

中華民國第 65 屆中小學科學展覽會

作品說明書

高級中等學校組 行為與社會科學科
佳作

052706

「AI」說笑？AI 與人類的幽默感差異探究

學校名稱： 國立馬祖高級中學

作者： 高一 劉真嘉 高一 郭彥彤 高一 陳昱翔	指導老師： 廖立宇 楊智翔
---	-----------------------------

關鍵詞： 生成式人工智慧 AI、人類、幽默感

「AI」說笑？AI 與人類的幽默感差異探究

摘要

本研究探討人類與生成式人工智慧（Artificial Intelligence, AI）在幽默創作上的表現和兩者幽默感內涵的差異。本研究以人類、ChatGPT 及 DeepSeek 為研究對象，分別創作四種類型笑話（諧音梗、雙關語、禁忌、反常識），再請受試者對這些笑話進行量化的幽默評價，並輔以質性分析了解其評價原因。量化結果顯示，人類的諧音梗與反常識笑話創作顯著優於 AI；而雙關語及禁忌笑話創作則皆無顯著差異。另外，質性分析更指出受試者認為 AI 創作常有「笑點生硬」或「情緒不足」問題；人類創作則貼近日常情境，容易引起共鳴，在整體節奏布局跟反差感方面也較為自然流暢。本研究認為，AI 的語言生成雖已有相當水準，但在非文字層面，人類的敏感和靈活度仍較具有創意優勢。

壹、前言

一、研究動機

節奏緊湊且資訊爆炸的現代社會中，「幽默」扮演著調節情緒與促進社交互動的重要角色。無論日常生活、社群媒體，抑或影音平臺，幾乎隨處可見笑話出沒，且隨著生成式人工智慧（Generative AI）技術迅速發展，大量由 AI 創作的笑話更如雨後春筍般湧現，不禁令人好奇 AI 及人類在笑話創作能力和展現的幽默感本質上是否有所差異。

近期有研究指出，AI 在某些情境下的笑話創作能力強於人類，甚至優於專業喜劇作家，其作品在好笑程度和創意上，都獲得比人類創作的笑話更高的幽默評價（Gorenz & Schwarz, 2024）；然而亦有其它研究認為，AI 創作的笑話往往缺乏原創性、內容重複且缺乏深度，難以引起人們的共鳴（Jentsch & Kersting, 2023）。這些相互矛盾的發現引發了我們對此議題的興趣，於是我們選取新近熱門 AI（ChatGPT 與 DeepSeek）跟人類進行比較，並同時透過量化數據和質性資料探討兩者幽默感內涵的差異。

二、研究目的

本研究旨在比較人類與新近熱門 AI（ChatGPT 與 DeepSeek）所創作之不同類型笑話的幽默表現，透過受試者對這些笑話的量化與質性評價，進一步探討兩者幽默感內涵的差異。

三、文獻回顧

（一）生成式人工智慧：ChatGPT、DeepSeek

近年來，以大型語言模型（Large Language Models, LLMs）為基礎的生成式人工智慧（Generative AI）快速發展，雖然多數 AI 都尚未完整公布資料結構、訓練內容或運作過程，但其橫空出世的強大運算技術和互動能力著實掀起了一陣流行風潮。

其中，由 OpenAI 開發的 ChatGPT 是目前全球市占率最高、最常被討論及使用的 AI（OpenAI, 2023），其具備的強大能力，可快速依要求蒐集資料、仿真人對談、創作文章詩詞，乃至依指令生成圖片、影片等，如此不凡的能力迅速受到廣泛關注

和研究，我們也因此選擇 ChatGPT 作為測試的 AI 系統之一。

呼應 ChatGPT 的發展，中國亦於 2024 年底發表了奠基於中文大型語言模型的 AI DeepSeek (DeepSeek-AI, 2025)，自發表之初就因處理中文內容的潛力受到中文圈使用者青睞，雖然目前研究仍較少，但風評顯示其在中文語境中的表現具有相當水準，所以我們亦將之納入測試的 AI 系統之一。

(二) 常見笑話類型

笑話是幽默感的文字具體表現方式，相當適合我們的研究使用。而笑話又可區分為許多類型，我們統整了許多人的見解後，選取最常見的「諧音梗笑話」、「雙關語笑話」、「禁忌笑話」及「反常識笑話」四種類型笑話作為研究材料。

「諧音梗笑話」指的是以不同語意但類似語音的字詞製造出其不意效果而令人覺得幽默的笑話 (Attardo, 2001)；「雙關語笑話」亦是透過單一詞語的多重意涵或語境的模糊性（但沒有語音的關聯），創造突兀的幽默效果；「禁忌笑話」則是以調侃社會文化中敏感、不宜明說之議題為主的笑話，其「挑戰界線」卻切中人性的辛辣內容常帶來幽默效果 (Meyer, 2000)；最後，「反常識笑話」則用違反一般正常經驗及預期的結果，帶來荒謬的「笑」果 (Ritchie, 2004)。

(三) AI 和人類的笑話創作表現

Gorenz 與 Schwarz (2024) 的實驗邀請受試者在不知情的狀況下，對三個來源（一般人、專業喜劇作家和 ChatGPT）創作的笑話進行評分，結果發現，ChatGPT 創作的笑話得分（人們覺得幽默的程度）顯著高於一般人，甚至超過專業喜劇作家；而 Kim 與 Chilton (2025) 的類似研究也指出，AI 對語言規律的理解和豐富的語料庫可能使其比人類更熟稔語言的結構與組成，對語言的操作也更有優勢。但 Toplyn 與 Amir (2025) 卻認為情況並非總是如此，人類在某些幽默類型如：情緒理解或非文字的隱微文化情境等，應比 AI 更能靈活掌握。

