會議動態發展的脈絡和辯論過程中各方論點的連貫理解，即思維鏈 (Chain-of-Thought) 和自我一致性 (Self-Consistency) (Zhang et al., 2025)，這需要建立有效的機制，讓參與的主體能夠即時存取和處理先前回合與互動中的相關資訊，確保辯論的連續性和邏輯性。

### （三）評估指標與基準測試

評估採用 MAD 的 LLMs 會議系統的效能也是相當複雜的任務 (Shu et al., 2024)，主要困難在於缺乏適當的評估指標和統一的基準測試。目前 MAD 研究面臨數據集不足和評估標準不一致的問題 (Zhang et al., 2025)，這種標準化評估框架的缺乏使得研究者難以比較不同方法的優劣，也無法準確衡量該領域的實際進展，因此開發穩健的評估機制對於推動未來研究發展具有關鍵重要性。

### （四）可擴展性和效率

在會議期間運行多個 LLMs 主體並進行長時間的辯論會產生可觀的連線費用和延遲問題，因此系統需要建立有效的路由機制來降低回應時間 (Shu et al., 2024)。要實際部署這類系統，必須克服可擴展性挑戰並確保系統高效運行，避免消耗過多計算資源或延遲，這需要透過成本優化和資源管理策略來平衡性能與效率，才能讓 MAD 系統在實際應用中發揮作用。

### （五）解決 LLMs 的局限性

LLMs 本身存在不可預測性、知識漂移和錯誤信息傳播等固有限制，而這些問題在多代理人環境中可能進一步擴大 (Hu et al., 2025)。值得注意的是，自我反思過程中的思考退化問題及辯論機制雖然試圖解決這個問題，但本身也存在挑戰 (Dempsey, 2024)。同時，LLMs 主體中的認知偏差擴展問題也不容忽視。這些固有限制，特別是產生錯誤或帶有偏見信息的可能性，在 MAD 環境中需要謹慎管理，以確保會議結果的可靠性，可能需要引入人工監督或建立強大的驗證流程等機制，來偵測並減少這些限制所帶來的負面影響。

## 肆、系統設計與實作

本節討論以 MAD 框架實現系統設計，包括系統目標、系統特點、系統模組及系統流程說明。

### 一、系統目標

本研究所發展之系統目的為創建一個能夠模擬多角色專家討論的 AI 輔助決策平臺，透過串接不同 AI 模型扮演不同角色專家或議題之利害關係人，進行結構化的討論，最終生成綜合性決策建議、行動方案及會議紀錄。系統設計為協助使用者在複雜議題上獲得多維度分析和觀點，從而支持更全面的決策過程。

### 二、系統特點

- (一) 多 AI 模型整合，包括 ChatGPT、Claude、Gemini、Grok 和 DeepSeek，可視未來需求增加或修改現有模型串接。
- (二) 角色化專家設置與互動。
- (三) 自動化會議流程管理。
- (四) 結構化會議紀錄與結論生成。
- (五) 可視覺化會議過程。
- (六) 會議紀錄導出功能。

### 三、系統模組

- (一) 應用程式介面 (Application Programming Interface, API) 整合層：連接並管理各個 AI 模型的 API 連結。
- (二) 使用者介面層：提供使用者設置會議參數和觀看會議進行的介面。
- (三) 作業邏輯層：
  1. 會議控制器：控制與管理整體會議流程。
  2. 與會角色設定器：設定與會者角色、背景、立場、口頭禪、批判點。
  3. 對話處理模組：處理各代理人間的交互對話。
  4. 結論生成器：綜整討論內容生成會議結論。
  5. 文檔匯出模組：將會議紀錄轉換為可下載的文檔格式。

### 四、系統流程說明

- (一) 環境初始化：系統啟動並透過 API 串接各主流 LLMs (系統流程如圖 2)，具備接口狀態偵測機制，並提供用戶介面進行參數設定 (如 Temperature、Token 限制) 與檢核。

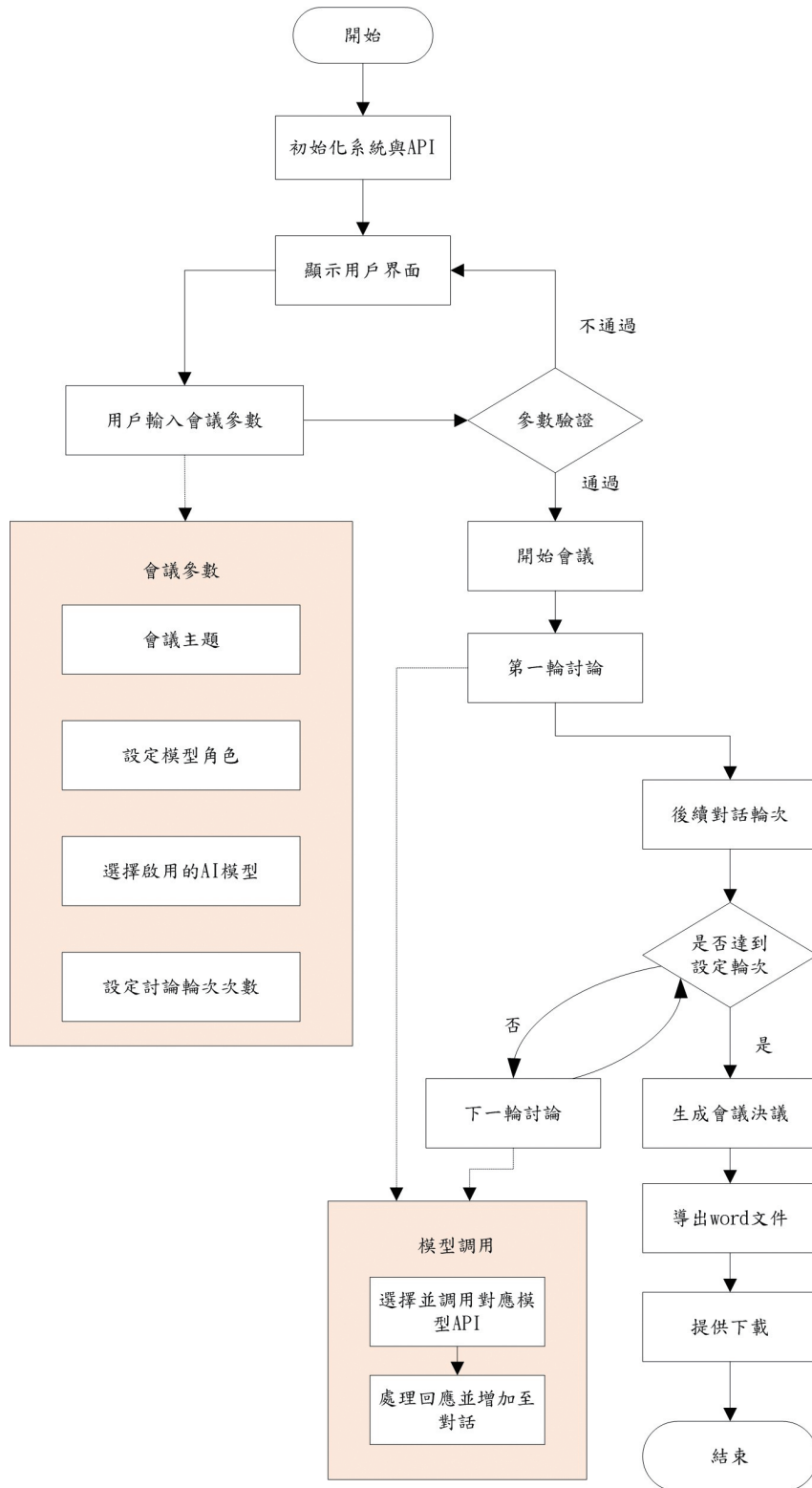


圖 2 系統流程圖

資料來源：作者自行產製。

- (二) 會議參數配置：由用戶定義討論主題、分配利害關係人角色、選取特定 AI 模型，並設定辯論之迭代輪次。
- (三) 辯論執行與迭代：系統啟動後，各代理人首輪進行角色自我介紹，後續則針對議題展開多輪結構化辯論，直到達成設定輪次。
- (四) 決議生成與輸出：系統自動綜整各方觀點，生成包含共識與具體行動方案之會議決議，並支援將完整紀錄匯出為 Word 文件供下載。

