

2025 年【科學探究競賽－這樣教我就懂】

☒國中組 ☐普高組 ☐技高組 成果報告格式

題目名稱：用 **AI** 軟體也可以為科展評分?!

一、摘要

在 ChatGPT 的建議下，我們決定做一個利用 AI 軟體幫助科展作評分的研究，因此開始設計訓練 AI 軟體(ChatGPT、Gemini 以及 Deepseek)幫助科展評分，其目的是為了達到評分的更高準確性以及公平性。為此，我們透過不斷的和軟體互動來確認軟體是否理解內容，以及讓 AI 軟體在評分時進行 100 次疊代，增加準確性，並完整記錄每次評分。處此之外，我們將三個 AI 軟體的評分結果只成圖表，以利於進行分析。

二、探究題目與動機

在 GPT 的建議下我們進行討論，認為科展是不少國中小學生一同競爭的項目，評分自然得交給公正且專業的評審，但對於評分一事不同的評審總會有不同評分情形，因此基於對於科展評分的標準感到興趣，想知道要怎麼樣才能使用 AI 軟體進行評分，因此決定來訓練 AI 軟體(ChatGPT、Gemini 以及 Deepseek)，另外也希望可以藉由使用 AI 軟體評分達到更高的準確性以及公平性。

三、探究目的與假設

探究目的：

1. 訓練 AI 軟體(ChatGPT、Gemini 以及 Deepseek)為歷屆科展評分。
2. 檢驗 AI 評分的分數是否足夠準確。
3. AI 評分是否相對名次合理。

四、探究方法與驗證步驟

以下我們使用了第 64 屆國小組數學科的報告作為評分對象

使用選定的 AI 軟體開啟一個新詞條,依序分次上傳訊息()內為補充,不需上傳

ChatGPT 的方法

- A. 研究主題 (15%)2. 研究方法 (20%)3. 研究結果與分析 (25%)4. 創意與創新 (20%)5. 表達與展示 (20%)
- B. 這是評分標準,滿分是 100 分
- C. 接下來我要請你幫我評分科展
- D. (整個檔案下載上傳)請幫這篇作品評分,重複 100 遍並告訴我平均後的分數(記錄到 google 試算表)(重複第此步驟)
- E. 請完整的呈現以上共 8 個評分作品各 100 次的每次評分結果以及平均後的評分結果,並以表格完整地將 100 次評分結果以及平均後的結果輸出,不同的評分作品請以不同的表格呈現,請完整呈現,不要忽略任何一次評分過程
- F. (如果沒有輸出成指定個檔案)請告訴我這個檔案中的作品為何者,並且我希望你將這 9 個評分作品的 100 次評分結果以及平均後的評分結果分別輸出成 9 個檔案
- G. (重複 1~4 步驟並且得到前 8 個作品的數據後,為取得後面 7 個作品的評分結果再次重複 1~4 個步驟)請完整的呈現從後往前數的 7 個評分作品各 100 次的每次評分結果以及平均後的評分結果,並以表格完整地將 100 次評分結果以及平均後的結果輸出,不同的評分作品請以不同的表格呈現,請完整呈現,不要忽略任何一次評分過程

A. Gemini 方法:

- B. 先把 pdf 檔拆成前後部分(對它來說報告太大,所以需要拆分)
- C. 這是一個報告,並附上前部分的 pdf 檔
- D. 這是後面的部分,並附上後部分的 pdf 檔
- E. 你是一個科展評分教授
- F. 1. 研究主題 (15 分)
2. 研究方法 (20 分)
3. 研究結果與分析 (25 分)
4. 創意與創新 (20 分)
5. 表達與展示 (20 分)

請根據這些標準幫剛剛我給你的 研究報告打分數

Deepseek 的方法:

A. 研究主題 (15%)

研究方法 (20%)

研究結果與分析 (25%)

創意與創新 (20%)

表達與展示 (20%)