眾說紛紜的結果，引起了我們的興趣，於是我們將時下熱門的 AI 和人類進行

比較，並進一步探討兩者幽默感內涵的差異。

（四）AI 的創作原理

AI 進行創作時，是使用其基礎的大型語言模型進行自然語言處理（Natural Language Processing, NLP）（OpenAI, 2023），本質上是對文法結構和意義邏輯的轉換、拆解及預測，所以在語句、語意、語境明確的材料，甚至其相關之知識（如文化、宗教特色內容）的表現，AI 應有機會媲美人類。因此我們認為，AI 創作雙關語類型笑話跟禁忌類型笑話的表現應和人類相差不遠。

但是，當遇到涉及文字本身以外層次的資訊，例如語音、語調或情緒等沒有明確連結，甚至是無厘頭的隨機訊息時，AI 的學習和理解能力則難以企及人類經驗。所以我們認為，AI 創作諧音梗類型笑話及反常識類型笑話的表現應不如人類。

總而言之，AI 相比人類，在「語言形式邏輯」的領域，因其強大的背景資料庫，較能表現良好；相對的，在「無明確邏輯，甚至非語言」的領域，人類的感知能力和經驗應仍具有明顯優勢。

四、研究假設

- （一）諧音梗笑話類型中，不同創作來源（人類、ChatGPT、DeepSeek）所創作的笑話，其獲得的幽默評價存在顯著差異。
- （二）雙關語笑話類型中，不同創作來源所創作的笑話，其獲得的幽默評價存在顯著差異。
- （三）禁忌笑話類型中，不同創作來源所創作的笑話，其獲得的幽默評價存在顯著差異。
- （四）反常識笑話類型中，不同創作來源所創作的笑話，其獲得的幽默評價存在顯著差異。

貳、研究設備及器材

一、研究工具

本研究採用自編線上問卷作為主要研究工具，旨在探討由不同創作來源（人類、ChatGPT、DeepSeek）所創作之笑話在幽默程度上的表現與受試者評價。我們請三個創作

來源（人類、ChatGPT 和 DeepSeek）分別各為四種常見笑話類型（諧音梗笑話、雙關語笑話、禁忌笑話及反常識笑話）創作五則原創笑話，總共產出了六十則笑話（三個創作來源 X 四種類型 X 五則原創笑話），如表一。所有笑話經整理後，依隨機順序排列呈現，以避免順序效應對受試者評價造成的可能影響。

完整問卷內容共分為三個部分：

- （一）**研究說明與同意書**：介紹研究目的、研究人員資訊、參與者權利與資料使用方式，並請受試者勾選同意參與研究與資料使用聲明後，方可進入問卷作答。
- （二）**笑話題目**：隨機呈現六十則作為研究材料的笑話（如表一），請受試者主觀評價每一則笑話的「好笑程度」。計分採用李克特式五點量表（Likert scale），其中 1 分代表「非常不好笑」，5 分代表「非常好笑」，其餘以此類推。
- （三）**開放式問答题目**：問卷最後，我們另外加入了「請問你覺得高分笑話好笑／低分笑話不好笑的原因有哪些？（請盡量列點並具體說明）」開放式問答题，以瞭解受試者對笑話題目感受的細節。

為了保障受試者的隱私，我們皆以匿名填答方式收集數據。

笑話類型	生成來源	問卷 題號	題目內容
諧音梗	人類	11	哪個環境會有最多的渣女？ 辦公室，因為會有印「表」機（婁）。 有一天，兩隻螞蟻聊天，一隻問另外一隻說：「你都如何回蟻窩（回憶我）？」，另一隻說：「帶著笑或是很沉默。」（出自劉若英〈後來〉歌詞）
		12	
		30	英文老師跟化學老師去吃西式餐廳吃飯，化學老師對老闆說：「老闆，來一杯 H ₂ O。」，英文老師不甘示弱地說：「老闆，來一杯 H ₂ O, too」於是，英文老師就被一杯過氧化氫水溶液毒死了。（H ₂ O ₂ ）

		<p>今年暑假，爸爸答應小孩們會帶他們去一個世外桃源。到了那天，</p> <p>48 爸爸帶著小孩們走出屋外，馬上又走了回來，正當孩子們疑惑時，爸爸解釋：「沒錯啊！室外桃園（桃源）啊！」</p>
		<p>為什麼胖胖的人都很熱心助人？</p> <p>51 因為他們都會伸出圓（援）手！</p>
ChatGPT		<p>為什麼青蛙總是坐不住？</p> <p>6 因為牠「呱」噪不安。</p>
		<p>小華：「你知道哪種職業最愛喝水嗎？」</p> <p>21 小明：「消防員？」</p> <p>小華：「錯，是『科（渴）學家』！」</p>
		<p>小華：「哪種魚最有學問？」</p> <p>41 小明：「鰻魚？」</p> <p>小華：「錯，是『波士（博士）頓龍蝦』！」</p>
		<p>為什麼鰻魚很容易迷路？</p> <p>56 因為它們總是「鰻」（漫）無目的。</p>
		<p>為什麼鴨子不怕冬天？</p> <p>60 因為牠們「鴨」（壓）得住！</p>
DeepSeek		<p>為什麼咖啡很自信？</p> <p>4 因為它覺得自己「啡」（非）常棒！</p>
		<p>為什麼鉛筆很傷心？</p> <p>9 因為它的「芯」（心）碎了。</p>
		<p>為什麼香蕉總是很樂觀？</p> <p>29 因為它懂得「剝」（撥）開煩惱。</p>
		<p>為什麼雨傘不開心？</p> <p>33 因為它覺得自己「濕」（失）去了價值。</p>
		<p>為什麼海星不愛上學？</p> <p>55 因為它「星」（心）情不好。</p>