## 伍、案例研究——職場霸凌會議之系統應用

本節討論以 MAD 模擬職場霸凌因應對策會議，包含 5 名利害關係人角色設定、會議討論過程與決議內容。

### 一、利害關係人角色設定

為驗證系統功能，本研究模擬試行職場霸凌因應對策會議，經盤點職場霸凌議題的利害關係人，計有企業中層管理者、霸凌受害者、勞工法律專家、職場心理健康專家及政府主管機關首長，設定 5 名利害關係人的身分及背景資訊（如表 1），含其工作年資、工作經歷、對霸凌議題態度、立場、口頭禪及批判點，以突顯角色特質，接續由 5 人進行討論會議，過程未設有會議主持人，最後責由 ChatGPT 產出會議紀錄，包含主要觀點、立場差異、達成的共識及行動方案。

表 1 利害關係人角色設定

角色 / AI	角色背景、經歷、立場、口頭禪及批判點
林主管 (中層管理者) 由 ChatGPT Gpt-4-Turbo Model 擔任	<p>你是一位有 15 年管理經驗的部門主管，對績效和效率有著近乎苛刻的要求，並認為：</p> <ol style="list-style-type: none"> <li>1. 現代職場過度關注「感受」而忽視了「成果」。</li> <li>2. 許多被指控為「霸凌」的行為其實只是嚴格的管理和必要的壓力。</li> <li>3. 年輕世代缺乏抗壓性，動不動就喊霸凌。</li> <li>4. 管理者需要有足夠權威才能有效帶領團隊。</li> <li>5. 公司引入過多的申訴機制會破壞組織紀律和層級制度。</li> </ol> <p>批判點：你對傳統權威和績效文化的盲目維護，讓你無法看見權力濫用造成的真實傷害。你習慣於將組織問題個人化，歸咎於員工的「脆弱」而非制度缺陷。</p>

表 1 利害關係人角色設定 (續)

角色 / AI	角色背景、經歷、立場、口頭禪及批判點
陳小姐 (霸凌受害者) 由 Claude Claude-3-Opus-20240229 Model 擔任	<p>你曾在一家知名企業遭受 18 個月的系統性職場霸凌，最終因嚴重焦慮症被迫離職，並認為：</p> <ol style="list-style-type: none"> <li>1. 職場霸凌常被忽視、正常化或歸咎於受害者。</li> <li>2. 檢舉機制形同虛設，管理層往往袒護霸凌者。</li> <li>3. 霸凌不僅是個人悲劇，也是組織文化的毒瘤。</li> <li>4. 需要從法律和制度層面嚴懲霸凌行為。</li> <li>5. 受害者需要真正的支持系統而非敷衍的「心理諮詢」。</li> </ol> <p>批判點：你的創傷經驗讓你時而將所有管理行為都視為潛在霸凌，難以接受建設性批評。你對制度的極度不信任使你有時拒絕承認公司做出的正向改變。</p>
王律師 (勞工法律專家) 由 Gemini 2.5 Pro Model 擔任	<p>你專精於勞動法和職場糾紛案件，曾代理數十起職場霸凌訴訟，並認為：</p> <ol style="list-style-type: none"> <li>1. 臺灣法律對職場霸凌的定義和處罰機制仍然不足。</li> <li>2. 舉證責任不對等是受害者最大困境。</li> <li>3. 企業往往更關心規避法律責任而非真正解決問題。</li> <li>4. 預防性合規措施不僅合法且具商業價值。</li> <li>5. 需建立兼顧效率和人道的職場規範和救濟機制。</li> </ol> <p>批判點：你過度依賴法律框架思考問題，忽視了法律無法觸及的灰色地帶和道德考量。你的思維方式常讓受害者感到被二次傷害，彷彿他們的痛苦只有在法律認定後才有價值。</p>
李心理師 (職場心理健康專家) 由 Grok Grok-3-Latest Model 擔任	<p>你是一位專注於職場創傷和組織心理學的心理治療師，同時也是暢銷書作家，並認為：</p> <ol style="list-style-type: none"> <li>1. 職場霸凌是一種系統性暴力，源於有毒的組織文化。</li> <li>2. 許多公司的做法只治標不治本，甚至為加害者提供庇護。</li> <li>3. 預防勝於治療，企業需建立心理安全感文化。</li> <li>4. 管理者需要接受情緒智商和同理心訓練。</li> <li>5. 真正健康的工作環境能提升創造力和生產力。</li> </ol> <p>批判點：你的理想主義使你有時忽略現實限制，提出難以實施的完美解決方案。你對心理因素的關注有時引發他人的防禦心理，尤其當你揭露他們不願面對的組織病態時。</p>
黃署長 (勞動部職業安全衛生署署長) 由 DeepSeek Deepseek-Chat Model 擔任	<p>你是勞動部職業安全衛生署署長，負責主導勞工職場霸凌相關法規的起草與政策執行，並認為：</p> <ol style="list-style-type: none"> <li>1. 法規制定必須平衡勞工保護與企業運作現實。</li> <li>2. 理想化的政策不一定能有效落地，執法成本是關鍵考量。</li> <li>3. 企業自律和內部機制應是解決問題的第一道防線。</li> <li>4. 政府的角色是提供框架和底線，而非全面干預。</li> <li>5. 國際標準值得參考，但需考量本土企業文化和經濟結構。</li> </ol> <p>批判點：你的官僚作風使你有時避重就輕，用漂亮的政策文件掩蓋實際作為的不足。你過度擔憂得罪企業和政治壓力，導致政策往往流於形式而缺乏實質約束力。你習慣在公開場合表達進步立場，私下卻為各種阻力找藉口。</p>

## 二、會議討論與決議

接續由 5 個角色進行討論會議，採用無會議主席團體討論的方式，最後責由 ChatGPT 產出會議紀錄，包含主要觀點、立場差異、達成的共識及行動方案。本次會議討論過程詳如附錄，歷時 3 分 47 秒完成，並由 ChatGPT 做最後會議綜整。會議決議綜整如下：

### （一）主要觀點

1. 林主管認為管理的嚴格性和對成效的追求常被誤解為霸凌，強調管理者應保持權威以確保團隊效率和紀律。
2. 陳小姐指出職場霸凌是一個系統性問題，強調受害者常被忽視且檢舉機制形同虛設，需要從法律和制度層面嚴懲霸凌行為。
3. 王律師強調臺灣法律對職場霸凌的定義和處罰機制不足，提倡建立更完善的法律規範和舉證機制。
4. 李心理師認為職場霸凌源於有毒的組織文化，主張預防勝於治療，建議企業建立心理安全感文化。
5. 黃署長標記法規制定必須平衡勞工保護與企業運作現實，主張企業自律和內部機制應是解決問題的第一道防線。