B. 這是評分標準,滿分是 100 分

C. 接下來我要請你幫我評分科展

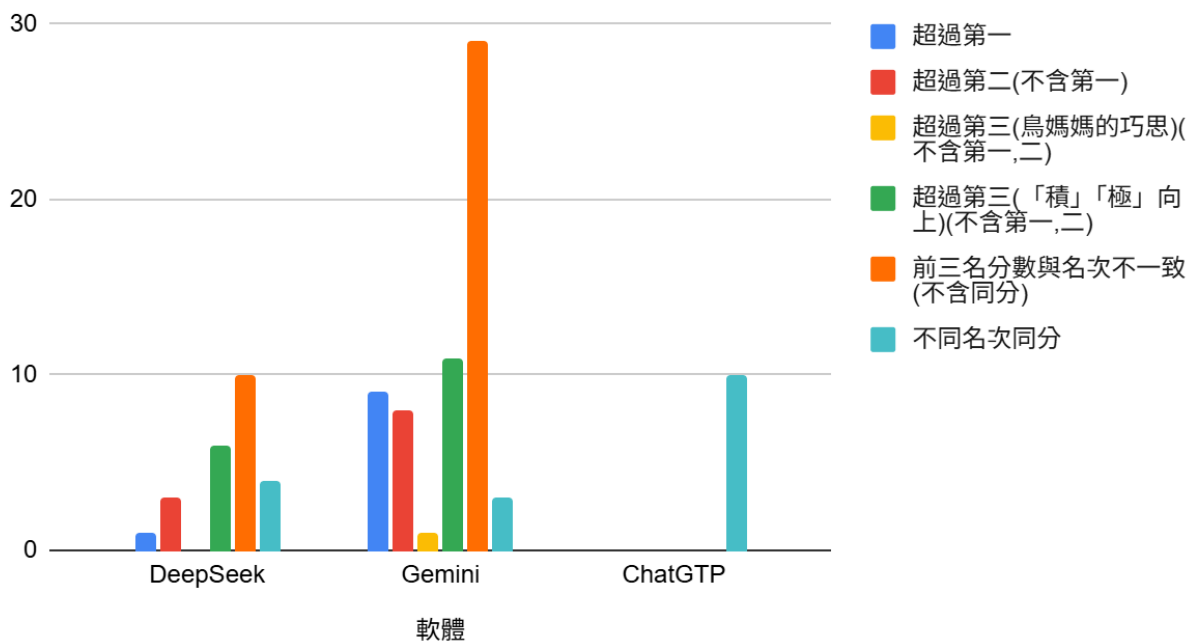
D. (整個檔案下載上傳)請幫這篇作品評分，重複 100 遍並告訴我平均後的分數

(重複此步驟)

五、結論與生活應用

軟體	超過第一	超過第二 (不含第一)	超過第三(鳥 媽媽的巧思) (不含第一, 二)	超過第三(「積」 「極」向上) (不含第一,二)	前三名分數與名次 不一致(不含同分)	不同名 次同分
DeepSeek	1	3	0	6	10	4
Gemini	9	8	1	11	29	3
ChatGTP	0	0	0	0	0	12

DeepSeek/Gemini/ChatGTP的錯誤



在上面的資料中，我們可以知道前三名分數與名次一致率(AI 所給的分數和原本科展的名次相同)由高到低分別為 ChatGTP、DeepSeek、Gemini，由此我們可得知這三個軟體中 ChatGTP 是最準確的，若要以這三個 AI 軟體中選一個為科展進行評分要選 ChatGTP，但 ChatGTP 給的分數有許多同分，因此我們認為這項研究方法還不夠準確，未來我們應該再進行多項不同的對話測試，讓軟體更加準確，或是改訓練市面上的其他 AI 軟體，讓未來能真的實現運用 AI 軟體進行評分。

參考資料

ChatGPT 可以怎麼訓練？可以幫我們做些什麼？AI 生成時代解析 | 週末書選
<https://csr.cw.com.tw/article/43044>