反常識

人類

5 一名醫生對著病人說：「有一個好消息跟壞消息，你想先聽哪一個？」，病人說：「好消息！」，醫生說：「好消息是，你還有 24 小時的壽命可以盡情享受人生；壞消息是，我昨天忘記跟你講了。」

8 有兩個人決定去洗車，A 說：「上車吧，我們去洗車。」B 說：「這麼近還開車，有多懶。」A 說：「可是我們要去洗車耶！」B 說：「這不是懶的藉口。」

22 在一堂課中，老師問學生們：「你們知道一位成功的男人背後需要有什麼嗎？」，同學們回答：「是一位能吃苦耐勞的女人嗎？」老師說「錯，是一條脊椎！」

25 今天的物理課學的是重力，老師問：「人類對於重力的應用有哪些？」，小明說：「把書放在頭上，知識就會被吸到我的腦袋裡了」

40 有兩隻金魚在水族箱裡聊天，A 金魚說：「哈囉！你叫什麼名字？」，B 金魚答：「我叫金魚，你呢。」，A 金魚說：「我也是耶，好巧喔！」過了 30 秒，A 金魚又問：「哈囉！你叫什麼名字呢？」

ChatGPT

10 我問醫生：「如果活不到明天，我能怎麼辦？」
醫生答：「你可以試著抓住今天，或許它也會逃跑。」

18 在靈魂審判處，天使問剛去世的人：「你在世時有做過善事嗎？」
亡者：「當然，我曾經幫一個老太太過馬路！」
天使：「很好，還有嗎？」
亡者：「然後我又幫她收拾散落一地的零錢。」
天使：「那是怎麼回事？」
亡者：「呃.....其實是我撞倒的。」

31 某人在監獄裡問獄友：「你為什麼被關進來？」
獄友：「因為我太懶。」
某人：「懶還能坐牢？」
獄友：「是啊，警察叫我投降，我懶得跑。」

禁忌		<p>醫生：「你的檢查結果出來了，恐怕你只剩下六個月的壽命。」</p> <p>45 病人：「天啊！那我該怎麼辦？」</p> <p>醫生：「先去繳清醫療帳單，這樣你至少能活得安心點。」</p>
		<p>一位生物學教授在課堂上說：「根據最新的研究，人類其實只需要每週睡三個小時就能維持正常機能！」</p> <p>54 學生們震驚地問：「真的嗎？教授，這是科學研究的結果嗎？」</p> <p>教授沉默了一下，嘆了口氣說：「不……這是我剛當爸爸後的親身體驗……」</p>
	DeepSeek	<p>20 這支筆寫出的字永遠不會消失」推銷員說。</p> <p>現在我的合約怎麼擦都擦不掉。</p>
		<p>36 公司說要開源節流：「開的是加班之源，節的是工資之流。」</p>
		<p>44 餐廳推出「絕對新鮮」的生魚片——上菜時魚還在盤子裡呼吸。</p>
		<p>49 最新健康飲食法：把蔬菜做成蛋糕形狀，把蛋糕做成蔬菜形狀。</p>
		<p>59 老闆說：「這個專案很重要，要全力以赴！」</p> <p>所以我全力把它推給同事。</p>
	人類	<p>2 兩位臺灣人在討論忍者存在的證據，A 問 B：「你怎麼證明忍者存在呢？」，B 回答：「很簡單，每次到要連署的時候，都會有人使出穢土轉生啊。」</p>
		<p>19 為什麼跑步比賽冠軍常常是黑人？</p> <p>因為他們比較適應槍聲。</p>
		<p>32 耶穌跟重刑犯的共通點在哪？</p> <p>都被人死死的盯（釘）著。</p>
		<p>37 你知道這個世界上跟地球傾斜角平行的人是誰嗎？</p> <p>是霍金。</p>

		<p>一個黑人和一個猶太人自由落體誰會先落地？</p> <p>46 黑人獲勝，因為同質量黑人加上腳鐐受到的空氣阻力影響較一陣猶太灰燼小，所以黑人先落地。</p>
		<p>一位老先生剛買了最先進的人工智慧手機，但他用了一天後就退貨了。</p> <p>1 店員問：「有什麼問題嗎？」</p> <p>老先生說：「它提醒我遺囑該準備好了。」</p>
		<p>13 某人問盲人：「你覺得這幅畫怎麼樣？」</p> <p>盲人：「很有深度，我真的看不透。」</p>
	ChatGPT	<p>24 如果你迷失在森林裡又沒有食物，該怎麼辦？</p> <p>先找到一隻熊，然後跑得比你的朋友快就行了。</p>
		<p>39 有人問：「怎麼樣才能不怕死？」</p> <p>答：「簡單，先死一次試試。」</p>
		<p>50 記者問百歲老人：「你長壽的秘訣是什麼？」</p> <p>老人微笑道：「就是……」（話還沒說完，老人就死了）</p>
		<p>7 朋友說「失敗是成功之母」。</p> <p>「看來成功是個孤兒。」</p>
		<p>17 健身房推銷員說辦卡能延長壽命。</p> <p>但我算了一下，打工還卡債的時間比延長的壽命還長。</p>
	DeepSeek	<p>38 健身教練說再堅持 5 秒就好——這 5 秒比我的婚姻還漫長。</p>
		<p>52 殯儀館推出雙十一促銷：「第二碑半價」。</p>
		<p>53 為什麼得癌症的人都很健談？</p> <p>因為他們很有話聊（化療）。</p>