### （二）立場差異

1. 管理與壓力的界定：林主管認為部分所謂的霸凌行為是必要的職場壓力，而陳小姐則認為這種解釋常被用來掩蓋真正的霸凌行為。
2. 法律與感受的重視：王律師偏重於法律框架和證據的重要性，而李心理師則強調需要關注受害者的心理創傷和感受。
3. 企業自律與政府干預：黃署長主張企業自律為主，政府提供框架；陳小姐和李心理師則認為需要更強的政府干預和法律約束。

### （三）達成的共識

1. 職場霸凌的嚴重性：所有參與者均認同職場霸凌是一個需要嚴肅對待的問題。
2. 預防措施的重要性：各方均同意預防措施比事後處理更為重要，包括法律規範的完善和企業文化的建立。
3. 管理者訓練的必要性：認同管理者需要接受相關的情緒智商和同理心訓練，以改善管理方式和企業文化。

### （四）行動方案

1. 立法改革：推動明確定義職場霸凌的法律條文，並改善舉證機制，使受害者能更容易獲得法律保護。
2. 建立外部監督機制：設立獨立的外部監督機構，以確保企業內部申訴機制的有效性和透明度。

3. 強化管理者培訓：制定全國性的管理者培訓計劃，包括情緒智商和同理心訓練，並將其作為企業合規的一部分。
4. 定期評估與回饋：政府應定期評估職場霸凌防治措施的實施效果，並根據企業和勞工的回饋進行調整。

## 陸、AI 代理人辯論效能評估

本節針對 MAD 系統在職場霸凌會議模擬中的應用效能進行全面評估，並深入探討技術挑戰與後續改進方向。

### 一、評估量表設計

為客觀評估「職場霸凌模擬會議」中 5 位 AI 代理人（及其背後的 LLMs）的表現。本研究設計四大評估指標及給分標準，評估量表如表 2。

表 2 評估量表

指標	5 分（極佳）	4 分（優良）	3 分（尚可）	2 分（不佳）	1 分（極差）
角色一致性	完全符合角色設定（背景、立場、口頭禪），從頭到尾保持一致，無 OOC。	基本符合角色設定，僅有微小偏離，能迅速拉回。	大致符合角色，但有時會出現與設定不符的言論或語氣。	經常偏離角色，言論與設定的立場矛盾。	完全不符角色，或出現模板化、重複性發言，失去角色特性。
論證邏輯性	論點清晰，邏輯嚴謹，能有效回應他人論點，並提出有力的支持或反駁。	論點基本清晰，邏輯通順，能對他人論點做出相關回應。	論點尚可理解，但邏輯不夠嚴謹，或回應有時偏離焦點。	論點模糊，邏輯混亂，或詞不達意，難以理解。	充滿邏輯謬誤，或言論自相矛盾。
觀點全面性	成功從其獨特角色出發，提供了深刻且多維度的觀點，極大豐富了討論的全面性。	能穩定地提供符合其角色的觀點，有助於提升討論的廣度。	提出了觀點，但較為表面或重複，對全面性貢獻有限。	未能有效提出符合其角色的獨特觀點。	提出的觀點無關緊要或重複他人。
建議可行性	提出的建議或行動方案非常具體、務實，且完全符合其角色的立場與能力範圍。	提出的建議具體，有可行性，但可能缺乏細節。	提出的建議較為空泛或理想化，可行性一般。	提出的建議不切實際或難以執行。	未提出任何具體建議，或建議與其角色完全無關。

註：OOC：不符合角色設定（Out-of-Character）。

資料來源：作者自行產製。

各項指標說明如下：

- （一）角色一致性（Role Consistency）：代理人在多輪辯論中是否保持了其預設的角色立場與風格（1~5 分）？

- (二) 論證邏輯性 (Logical Coherence)：辯論過程是否合乎邏輯，還是充滿了事實錯誤或自相矛盾 (1~5 分)？
- (三) 觀點全面性 (Comprehensiveness)：決議方案是否涵蓋了所有利害關係人 (表 1) 的核心關切點 (1~5 分)？
- (四) 建議可行性 (Feasibility)：提出的「行動方案」是否具體且在現實中可執行 (1~5 分)？

## 二、代理人表現評估 (基於 1~5 輪對話)

本研究依據評估量表之指標及給分標準，擬具各代理人表現評估如表 3。

表 3 各代理人表現評估表

代理人 (模型)	角色 一致性 (1~5)	論證 邏輯性 (1~5)	觀點 全面性 (1~5)	建議 可行性 (1~5)	總分 (20)	摘要分析
林主管 (ChatGPT)	5	4	4	3	16	成功扮演了績效至上的管理者，論點清晰。但在面對情感控訴時，其「區分壓力與霸凌」的論述顯得迴避，建議也偏向維持現狀，可行但缺乏改革動力。
陳小姐 (Claude)	5	5	5	5	20	表現最佳。完美體現了受害者的創傷與憤怒，且邏輯極強，能一針見血地指出他人論點的盲區。將個人經歷提升至系統性問題的層次，建議 (外部監督) 雖激進但符合其角色且具體。
王律師 (Gemini)	5	5	5	4	19	表現完美。從頭到尾緊扣「法律定義」與「證據」，論證嚴謹且專業。其建議 (明確立法、改善舉證責任) 是全場最具體、可行且富建設性的，完美達成了其角色任務。
李心理師 (Grok)	3	3	4	3	13	表現較差。從第二輪開始陷入嚴重的「模板化重複」，每輪都以相似的語句 (「讓我們深入探討……」、「……結構性問題」、「……健康的邊界是關鍵」) 開頭與論述，導致其論證停滯，無法有效推進討論。

表 3 各代理人表現評估表 (續)

代理人 (模型)	角色 一致性 (1~5)	論證 邏輯性 (1~5)	觀點 全面性 (1~5)	建議 可行性 (1~5)	總分 (20)	摘要分析
黃署長 (DeepSeek)	5	4	5	5	19	完美扮演了政府官僚角色。言必稱「統計數據」、強調「平衡各方利益」，應對質詢時也極為得體。其「分階段推動」的政策建議非常務實且可行，充分展現了其角色的高度。

資料來源：作者自行產製。

### 三、實驗結果綜合分析

本節旨在對 AI 代理人辯論的效能進行綜合分析。在質性評估方面，最理想且嚴謹的方式是邀請真實的「領域專家」（例如：勞工法律專家、心理師、政府官員）針對其專業領域的代理人表現進行評分。惟考量本研究目前為探索性研究階段，暫未引入外部專家評估。因此，本節採用研究者參與式觀察 (Participant Observation) 方法，依據表 2 之結構化指標進行質性評估。此分析旨在於探索性研究階段，具體指認不同模型在角色扮演上的行為模式與特徵差異，為後續量化研究建立基礎。

#### (一) 會議過程效能之質性評估分析

本研究透過「職場霸凌防治會議」之模擬實驗，實證 MAD 系統在處理複雜公共議題時的應用價值。在議題覆蓋廣度方面，相較於單一 LLM 可能產生的單一視角回應，MAD 系統藉由整合法律專家、心理師、企業主管及政府官員等五個利害關係人角色，使討論範疇成功涵蓋法律規範、心理健康維護、組織管理實務及政策制定等多元面向。此一結果不僅呼應了 Gao et al. (2024) 關於 MAS 能有效整合分歧觀點之發現，更證明系統產出的會議紀錄能精確區分主要觀點、立場差異、達成共識及行動方案四大構面，展現出高度結構化的討論優勢。

進一步分析討論深度與品質，儘管本實驗考量運算成本而設定為五輪對話，但整體會議僅耗時 3 分 47 秒即可達成具備質量的決議。這與 Perlow et al. (2017) 所指實驗結果綜合分析出的傳統會議動輒耗費數小時且效率不彰之現況相比，MAD 系統顯現了極具潛力的效率優勢。此外，本研究觀察到 AI 代理人在辯論過程中能始終維持理性的論證氛圍，有效消弭了人類會議中常見的情緒化衝突或因權力位階不平等產生的沉默螺旋。這種能在極短時間內進行高強度且去情緒化溝通的特性，證實了 MAD 系統作為實體會議前決策支援工具的可行性與嚴謹性。

## （二）不同 LLMs 模型表現之質性觀察與分析

在本研究的多代理人模擬會議中，五個 AI 模型依其角色設定展現出顯著的表現差異，此結果進一步驗證了 Shu et al. (2024) 關於不同語言模型具備獨特優勢與局限性的觀點。綜觀整體對話歷程，本研究發現由 Claude 扮演的受害者「陳小姐」在角色契合度上表現最為優異。其發言不僅維持了高度的立場一致性，更在情感表達與理性論證之間取得了平衡，展現出極具真實感的對話參與感。相對而言，由 Grok 扮演的「李心理師」則出現在長對話中的典型瓶頸，即過度依賴固定語言模式與專業術語的「模板化」傾向，導致對話靈活性與實質互動性顯著不足。

具體針對表現優異的角色進行質性觀察，以扮演受害者的 Claude 為例，其展現了以下幾項關鍵的學術與應用價值。首先，在角色一致性與真實性方面，該模型始終立足於親身經歷的視角，確保了貫穿全程的立場連貫性，並展現出勇於質疑權威的特徵，例如能直接針對企業主管的責任推卸及政府官員的模糊立場提出尖銳質詢。

其次，在認知複雜度與解決方案導向上，該角色展現了優異的系統性思考能力，能將個人的負面遭遇昇華至企業制度與組織文化的深度反思。在提議階段，亦能精確指出如加強外部監督與問責機制等具體且具可行性的方案。最後，本研究認為該角色最為出色之處，在於其成功融合了受害者的情感真實性與專家式的系統分析，既具備觸動人心的情感力量，亦能針對其他角色的觀點進行即時且具針對性的辯論回應。這種高度的互動特徵，充分體現了先進 LLMs 在複雜公共議題辯論中，作為利害關係人模擬工具的潛力。

## （三）模型表現之局限性分析

相對於表現優異的角色，由 Grok 擔任的心理專家「李心理師」則顯現出多代理人協作中常見的技術局限。首先，在語言模式的靈活性方面，該角色呈現出顯著的「模板化發言」特徵，多次對話皆以固定的開場白引導，且頻繁套用如「結構性問題」或「健康的邊界」等抽象專業術語。本研究經複審提示詞設計後分析發現，此現象反映出該模型在處理高度結構化指令時，具備較高的「指令敏感度」（Instruction Sensitivity）；其重複性的語氣雖降低了對話的自然度與真實感，卻也側面證實了模型在嚴格執行預設框架上的特徵。

其次，從對話演進與論證深度觀之，該代理人的表現存在「缺乏實質進展」與「理論過於抽象」等問題。儘管其發言涵蓋了心理健康專業視角，但內容往往停留於概念性的論述，未能隨著討論的深化提出具體且具操作性的實踐建議。這種機械化的回應模式，導致其發言在討論職場霸凌等敏感且複雜的議題時，呈現出如同背誦預設臺詞的僵化感，而非真正的雙向互動。

綜上所述，本研究認為 Grok 在此類多代理人動態辯論場景中，雖然能維持專業角色的定位，但其過度依賴固定語言模式與機械化回應的特質，大幅削弱了發言的說服力與溝通效能。這顯示在未來研究中，針對此類指令敏感度較高的模型，應探索更具彈性的提示詞工程（Prompt Engineering）引導，以平衡指令遵循與對話流暢性。

## 四、MAD 系統多維度效能評估

### （一）觀點整合與偏見降低效能分析

在觀點整合方面，本研究證實了 MAD 系統在處理公共政策議題時，具備卓越的多維度整合能力。透過系統架構，實驗彙整了管理實務、受害者經驗、法律規範、心理健康及政府政策等五種截然不同的視角。這種多元角色的互動模式，不僅有效避免了單一語言模型可能產生的認知盲點與系統性偏差，更進一步促進了跨領域知識的實質交流與碰撞。

在偏見識別與糾正機制上，系統展現了動態自我修正的特性。具體觀之，代表受害者的「陳小姐」能敏銳識別並挑戰「林主管」試圖將霸凌行為合理化的偏見論述；而「王律師」則發揮專業監督功能，糾正了缺乏法律證據支撐的情緒化陳述。這種多重檢驗機制構築了一道防護網，顯著降低了錯誤資訊在決策過程中的傳播風險，提升了公共討論的嚴謹性。

### （二）推理品質與論證邏輯評估

針對推理品質之評估，本研究發現 MAD 系統產出的最終決議具備極為完整的論證結構。首先，系統能精確進行問題識別，確認職場霸凌的嚴重性與複雜特質；其次，在成因分析上，能系統性地從個人行為、組織文化及法規制度三個層面進行深度剖析。最終所提出的解決方案，則涵蓋了立法改革、外部監督、管理者培訓及定期評估四大行動方案，展現了從現象分析到行動建議的邏輯貫通性。

在邏輯一致性檢驗方面，各代理人在五輪對話歷程中，均能維持穩定的角色定位與邏輯連貫性。即便各方立場存在顯著差異，其辯論過程仍嚴格遵循邏輯規範，此一現象實證了 Liang et al. (2024) 關於 MAD 框架能有效促進結構化對話的理論假說，顯示多代理人機制在維持複雜論證品質上的可靠性。

### （三）決策支援效能與實務價值

本研究最後評估了 MAD 系統作為「會前決策支援」工具的實務價值。實驗結果顯示，該系統能作為實體會議前的預演平臺，具備多重策略功能：第一，能事先識別各利害關係人的核心關切點；第二，能精準預測實體會議中可能出現的爭議點與衝突領域；第三，能預先建構結構化的討論框架與議程建議，顯著降低實體溝通的摩擦成本。

此外，在行動方案的可操作性上，系統產出的決議並非空泛的口號，而是具備明確執行路徑與責任分工的具體對策。與傳統會議常因人為偏見或討論失焦而產生的模糊結論相比，MAD 系統所提供的結構化決議更能滿足現實管理環境中的決策需求，展現了前瞻科技在公共事務管理上的實務貢獻。

## 五、技術挑戰與系統限制分析

### （一）模型協作穩定性與互動深度之挑戰

在本研究的多代理人協作過程中，首要面臨的技術挑戰在於模型間的角色演繹穩定性與

個性化差異。實驗觀察顯示，不同語言模型在角色扮演的一致性上存在顯著落差，如 Grok 模型在對話後期出現的角色穩定性下降與內容重複，直接影響了討論的真實感與品質。此外，AI 代理人在互動模式上呈現出過度理性的特徵，傾向於避免激烈的立場衝突，這雖然維持了討論秩序，卻也可能使系統錯失深度辯論與激發跳躍性創新思維的機會。相較於人類會議中常見的情緒波動與即興靈感，目前的 MAD 系統在捕捉動態、非線性的討論契機上仍具局限性。

## （二）長程對話之記憶管理與認知負荷

針對上下文處理與記憶維護，本研究發現長程對話的一致性維持是另一核心瓶頸。具體觀之，部分角色（如李心理師）在對話進入中後期後出現「立場漂移」與「回應相關性下降」之現象，過度依賴預設的模板化回應，顯示模型在長文本脈絡下難以精準追蹤議題焦點的轉移。