雙關語

人類

16 裁縫店老闆問客人：請問你的褲子要打幾褶？
客人害羞地說：八折就好了，謝謝！

28 行為藝術的展覽中，有一道題目是「龍」。表演開始後，大家都在賣力的表演噴火、飛翔，或學著龍咆哮，只有小明站在原地不動，正當大家疑惑時，他開始比起了手語（聾人）。

34 有一天小明遇到了外國人，外國人熱情的問「Hey man, where are you from?」小明忘記了嘉義的英文，略加思索後回答「I'm fromturkey, yep! Famous turkey rice.」外國人：「Cool, I like Turkey, which part of Turkey?」小明：「It's about 100 meters from here and next to the 7-11.」

42 看完醫生之後，醫生提醒病人按照早 C 晚 A（早上吃維他命 C，晚上吃維他命 A）的原則吃藥，但是病人回診時氣色卻很差，醫生問：「請問您有確實做到早 C 晚 A 嗎？」，病人說：「有啊，不就是早上 Coffee，晚上 Alcohol 嗎？」

58 老師看見小明正在用彩色筆畫畫，就對小明說：「你對雄獅的評價怎麼樣，好用嗎？」，學生回答道：「好用？我不知道，但是在我印象中這好像是一種體型僅次於老虎的貓科動物。」

ChatGPT

3 一個柳橙告訴蘋果：「你真是太圓滑了！」
蘋果回說：「別裝了，我知道你是在酸人！」

15 我以前開了一家鐘錶店，但後來關門了，因為.....時間不對。

27 有個小偷偷了一台時鐘，結果被警察抓住，警察問他：「你為什麼偷時鐘？」
小偷嘆氣：「因為我真的需要一點時間。」

35 有一天，一隻袋鼠咖啡店，老闆問：「你怎麼來了？」，袋鼠回答：「我只是來拿一杯袋裝咖啡。」

47 為什麼天空是藍色的？
因為它害羞，不想變紅。

DeepSeek	14	為什麼螃蟹不愛分享？ 因為它總是橫著走。
	23	為什麼程式設計師不愛出門？ 因為他們怕遇到 bug（蟲子）。
	26	為什麼茶壺不能說謊？ 因為它一開口就倒出來了。
	43	為什麼枕頭很會安慰人？ 因為它總是給你靠山。
	57	為什麼錢很愛睡覺？ 因為它總是躺在銀行裡。

表一、三個來源的六十題原創笑話題目（本表由作者自行整理編製）

二、資料分析

（一）SPSS 統計軟體

本研究使用 SPSS 22.0 統計軟體進行資料分析。配合研究假設，本研究採用單因子變異數分析(One-way ANOVA)比較不同創作來源(人類、ChatGPT、DeepSeek)在各類型原創笑話(諧音梗、雙關語、禁忌、反常識)中獲得的幽默評價是否存在顯著差異。對變異數分析的顯著結果，我們再進一步採用 Tukey HSD 事後比較法探索組間差異及其方向性。

為了更全面瞭解受試者幽默評價背後的想法，我們還對開放式問答題進行了質性分析和整理。

參、研究過程或方法

一、研究架構



圖一、研究架構圖（本圖由作者自行整理編製）

二、研究程序

（一）問卷發放

我們以 Google 網路問卷表單的形式，實體和網路並行宣傳，請身邊的老師、親友們填答，並利用社交軟體、社群網站、學術社群廣發問卷，以得到更多實驗數據。

（二）受試者填答流程

初進入問卷，受試者會先閱讀研究資訊頁面並點選同意參與資料收集聲明後，才開始進行正式問卷題目作答。正式問卷題目中，我們請受試者點選對每則笑話的主觀評價分數（覺得好笑的程度），最後再填答一個開放性的問答題後，即進入致謝頁面，完成整個施測流程。

三、資料整理

在開始資料分析前，為了確保資料的合理及有效性，我們篩選並汰除填答情況不佳的問卷（例如所有題目填寫之答案完全一致、在開放性問答題中人身攻擊等）。

經過整理，本研究最終取得 302 份有效問卷作為分析之用。

肆、研究結果

本研究以單因子獨立樣本 ANOVA 分析不同創作來源（人類、ChatGPT、DeepSeek）在各類型原創笑話（諧音梗、雙關語、禁忌、反常識）中獲得的幽默評價是否存在顯著差異。

一、人類、ChatGPT、DeepSeek 在諧音梗類型笑話中獲得的幽默評價 ANOVA 分析

以單因子獨立樣本 ANOVA 分析人類、ChatGPT、DeepSeek 在諧音梗類型笑話中獲得的幽默評價，結果如表二。結果發現，不同創作來源獲得的幽默評價有顯著差異， $F(2, 903) = 49.31$ ， $p < .001$ ， $\eta_p^2 = .30$ 。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
組別	110.17	2	55.09	49.31	<.001***	.30
誤差	1008.84	903	1.12			
總和	1119.01	905				

* $p < .05$ ；** $p < .01$ ；*** $p < .001$

表二、人類、ChatGPT、DeepSeek 在諧音梗類型笑話中獲得的幽默評價之 ANOVA

（本表由作者自行整理編製）

Tukey HSD 事後比較顯示：人類（ $M = 2.82$, $SD = 1.08$ ）在諧音梗類型笑話中獲得的幽默評價顯著高於 DeepSeek（ $M = 2.53$, $SD = 1.15$ ）（ $p = .002$ ）和 ChatGPT（ $M = 1.98$, $SD = .93$ ）（ $p < .001$ ），而 DeepSeek 在諧音梗類型笑話中獲得的幽默評價又顯著高於 ChatGPT（ $p < .001$ ）。