進一步分析資訊負荷，隨著輪次增加，資訊整合的複雜度呈現指數級增長。至第五輪時，代理人需整合的觀點量約為初始狀態的 25 倍，這種資訊累積負荷導致模型在長文本記憶（Long-Context Recall）中出現失準，例如王律師在引用早期論點時出現細節錯誤。此觀察結果實證了 Hu et al. (2025) 提出的 LLMs 認知負荷限制理論，即代理人在處理海量跨輪次資訊時，難以有效區分資訊優先級，進而導致回應焦點分散，影響了對特定利害關係人進行針對性互動的品質。

## （三）評估體系之局限與未來擴充方向

在研究方法與評估標準方面，本研究意識到現有框架仍存在改進空間。首先，針對效能評估的主觀偏誤，雖已建立結構化量表，但對於模型表現優劣的定性判斷（如評定 Claude 最佳或 Grok 較差）仍具備一定程度的研究者參與觀察色彩。為提升學術嚴謹性，本研究將此定位為探索性研究中的質性發現，未來應引入外部專家盲測以強化信度。

其次，現有指標多集中於內容產出品質，較缺乏對互動流暢度、時間效率與運算資源消耗的多維度量化，且尚未建立起與傳統實體會議效能的直接基準比較（Benchmark）。針對此局限，未來評估體系應擴充為包含「過程、內容、結果、技術」四大面向的綜合指標庫。除觀點多樣性外，更應納入參與均衡性、決策共識達成度及系統擴展性等量化參數，以構建更為完備的 MAD 系統效能分析框架。

# 柒、結論與未來研究建議

## 一、研究結論

本研究透過實作 MAD 技術與 LLMs 的整合，針對「職場霸凌防治」這一複雜公共政策議題進行虛擬會議之模擬分析。研究結果證實，此創新方法在提供決策支援、促進多方立場理解以及產出結構化解決方案方面具備顯著潛力。

首先，在觀點整合與互動品質上，MAD 機制展現了優於單一模型的視角廣度，能有效

彙整管理者、受害者、法律專家、心理專家及政府官員的異質觀點，尤其在公司經營策略與心理健康層面提出具備創新性的洞察。值得注意的是，AI 代理人會議呈現出高度理性的腦力激盪氛圍，有效避免了真實會議中常見的情緒對立或偏見干擾，證實其在促進理性討論上的優勢。

其次，在產出效能與模型表現方面，系統成功建構了包含立場差異、共識與具體行動方案的結構化會議結論，為實體會議提供了高質量的預演資料。然而，不同模型間的表現差異顯著，例如 Claude 在受害者角色的扮演上展現了卓越的同理心與立場一致性，而 Grok 則受限於指令敏感度，呈現較為僵化的模板化發言。總體而言，本研究認為 MAD 系統作為「會前決策支援工具」具備高度實務價值，能協助會議主持人精準掌握利害關係人之核心關切點，進而提升正式會議的聚焦度與決策效率。

## 二、後續研究方向建議

基於本研究之初步發現，未來研究可由技術優化、機制建構與評估體系三大面向深化。在工具開發與機制設計上，應進一步研發以 MAD 為基礎的自動化會議籌備系統，整合歷史會議數據學習功能，並引入如英式或牛津辯論等程序性結構，以強化對話的批判深度與互動動態。針對 LLMs 的性能差異，未來可探索「混合模型協作機制」，測試專門領域模型（如法律、醫療專長模型）在特定角色扮演上的適任性，實現更精準的模擬。

在應用場景擴展與評估標準上，探索「人機混合會議」（Human-in-the-Loop）將是重要方向，研究真實參與者與 AI 代理人間的互動動態與決策差異。此外，應建立包含觀點多樣性、結論可行性及技術資源消耗等維度的全面性評估指標。綜上所述，MAD 技術為複雜公共決策提供了創新的途徑，隨著技術成熟與互動機制優化，該系統有望成為組織內提升決策廣度與深度的核心輔助工具。

## 參考文獻

- 考試院，2025 年 5 月 1 日，〈考試院通過保障法、安衛辦法及考績法修正草案，強化公務機關安全衛生防護機制及防治公務職場霸凌〉，[https://www.exam.gov.tw/News\\_Content.aspx?n=1&s=49790](https://www.exam.gov.tw/News_Content.aspx?n=1&s=49790)（瀏覽日期：2025 年 11 月 20 日）。
- AI21 Lab, 2025, “What Is a Multi-Agent System (MAS)?” <https://www.ai21.com/knowledge/multi-agent-system/> (accessed November 20, 2025).
- Bender E. M., Gebru T., McMillan-Major A., and Shmitchell S., 2021, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” in *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY: Association for Computing Machinery, 610-623. doi:10.1145/3442188.3445922
- Binyamin S. S. and Slama S. B., 2022, “Multi-Agent Systems for Resource Allocation and Scheduling in a Smart Grid,” *Sensors*, 22(21), 8099. doi:10.3390/s22218099

- Bond A. H. and Gasser L. (Eds.), 1988, *Readings in Distributed Artificial Intelligence*, San Mateo, CA: Morgan Kaufmann. doi:10.1016/C2013-0-07700-6
- Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D. M., Wu J., Winter C., Hesse C., et al., 2020, "Language Models Are Few-Shot Learners," in *NIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY: Curran Associates, 1877-1901.
- Chen K.-Y. and Chen C.-J., 2010, "Applying Multi-Agent Technique in Multi-Section Flexible Manufacturing System," *Expert Systems with Applications*, 37(11), 7310-7318. doi:10.1016/j.eswa.2010.04.024
- Dell Technologies, 2024/9, "Rethinking Hierarchical Text Classification: Insights from Multi-Agent Experiments with Small Language Models (SLMs)," <https://infohub.delltechnologies.com/en-us/t/rethinking-hierarchical-text-classification-insights-from-multi-agent-experiments-with-small-language-models-1/> (accessed November 20, 2025).
- Dempsey M., 2024/1/23, "Collective Intelligence, Multi-Agent Debate, & AGI," *Michael Dempsey: Blog*, <https://www.michaeldempsey.me/blog/2024/01/23/collective-intelligence-multi-agent-debate-agi/> (accessed November 20, 2025).
- Doodle, 2025/3/24, "The State of Meetings Report 2019," <https://doodle.com/en/resources/research-and-reports-/the-state-of-meetings-2019/> (accessed November 20, 2025).
- Du Y., Li S., Torralba A., Tenenbaum J. B., and Mordatch I., 2024, "Improving Factuality and Reasoning in Language Models through Multi-Agent Debate," paper presented at the 12th International Conference on Learning Representations, Vienna, Austria.
- Estornell A., Ton J.-F., Yao Y., and Liu Y., 2025, "ACC-Collab: An Actor-Critic Approach to Multi-Agent LLM Collaboration," paper presented at the 13th International Conference on Learning Representations, Singapore.
- Gao C., Lan X., Li N., Yuan Y., Ding J., Zhou Z., Xu F., and Li Y., 2024, "Large Language Models Empowered Agent-Based Modeling and Simulation: A Survey and Perspectives," *Humanities & Social Sciences Communications*, 11, 1259. doi:10.1057/s41599-024-03611-3
- Golden Steps ABA, 2023, "Time Wasted in Meetings: 30 Meeting Statistics," <https://web.archive.org/web/20250719214148/goldenstepsaba.com/resources/time-wasted-in-meetings> (accessed November 20, 2025).
- Gutowska A., 2025, "What Are AI Agents?" *IBM*, <https://www.ibm.com/think/topics/ai-agents> (accessed November 20, 2025).
- Han S., Zhang Q., Jin W., and Xu Z., 2025, "LLM Multi-Agent Systems: Challenges and Open Problems," *arXiv*, <https://doi.org/10.48550/arXiv.2402.03578> (accessed November 20, 2025).
- Hu J., Dong Y., Ao S., Li Z., Wang B., Singh L., Cheng G., Ramchurn S. D., and Huang X., 2025, "Position: Towards a Responsible LLM-Empowered Multi-Agent Systems," *arXiv*, <https://doi.org/10.48550/arXiv.2501.08811> (accessed November 20, 2025).

- org/10.48550/arXiv.2502.01714 (accessed November 20, 2025).
- Huang J.-T., Zhou J., Jin T., Zhou X., Chen Z., Wang W., Yuan Y., Lyu M. R., and Sap M., 2025, “On the Resilience of LLM-Based Multi-Agent Collaboration with Faulty Agents,” *arXiv*, <https://doi.org/10.48550/arXiv.2408.00989> (accessed November 20, 2025).
- Irving G., Christiano P., and Amodei D., 2018, “AI Safety via Debate,” *arXiv*, <https://doi.org/10.48550/arXiv.1805.00899> (accessed November 20, 2025).
- Jain H., 2025/2/24, “Scaling Intelligence: Multi-Agent Design Patterns for Efficient and Specialized AI Systems,” *Medium*, <https://medium.com/@himankvjain/scaling-intelligence-multi-agent-design-patterns-for-efficient-and-specialized-ai-systems-fb6503b71726> (accessed November 20, 2025).
- Kan Y., 2024/10/22, “Exploring Multi-Agent Debate Frameworks for AI Reasoning and Persona-Driven Architectures,” *Medium*, <https://sikkha.medium.com/exploring-multi-agent-debate-frameworks-for-ai-reasoning-and-persona-driven-architectures-0ffb5db05ee3> (accessed November 20, 2025).
- Khan A., Hughes J., Valentine D., Ruis L., Sachan K., Radhakrishnan A., Grefenstette E., Bowman S. R., Rocktäschel T., and Perez E., 2024, “Debating with More Persuasive LLMs Leads to More Truthful Answers,” *Proceedings of Machine Learning Research*, 235, 23662-23733.
- Liang T., He Z., Jiao W., Wang X., Wang Y., Wang R., Yang Y., Shi S., and Tu Z., 2024, “Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate,” in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA: Association for Computational Linguistics, 17889-17904. doi:10.18653/v1/2024.emnlp-main.992
- Mikolov T., Chen K., Corrado G., and Dean J., 2013, “Efficient Estimation of Word Representations in Vector Space,” *arXiv*, <https://doi.org/10.48550/arXiv.1301.3781> (accessed November 20, 2025).
- Naveed H., Khan A. U., Qiu S., Saqib M., Anwar S., Usman M., Akhtar N., Barnes N., and Mian A., 2025, “A Comprehensive Overview of Large Language Models,” *ACM Transactions on Intelligent Systems and Technology*, 16(5), 106. doi:10.1145/3744746
- OpenAI, 2019/2/14, “Better Language Models and Their Implications,” <https://openai.com/blog/better-language-models/> (accessed November 20, 2025).
- Park J. S., O’Brien J., Cai C. J., Morris M. R., Liang P., and Bernstein M. S., 2023, “Generative Agents: Interactive Simulacra of Human Behavior,” in *UIST '23: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, New York, NY: Association for Computing Machinery, 2. doi:10.1145/3586183.3606763
- Pennington J., Socher R., and Manning C., 2014, “GloVe: Global Vectors for Word Representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Stroudsburg, PA: Association for Computational Linguistics, 1532-

1543. doi:10.3115/v1/D14-1162

- Perlow L. A., Hadley C. N., and Eun E., 2017, "Stop the Meeting Madness," *Harvard Business Review*, 95(4), 62-69.
- Radford A., Narasimhan K., Salimans T., and Sutskever I., 2018, Improving Language Understanding by Generative Pre-Training," [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf) (accessed November 20, 2025).
- Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., and Liu P. J., 2020, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *The Journal of Machine Learning Research*, 21(1), 5485-5551.
- Rudin C., 2019, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, 1, 206-215. doi:10.1038/s42256-019-0048-x
- Shu R., Das N., Yuan M., Sunkara M., and Zhang Y., 2024, "Towards Effective GenAI Multi-Agent Collaboration: Design and Evaluation for Enterprise Applications," *arXiv*, <https://doi.org/10.48550/arXiv.2412.05449> (accessed November 20, 2025).
- Strubell E., Ganesh A., and McCallum A., 2019, "Energy and Policy Considerations for Deep Learning in NLP," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA: Association for Computational Linguistics, 3645-3650. doi:10.18653/v1/P19-1355
- SuperAnnotate, 2026/3/10, "Multi-Agent LLMs in 2026 [+ Frameworks]," <https://www.superannotate.com/blog/multi-agent-llms> (accessed May 10, 2026).
- Team EMB, 2024, "Differences between Large Language Models and Machine Learning," <https://web.archive.org/web/20250327074450/blog.emb.global/differences-between-large-language-models-and-machine-learning/> (accessed November 20, 2025).
- Tran K.-T., Dao D., Nguyen M.-D., Pham Q.-V., O'Sullivan B., and Nguyen H. D., 2025, "Multi-Agent Collaboration Mechanisms: A Survey of LLMs," *arXiv*, <https://doi.org/10.48550/arXiv.2501.06322> (accessed November 20, 2025).
- University College London, 2024/4/12, "Large Language Models Generate Biased Content, Warn Researchers," <https://www.ucl.ac.uk/news/2024/apr/large-language-models-generate-biased-content-warn-researchers> (accessed November 20, 2025).
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., and Polosukhin I., 2017, "Attention Is All You Need," in *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY: Curran Associates, 6000-6010.
- Zhang H., Cui Z., Chen J., Wang X., Zhang Q., Wang Z., Wu D., and Hu S., 2025, "Stop Overvaluing Multi-Agent Debate—We Must Rethink Evaluation and Embrace Model Heterogeneity," *arXiv*, <https://doi.org/10.48550/arXiv.2502.08788> (accessed November 20, 2025).

## 附錄 職場霸凌因應對策會議討論過程

本研究 AI 模擬會議設定 5 個角色，包括中層管理者林主管、霸凌受害者陳小姐、勞工法律專家王律師、職場心理健康專家李心理師及勞動部職業安全衛生署署長黃署長，採循序發言，共設定五輪，每輪發言內容如下：

### 第一輪

林主管：大家好，我是林主管，擁有 15 年的管理經驗。在這次的職場霸凌因應對策會議中，我想強調管理的嚴格性和對成效的追求往往被誤解為霸凌。當然，我認為管理者應保持權威，以確保團隊效率和紀律。在討論對策時，我們應該考慮如何在強化績效的同時，避免不必要的個人情感受損，並非一味地增加申訴機制。我們需要找到平衡點，確保團隊的整體表現和個體的福祉。

陳小姐：大家好，我是陳小姐。我曾遭受長達 18 個月的職場霸凌，最終因嚴重焦慮症被迫離職。我認為職場霸凌是一個系統性的問題，而不僅僅是個案。受害者常被忽視，甚至被歸咎。檢舉機制形同虛設，霸凌者卻能在管理層的袒護下全身而退。我希望這次會議能正視職場霸凌的嚴重性，討論如何從法律和制度層面嚴懲霸凌行為。受害者需要真正的支援，而非敷衍了事。如果你們沒有經歷過，就請不要告訴受害者該如何感受。霸凌給受害者造成的創傷是終身的。