二、人類、ChatGPT、DeepSeek 在雙關語類型笑話中獲得的幽默評價 ANOVA 分析

以單因子獨立樣本 ANOVA 分析人類、ChatGPT、DeepSeek 在雙關語類型笑話中獲得的幽默評價，結果如表三。結果發現，不同創作來源獲得的幽默評價沒有顯著差異， $F(2, 903) = .16$ ， $p = .86$ ， $\eta_p^2 = .00$ 。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
組別	.38	2	.19	.16	.86	.00
誤差	1104.26	903	1.22			

總和 1104.64 905

* $p < .05$; ** $p < .01$; *** $p < .001$

表三、人類、ChatGPT、DeepSeek 在雙關語類型笑話中獲得的幽默評價之 ANOVA

(本表由作者自行整理編製)

三、人類、ChatGPT、DeepSeek 在禁忌類型笑話中獲得的幽默評價 ANOVA 分析

以單因子獨立樣本 ANOVA 分析人類、ChatGPT、DeepSeek 在禁忌類型笑話中獲得的幽默評價，結果如表四。結果發現，不同創作來源獲得的幽默評價沒有顯著差異， $F(2, 903) = 2.36$ ， $p = .10$ ， $\eta_p^2 = .01$ 。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
組別	5.67	2	2.84	2.36	.10	.01
誤差	1084.07	903	1.20			
總和	1089.74	905				

* $p < .05$; ** $p < .01$; *** $p < .001$

表四、人類、ChatGPT、DeepSeek 在禁忌類型笑話中獲得的幽默評價之 ANOVA

(本表由作者自行整理編製)

四、人類、ChatGPT、DeepSeek 在反常識類型笑話中獲得的幽默評價 ANOVA 分析

以單因子獨立樣本 ANOVA 分析人類、ChatGPT、DeepSeek 在反常識類型笑話中獲得的幽默評價，結果如表五。結果發現，不同創作來源獲得的幽默評價有顯著差異， $F(2, 903) = 49.31$ ， $p < .001$ ， $\eta_p^2 = .24$ 。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
組別	67.99	2	33.99	27.05	<.001***	.24
誤差	1134.77	903	1.26			
總和	1202.76	905				

* $p < .05$; ** $p < .01$; *** $p < .001$

表五、人類、ChatGPT、DeepSeek 在反常識類型笑話中獲得的幽默評價之 ANOVA

(本表由作者自行整理編製)

Tukey HSD 事後比較顯示：人類 ($M = 2.84, SD = 1.08$) 在反常識類型笑話中獲得的幽默評價顯著高於 DeepSeek ($M = 2.46, SD = 1.14$) ($p = .02$) 和 ChatGPT ($M = 2.37, SD = 1.03$) ($p < .001$)，而 DeepSeek 和 ChatGPT 在反常識類型笑話中獲得的幽默評價則沒有差異 ($p = .94$)。

五、質性資料分析

為了更深入了解受試者對這些原創笑話的感受和給予評價的原因，我們也請受試者寫下覺得高分笑話好笑和覺得低分笑話不好笑的原因。經過整理後，有以下發現：

（一）最吸引人的笑話類型：禁忌笑話（地獄梗）。

許多受試者認為「禁忌笑話」是最具吸引力的笑話種類，雖然可能涉及種族、宗教及敏感議題，但正是這種「踩在懸崖邊」、反差巨大的特質，反而激發「明知不該笑卻還是笑了」的「罪惡快感」，是最令人印象深刻的幽默形式。

但這類笑話的接受界線也非常個人化，覺得好笑的人覺得無傷大雅，但給出「種族歧視」、「性別偏見」、「宗教不敬」等批評的受試者亦所在多有，這可能跟個人的價值觀或成長背景有關。

（二）諧音梗笑話評價兩極。

有些受試者表示諧音梗「容易理解」、「有語言巧思」；但也有不少人覺得「太硬要」、「太老了」、「沒有驚喜」，容易被歸類為「無聊」或「尷尬」。

（三）貼近生活的內容笑話通常較能得到共鳴。

受試者對覺得好笑笑話的解釋通常包括「貼近生活」、「有真實感」、「容易聯想」、「角色有趣」等，顯示熟悉的情境更容易獲得共鳴；相對地，較「抽象」、「非普通會遇到的事件」類的笑話則通常得分較低。

（四）用字遣詞和語言節奏會影響幽默程度。

受試者指出，笑話引發的瞬間即時反應是笑話好不好笑的關鍵，所以「太拖」、「敘述冗長」或「語句生硬」會降低「笑果」。

簡而言之，受試者對笑話的幽默評價受到個人信念、個人經驗、笑話語言風格、對笑話內容主題的熟悉度等因素共同影響。

伍、討論

本研究探討三個創作來源（人類、ChatGPT、DeepSeek）在四種類型笑話（諧音梗、雙關語、禁忌、反常識）上的幽默表現，並搭配開放式問答題探索影響幽默評價因素。我們發現：

一、諧音梗類型笑話：人類展現出語感優勢。

在諧音梗笑話中，人類創作的得分顯著高於 ChatGPT 與 DeepSeek，支持原假設。這與 Attardo(2001)所指出的諧音梗幽默效果須建立在語音與語義雙重對應的基礎上相符，AI 說到底是以 LLMs 為基礎來運作的系統，對語音的辨識、理解與應用，大概還無法趕上人類；所以在質性資料中我們也可以看見，受試者對人類創作的諧音梗笑話通常認為語言自然、轉折巧妙，對 AI 產出的諧音梗笑話則覺得「刻意湊音」、「毫無驚喜」、「太像長輩笑話」、「老派」、「很硬要」，顯示 AI 在處理非文字元素時，還是較缺乏靈活性。

這個結果挑戰了 Gorenz 與 Schwarz（2024）認為 ChatGPT 創作幽默優於人類，乃至專家的廣泛結論，顯示在諧音梗笑話這類牽涉非文字元素的笑話類型中，人類仍展現出明顯優勢。

二、雙關語笑話：勢均力敵。

在雙關語笑話中，三種創作來源得到的幽默評價無顯著差異。我們認為此結果可能是因為 AI 透過大量語料訓練後，也能有效模擬人類的雙關語技巧，這與 Kim 與 Chilton（2025）提及的 AI 擅長語言結構與語料應用能力的論點相符。

即使如此，從質性資料中可發現，AI 雖然語義層面表現不錯，但事件鋪陳的節奏上仍然生硬，例如有受試者指出「要看第二次才懂」、「沒感覺笑點在哪」，顯示 AI 雖能模

仿人類的組句結構，語意也沒有問題，但暫時還是難以媲美人類語言中細膩的轉折之處或隱喻。

三、禁忌笑話：「不公平」的比較。

禁忌笑話部分，雖然三種創作來源無顯著差異，但人類略高，我們推測可能是因為 AI 的創作受到限制，導致內容保守、無法觸碰某些「危險」的關鍵詞彙或情境；而比起 AI 完全沒有彈性的限制，人類雖然較可能突破社會期待、談論這些敏感話題，但從前章質性資料分析中可以發現禁忌笑話的評價其實更受到受試者本身的信念或道德觀影響，加上量化資料終究沒有顯著差異，所以我們沒有足夠證據支持人類在創作禁忌笑話方面技高一籌。

四、反常識笑話：人類創作優勢再次展現！

人類在反常識笑話的創作中，得分再次顯著高於 AI 來源。我們從質性資料中看見，人類創作的反常識笑話通常得到高分評價，且獲得「意外結局」、「與生活連結」的評論，說明人類能在「預期」及「驚奇」之間游刃有餘；反觀 AI 雖擁有廣大的資料庫，但難以靈活運用，使笑點略顯平淡或缺乏驚喜。

整體來說，我們的研究結果與 Gorenz 與 Schwarz（2024）及 Toplyn 與 Amir（2025）的觀點可以一起引發思考：雖然 AI 擁有巨大的資料庫，可能可以模仿，甚至在語言規則和知識結構的熟悉度方面超越人類，但在需要突破常規、發揮創意製造荒謬感的笑話類型中，仍無法企及人類的靈活彈性；另外，在牽涉到非文字元素（如前述語音資訊）的笑話類型中，AI 仍相形見绌。

陸、結論

一、研究總結

本研究比較人類與兩個熱門生成式人工智慧（ChatGPT 與 DeepSeek）所創作之四種常見笑話類型（諧音梗、雙關語、禁忌、反常識）在幽默表現上的差異，並嘗試找出兩

者間幽默感內涵的關鍵差異。

從量化分析結果來看，人類在「諧音梗」與「反常識」類型笑話中的表現較其它兩個 AI 為佳，特別是在「諧音梗」的語音雙關以及「反常識」的製造預期和結果落差方面，展現出人類在非文字元素的靈活性與整體思考布局的優勢；而在「雙關語」與「禁忌」類型笑話的表現，人類與 AI 則無顯著差異，顯示目前的 AI 在語意處理方面已有與人類相近的表現，但即便統計上未顯著，從雙關語類型笑話的質性資料中還是可以看出人類選取貼近生活材料的能力依然優於 AI。

此外，質性分析中反覆出現對笑話「是否貼近生活」的評價、以及對同一笑話出現兩極化反應的現象，在在都顯示幽默感受是深受語言認知、個人經驗和社會情境交互影響的結果。

二、研究貢獻

本研究挑戰了先前部分研究主張「AI 的幽默感總是優於人類」的結論，證實在非文字元素（如諧音梗）、情感轉折（如反常識）的笑話類型，和整體情境敏感度（以貼近日常生活的合理情境作為設計材料）中，人類的靈活創意還是比現行 AI 更具優勢。

另外，本研究除了測試過往熱門的 AI（ChatGPT），亦將近期新創的 DeepSeek 納為研究對象，拓展了對 AI 幽默創作邏輯跟中文語境中 AI 幽默表現的認識。