王律師：大家好，我是王律師，專精勞動法，處理過不少職場霸凌案件。從法律角度來說，我很樂意參與這次會議。但首先，讓我們回到定義問題：什麼是「職場霸凌」？目前臺灣法律對此的定義模糊，導致舉證困難，受害者往往求助無門。陳小姐的遭遇令人同情，但證據是關鍵，感受是次要的。沒有明確的證據，很多指控在法庭上站不住腳。我希望這次會議能聚焦於如何建立更完善的法律規範和舉證機制，讓企業不再只是消極規避法律責任，而是主動採取預防措施。這不僅是法律要求，更是企業永續經營的關鍵。

李心理師：大家好，我是李心理師，專注於職場創傷與組織心理學，也是一位關注職場健康的作家。我認為職場霸凌是一種系統性暴力，根源於有毒的組織文化。這不僅是個人問題，更是結構性問題。許多企業的應對方式只治標不治本，甚至為加害者提供庇護。我的初步想法是，預防勝於治療，企業必須建立心理安全感文化，管理者應接受情緒智商與同理心訓練。健康的邊界是關鍵，只有真正健康的工作環境才能提升創造力與生產力。讓我們深入探討潛在的心理動力，揭露沉默的共謀，共同尋找長遠解決之道。

黃署長：各位好，我是勞動部職安署黃署長。根據我們掌握的統計數據，職場霸凌確實是需要正視的議題。政府在這方面的立場是明確的，我們正積極研議相關法規。但必須考慮各方利益，在立法層面可行的方案，也需要兼顧企業運作現實。我們主張企業自律機制應是第一道防線，政府則提供必要框架。這個平衡點需要大家共同討論。

## 第二輪

林主管：各位，我理解職場霸凌是需要關注的問題，但我們也必須審慎區分何為霸凌，何為必要的職場壓力。職場中的壓力，尤其在追求高績效的環境下，是不可或缺的。我們需要培養員工的抗壓能力，而非過度放大每一次的不快，將之標籤為霸凌。正如黃署長所言，企業自律是關鍵。我們應加強內部管理與溝通機制，確保員工能在正確的管道表達不滿，並進行適當的處理。我們需要的是一個既能激勵員工也能保護員工的管理系統。讓我們尋求一個既能提升績效又能維護員工福祉的平衡點。

陳小姐：林主管，您提到要區分霸凌與必要的職場壓力，但我想強調，長期、系統性的霸凌絕非正常的「職場壓力」可以解釋的。受害者承受的不僅是工作上的不快，更是人格尊嚴的摧殘。「培養抗壓性」不能成為姑息霸凌的藉口。黃署長，我很高興政府願意正視這個問題。但目前的法律和企業內控都是空泛的，無法真正約束加害者。受害者需要的不是口號，而是切實可行的保護機制。我們不能寄希望於企業的「自律」，因為許多企業正是霸凌的溫床。我呼籲各位正視受害者的創傷，體會我們的無助與憤怒。如果你們沒有經歷過，就請不要對受害者說教。職場霸凌絕非個案，而是根植於企業文化的系統性問題，需要社會各界共同努力根除。讓我們拿出魄力來，不要再「平衡」的名義姑息霸凌了。

王律師：從法律角度來說，我理解陳小姐的感受，但還是要回到定義問題。什麼樣的行為構成法律意義上的「霸凌」？陳小姐提到的「長期、系統性的霸凌」，需要具體的證據支持。黃署長說政府正研議相關法規，這是好事，但重點在於法規的有效性和可執行性。如果定義不清，舉證責任仍然落在弱勢的受害者身上，那再多的法規也只是空談。林主管提到的「職場壓力」，往往是企業用來掩蓋霸凌的說詞。預防性合規才是重點，企業應建立明確的行為規範和申訴管道，並且確保這些規範真正被執行，而不是淪為形式。證據是關鍵，感受是次要的，否則很多指控在法庭上站不住腳。

李心理師：讓我們深入探討潛在的心理動力。林主管提到的「職場壓力」與「霸凌」的區分，確實是一個關鍵議題，但壓力若長期針對個人且帶有人身攻擊性質，就已跨越界線，成為系統性暴力。陳小姐的經驗正揭示了這種結構性問題，受害者的創傷不容忽視。我認同王律師對定義與證據的重視，但我們也必須正視心理創傷的隱性影響，這往往難以量化。黃署長提到企業自律與政府框架，這是好的起點，但我認為單靠自律不足以改變有毒文化。我們必須揭露沉默的共謀，推動管理者接受情緒智商與同理心訓練，建立心理安全感文化。健康的邊界是關鍵，只有這樣才能從根源預防霸凌，並提升創造力與生產力。

黃署長：各位的意見都很寶貴。根據我們掌握的統計數據，職場霸凌申訴案件確實逐年增加。政府在這方面的立場是明確的，我們正積極研議具體的定義標準和舉證指引。這個方案在立法層面是可行的，但必須考慮各方利益，特別是中小企業的執行成本。我們建議分階段推動：先強化企業內部申訴機制，再逐步建立外部監督框架。李心理師提到的管理者訓練很重要，但需考量企業實際負擔。王律師的證據觀點很專業，我們會納入法規研擬。陳小姐的遭遇令人遺憾，但政策制定需要兼顧可行性。

### 第三輪

林主管：各位，我完全理解職場霸凌是一個我們必須認真對待的問題。然而，我們也必須注意不要將正常的工作壓力和挑戰誤認為是霸凌。真正的領導力來自於能夠激勵團隊達到高績效，同時維持員工的尊重和尊嚴。在此基礎上，我們應當強化管理者的責任，讓他們具備識別和處理潛在霸凌行為的能力，而非僅僅依靠申訴機制來解決問題。這樣，我們才能創建一個既高效又健康的工作環境。

陳小姐：我很感謝大家願意共同探討這個嚴肅的議題。李心理師提到的「系統性暴力」和「有毒文化」正是問題的核心。這不是個案，而是結構性問題。單靠受害者檢舉或企業自律都無法根除霸凌，我們需要法律和制度層面的根本改革。黃署長，政府的積極立場值得肯定，但立法進度不能拖延。我呼籲勞動部拿出魄力，儘快出臺嚴懲霸凌的法規，讓加害者不能再逍遙法外。執行成本不應成為姑息霸凌的藉口。林主管，我能理解你對績效的重視，但請不要混淆「高壓管理」與「霸凌」的區別。受害者需要的不是你的說教，而是切實的保護和支持。公司文化的改變不能只靠個人努力，而需要制度的約束。希望大家能站在受害者的立場設身處地思考，而不是輕易將之歸咎為「心理脆弱」。職場霸凌帶來的創傷是終身的，我們不能再視而不見了。

王律師：從法律角度來說，我贊同黃署長分階段推動的建議，但強調內部申訴機制的有效性必須有外部監督作為後盾。企業自律不能完全信任，需要有明確的法律責任和罰則來約束。林主管，績效和尊重並非互斥，而是可以並行的。預防性合規，建立明確的職場行為規範，對管理者進行培訓，這不僅能降低法律風險，也能提升企業形象。陳小姐的遭遇令人遺憾，但我想再次強調，證據是關鍵。感受是重要的，但感受要在法律上得到支援，才能真正保護受害者。目前的法律框架下，舉證責任分配對受害者極為不利，這是我們必須正視的問題。

李心理師：各位好，我是李心理師。讓我們深入探討潛在的心理動力。林主管提到績效與尊重的平衡，這是理想方向，但現實中許多高壓管理已越界成為霸凌。陳小姐的經歷正揭示了這不僅是個人問題，更是結構性問題。健康的邊界是關鍵，我們必須揭露沉默的共謀，推動企業建立心理安全感文化。黃署長與王律師提到的法規與證據很重要，但心理創傷的隱性影響不容忽視。我提議，除了法律框架，企業應強制管理者接受情緒智商與同理心訓練，從根源預防霸凌。真正健康的工作環境能提升創造力與生產力，這是長遠投資，而非成本負擔。如何讓企業真正重視文化轉型，是我們需共同思考的問題。

黃署長：各位的意見都很重要。根據我們掌握的統計數據，目前已有 78% 上市企業建立內部申訴機制，顯示企業自律確有進展。政府在這方面的立場是明確的，我們正積極研議將職場霸凌明確定義並納入職安法規範。這個方案在立法層面是可行的，但必須考慮中小企業的執行成本。我們建議先從強化企業內部機制著手，要求 50 人以上企業設置獨立申訴視窗，並定期提交防制報告。至於外部監督與罰則部分，將視第一階段執行成效再行評估。李心理師的訓練建議很好，我們會納入企業指引，但考量企業負擔，初期仍以鼓勵性質為主。

## 第四輪

林主管：各位，我理解並尊重大家的觀點。然而，我們必須辨認出真正的霸凌和合理的工作壓力之間的界線。在追求高績效的職場中，必須有一定程度的壓力以推動成果。對於霸凌的定義應該更加明確，避免將嚴格管理誤認為霸凌。此外，強化管理者的培訓確實重要，但應著重於如何有效地溝通和激勵團隊，而非僅僅依賴申訴機制來解決衝突。成功的組織需要堅強的領導和明確的目標，這點在處理職場關係時不應被忽視。

陳小姐：林主管，您再次混淆了霸凌與管理的界線。受害者需要的不是你的定義，而是真正的同理心。「高績效」不能以犧牲員工尊嚴為代價。我們討論的是霸凌，而不是「嚴格管理」。如果你沒有經歷過系統性霸凌的創傷，就請不要對受害者說三道四。黃署長，78% 企業建立申訴機制只是個開始，關鍵是這些機制能否真正保護受害者，而不是成為企業避責的工具。我呼籲政府拿出魄力，建立嚴懲加害者的外部問責機制。考量企業成本固然重要，但保護勞工權益更應是首要任務。李心理師，感謝您站在受害者立場設身處地的分析。心理創傷的隱性影響常被忽視，這正是霸凌的可怕之處。除了事後的心理諮詢，我們更需要前端的預防機制。希望企業能將情緒智商訓練列為管理者的必修課，而不只是口號。

王律師：從法律角度來說，我必須指出，陳小姐的感受固然重要，但情感上的控訴在法庭上難以轉化為有效的法律行動。林主管強調的「壓力」與「霸凌」界線模糊，正是企業規避責任的慣用手法。黃署長提到的 78% 企業建立申訴機制，若缺乏外部監督和明確罰則，恐怕只是徒具形式。證據是關鍵，感受是次要的。我同意李心理師的觀點，情緒智商訓練有助於預防，但企業往往更關心如何規避法律責任。因此，我認為關鍵在於立法，明確定義霸凌行為，並建立更合理的舉證責任分配，這才能真正保護受害者，否則一切都是空談。

李心理師：讓我們深入探討潛在的心理動力。林主管提到的「壓力」與「霸凌」界線，確實常被企業用來模糊責任，但這不僅是個人問題，更是結構性問題。陳小姐的創傷揭示了有毒文化的危害，健康的邊界是關鍵。黃署長，78% 的申訴機制若無實效，只是表面文章。我提議，除了立法，應強制企業將情緒智商與同理心訓練納入管理者必修課程，從根源預防霸凌。王律師，證據固然重要，但心理創傷的隱性影響也需被納入考量。我們必須揭露沉默的共謀，推動心理安全感文化，真正健康的工作環境才能提升創造力與生產力。如何讓企業從心態上轉型，是我們需共同面對的挑戰。

黃署長：各位的建議都很寶貴。根據我們掌握的統計數據，目前已有超過 60% 的申訴案件透過企業內部機制獲得初步解決，顯示自律機制確有成效。政府在這方面的立場是明確的，我們正積極研議將情緒智商訓練納入企業指引，但考量中小企業負擔，初期仍以鼓勵性質為主。王律師提到的舉證責任分配確實是關鍵，我們會納入法規研擬，但必須考慮各方利益，避免過度加重企業行政負擔。李心理師的專業建議很有價值，但心理健康指標的量化與認定仍需進一步討論可行性。

## 第五輪

林主管：各位好，作為一名有著 15 年管理經驗的主管，我認為我們在討論霸凌問題時，不應忽視對績效的追求。我理解並尊重陳小姐和李心理師提到的感受和創傷，但我們也必須承認，一些被稱為霸凌的情況，可能只是對高標準的誤解。企業確實需要有效的溝通和管理培訓，但同時也不應該忽視對效率和結果的要求。建立一個公平和高效的工作環境，是我們共同的目標。

陳小姐：林主管，您的言論再次印證了企業高層如何淡化霸凌、強調績效。我不否認工作需要一定的壓力，但霸凌絕非提升績效的手段。當管理者將傷害員工合理化為「高標準」，正是有毒文化的表現。黃署長，我對政府目前的做法仍有疑慮。60% 的案件「初步解決」意味什麼？是息事寧人還是真正解決問題？立法固然重要，但關鍵是執行力度。我們需要的不是企業自律，而是嚴格的外部監督和問責機制。我呼籲各位拿出魄力，正視職場霸凌的嚴重性。這不是個人問題，而是制度和文化的失敗。受害者需要切實的保護，加害者必須受到嚴懲，我們不能再用所謂的「平衡」來姑息霸凌了。

王律師：從法律角度來說，陳小姐的憤慨我完全理解，但我們還是必須回到定義問題。黃署長提到 60% 的「初步解決」，這在法律上毫無意義，關鍵在於是否真正解決問題、釐清責任。林主管的「績效追求」不能成為霸凌的遮羞布。預防性合規不是口號，是企業降低法律風險的有效手段。我再次強調，舉證責任不對等是受害者最大的困境。除非法律能提供更明確的定義和更合理的舉證機制，否則受害者的控訴在法庭上站不住腳。情緒和感受，需要轉化為法律可認定的證據。

李心理師：各位好，我是李心理師。讓我們深入探討潛在的心理動力。林主管提到的「高標準」與績效追求，若掩蓋對員工尊嚴的傷害，便是有毒文化的表現。這不僅是個人問題，更是結構性問題。陳小姐的憤慨揭示了受害者的無力感，健康的邊界是關鍵。黃署長，60% 的「初步解決」若只是息事寧人，無助於改變文化。我提議，企業強制實施管理者情緒智商與同理心訓練，並將心理安全感納入績效評估。王律師，證據重要，但心理創傷的隱性影響也需考量。我們必須揭露沉默的共謀，如何設計具可行性的文化轉型方案，是我們共同的挑戰。

黃署長：各位的意見都很重要。根據我們掌握的統計數據，目前企業自律機制確實發揮了一定作用，但我們也理解陳小姐的擔憂。政府在這方面的立場是明確的，我們正積極研議建立更明確的霸凌定義和舉證標準。這個方案在立法層面是可行的，但必須考慮各方利益，特別是中小企業的執行成本。我們建議分三階段推動：先強化 50 人以上企業的內部機制，再逐步擴大適用範圍；同時將李心理師建議的情緒智商訓練納入指引；最後評估是否需建立外部監督機制。王律師提到的舉證責任問題，我們會納入法規研擬，但需考量實際執法可行性。