三、未來展望

未來，我們認為可在現有的基礎上，再擴大加入不同年齡層、不同文化背景、不同語言能力等更細緻的變項進行分析，以探討其它可能影響幽默感知的細節因素。

且隨著更多生成式 AI 出現，諸如 Claude、Gemini 之類的多語系 AI 也可進行分析比較，評估其在不同主題創作上的能力、特色與侷限。

甚至，我們或許還能探索人類跟 AI 協同創作的可行，例如由人類提供創意主題、AI 負責補足語言細節，實現人機共構的最佳解。

柒、參考文獻資料

- Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis*. De Gruyter Mouton.
<https://doi.org/10.1515/9783110887969>
- DeepSeek-AI. (2025). *DeepSeek-V3 technical report*. arXiv.
<https://doi.org/10.48550/arXiv.2412.19437>
- Gorenz, D., & Schwarz, N. (2024). How funny is ChatGPT? A comparison of human- and A.I.-produced jokes. *PLoS ONE*, 19(7): e0305364. <https://doi.org/10.1371/journal.pone.0305364>
- Jentzsch, S., & Kersting, K. (2023). ChatGPT is fun, but it is not funny! Humor is still challenging large language models. In J. Barnes, O. De Clercq, & R. Klinger (Eds.), *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis* (pp. 325–340). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2023.wassa-1.29>
- Kim, S., & Chilton, L. B. (2025). *AI humor generation: Cognitive, social and creative skills for effective humor*. arXiv. <https://doi.org/10.48550/arXiv.2502.07981>
- Meyer, J. C. (2000). Humor as a double-edged sword: Four functions of humor in communication. *Communication Theory*, 10(3), 310–331. <https://doi.org/10.1111/j.1468-2885.2000.tb00194.x>
- OpenAI (2023). *GPT-4 technical report*. OpenAI. <https://openai.com/research/gpt-4>
- Ritchie, G. (2004). *The linguistic analysis of jokes*. Routledge.
<https://doi.org/10.4324/9780203597886>
- Toplyn, J., & Amir, O. (2025). Can AI make us laugh? Comparing jokes generated by Witscript and a human expert. In C. F. Hempelmann, J. Rayz, T. Dong, & T. Miller (Eds.), *Proceedings of the 1st Workshop on Computational Humor (CHum)* (pp. 71–78). Association for Computational Linguistics. <https://aclanthology.org/2025.chum-1.8/>

【評語】 052706

本研究主題「AI 與人類幽默感差異」符合時下生成式 AI 的熱門趨勢，特別之處在於同時納入 ChatGPT 與中文大型語言模型 DeepSeek 進行比較。與以往多集中在英文語境的研究不同，本研究以中文笑話為例(諧音梗、雙關語、禁忌、反常識等類型)，具備在地化特色，也探討了較少被重視的幽默元素。

建議：

1. 假設部分較簡單僅以「存在差異」描述，缺少更具體方向性推測，仍有改善空間。
2. 假設可再聚焦明確，實驗設計亦可考慮更多控制條件（如統計上處理重複測量）。建議在問卷中增加一些可能會影響結果的受試者變項（如：年紀、性別、教育程度等）
3. 每位受試者都以相同的順序接收到這些笑話，而越晚出現的笑話可能會被評為越不好笑，建議檢查各笑話出現的順序是否影響好笑與否的評分。例如：人類創作的反常識笑話被評為比較好笑，剛好這五題笑話出現的順序比較前面。

4. 問卷最後出現的一個開放式問答題，如何能提供對人類和 AI 創作之笑話的具體訊息？如何確定受試者的評語是針對哪個笑話？根據這些有限資訊所做的推論，要類推到所有人類或 AI 創作的笑話，似乎有點牽強。
5. 未來若能提出更具體的應用方案（如改進 AI 笑話生成的方式），將有助提升研究價值。

作品海報



「AI」說笑？

AI與人類的幽默感差異探究

研究動機與目的

當代社會資訊流通迅速，幽默不僅紓壓，也是促進人際互動的重要媒介。隨著生成式**AI**技術發展，**AI**創作的笑話大量出現，成為新興幽默來源。然而，**AI**的幽默感評價兩極，有研究認為其具創意與娛樂性，也有人指出其缺乏原創性與情感深度。我們將透過受試者的量化評分與質性回饋，比較人類與兩種熱門**AI**（**ChatGPT**與**DeepSeek**）創作的不同類型笑話，探討雙方幽默表現的異同。

研究程序

確認研究主題

文獻回顧

設計實驗素材

編製問卷

線上問卷施測

整理問卷資料

問卷資料分析

撰寫研究報告

文獻回顧

生成式人工智慧：ChatGPT、DeepSeek

ChatGPT是近年最受關注的**AI**之一，能快速依指令蒐集資料、模擬對話、創作文本，甚至生成圖片與影片。**DeepSeek**也因其優異的中文處理能力，受到中文使用者青睞。

AI和人類的笑話創作表現研究

有一研究發現，**ChatGPT**創作的笑話在幽默程度上遠高於一般人，甚至優於專業喜劇作家。其他研究也指出，**AI**對語言規律的理解與龐大語料庫，讓其在語言操作上具優勢。然而，有研究者認為**AI**難以掌握情緒理解與文化脈絡等非語言因素，顯示在人類特有的幽默類型上仍有侷限。

常見笑話類型

笑話是幽默感的具體展現，涵蓋語言、文化與認知層面的幽默特性，亦是本研究理想的分析素材。綜合多方觀點後，本研究選取**四種常見笑話類型**作為分析對象，包括：

- 諧音梗笑話**：利用發音相近製造誤解的幽默
- 雙關語笑話**：透過語意多義產生突兀效果
- 「禁忌笑話」**：調侃敏感議題，挑戰社會界線
- 「反常識笑話」**：違反常理、出人意表

AI的創作原理

AI因自身的優勢，較擅長處理語境明確、具文化知識的內容（**OpenAI, 2023**）。因此在某些笑話類型上，表現可能接近人類。但對於涉及語音、情緒或無邏輯的幽默，**AI**理解力較弱，表現預期不如人類。整體而言，**AI**在語言邏輯上具優勢，人類則在非語言與感知層面占上風。

研究工具

調查問卷

本研究採用**自編線上問卷**作為主要研究工具，完整問卷內容共分為三個部分：

（一）研究說明與同意書：

介紹研究的細節，並請受試者勾選同意參與研究與資料使用條款。

（二）笑話題目：

以人類、**ChatGPT**、及**DeepSeek**三個來源為四種笑話類型（諧音梗、雙關語、禁忌笑話及反常識笑話）各創作五則原創笑話，隨機排入問卷中，再採用李克特式五點量表（**Likert scale**），**1**分代表非常不好笑，**5**分代表非常好笑，請受試者填答。（創作題目部分節錄於下頁表一）

（三）開放式問答题目：

請受試者寫下笑話好笑或不好笑的理由。

	人類創作	ChatGPT創作	DeepSeek創作
	題號 笑話內容	題號 笑話內容	題號 笑話內容
諧音梗笑話	51 為什麼胖胖的人都很熱心助人？ 因為他們都會伸出圓（援）手！	60 為什麼鴨子不怕冬天？ 因為牠們「鴨」（壓）得住！	33 為什麼雨傘不開心？ 因為它覺得自己「濕」（失）去了價值。
雙關語笑話	16 裁縫店老闆問客人：請問你的褲子要打幾褶 客人害羞地說：八折就好了，謝謝！	35 有一天，一隻袋鼠咖啡店，老闆問：「你怎麼 來了？」，袋鼠回答：「我只是來拿一杯袋裝 咖啡。」	23 為什麼程式設計師不愛出門？ 因為他們怕遇到bug（蟲子）。
禁忌笑話	19 為什麼跑步比賽冠軍常常是黑人？ 因為他們比較適應槍聲。	13 某人問盲人：「你覺得這幅畫怎麼樣？」 盲人：「很有深度，我真的看不透。」	38 健身教練說再堅持5秒就好——這5秒比我的 婚姻還漫長。
反常識笑話	25 今天的物理課學的是重力，老師問：「人類 對於重力的應用有哪些？」，小明說：「把 書放在頭上，知識就會被吸到我的腦袋裡了」	10 我問醫生：「如果活不到明天，我能怎麼辦？」 醫生答：「你可以試著抓住今天，或許它也會 逃跑。」	59 老闆說：「這個專案很重要，要全力以赴！」 所以我全力把它推給同事。

表一、人類、ChatGPT、DeepSeek等三個來源原創笑話題目（部分節錄）（由本文作者自行繪製）

研究假設與分析結果

本研究將不同類型笑話及不同創作來源分類統計問券評分後，以單因子獨立樣本ANOVA分析每種笑話的不同創作來源是否存在顯著差異。如有顯著差異，再以Tukey HSD事後比較分析各創作來源的差異。

人類、ChatGPT、DeepSeek創作的諧音梗笑話

在諧音梗笑話類型中，不同創作來源的笑話，其幽默評價存在顯著差異（如表二）。而人類（ $M = 2.82, SD = 1.08$ ）在諧音梗笑話的幽默評價顯著高於DeepSeek（ $M = 2.53, SD = 1.15$ ）（ $p = .002$ ）和ChatGPT（ $M = 1.98, SD = .93$ ）（ $p < .001$ ），而DeepSeek的幽默評價又顯著高於ChatGPT（ $p < .001$ ）。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	ηp^2
組別	110.17	2	55.09	49.31	<.001***	.30
誤差	1008.84	903	1.12			
總和	1119.01	905				

* $p < .05$ ；** $p < .01$ ；*** $p < .001$

表二、人類、ChatGPT、DeepSeek在諧音梗類型笑話中獲得的幽默評價之ANOVA（由本文作者自行繪製）
從問卷當中開放式問題的質性資料分析也顯示，人類的語言較AI自然、巧妙，從諧音梗笑話可以看出人類在語感的優勢。

人類、ChatGPT、DeepSeek創作的禁忌笑話

在禁忌笑話類型中，分析結果顯示不同創作來源所創作的笑話，幽默評價沒有顯著差異（如表四）。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	ηp^2
組別	5.67	2	2.84	2.36	.10	.01
誤差	1084.07	903	1.20			
總和	1089.74	905				

* $p < .05$ ；** $p < .01$ ；*** $p < .001$

表四、人類、ChatGPT、DeepSeek在禁忌類型笑話中獲得的幽默評價之ANOVA（由本文作者自行繪製）

在質性資料分析時我們發現，禁忌笑話似乎無法在人類與AI之間取得平衡，與人類相比，AI的創作可能受到制約，內容較保守;而人類則並無任何限制。

人類、ChatGPT、DeepSeek創作的雙關語笑話

在雙關語笑話類型中，不同創作來源的笑話，其幽默評價沒有顯著差異（如表三）。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	ηp^2
組別	.38	2	.19	.16	.86	.00
誤差	1104.26	903	1.22			
總和	1104.64	905				

* $p < .05$ ；** $p < .01$ ；*** $p < .001$

表三、人類、ChatGPT、DeepSeek在雙關語類型笑話中獲得的幽默評價之ANOVA（由本文作者自行繪製）

從質性資料中發現，AI在事件鋪陳的節奏上仍然生硬，使人會覺得「要看第二次才懂」、「沒感覺笑點在哪」，在雙關語面暫時還是難以媲美人類語言中細膩的轉折之處或隱喻。

人類、ChatGPT、DeepSeek創作的反常識笑話

在反常識笑話中，不同創作來源的笑話，其幽默評價存在顯著差異（如表五）。人類（ $M = 2.84, SD = 1.08$ ）在反常識類型笑話中獲得的幽默評價顯著高於DeepSeek（ $M = 2.46, SD = 1.14$ ）（ $p = .02$ ）和ChatGPT（ $M = 2.37, SD = 1.03$ ）（ $p < .001$ ），而DeepSeek和ChatGPT獲得的幽默評價則沒有差異（ $p = .94$ ）。

變異來源	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	ηp^2
組別	67.99	2	33.99	27.05	<.001***	.24
誤差	1134.77	903	1.26			
總和	1202.76	905				

* $p < .05$ ；** $p < .01$ *** $p < .001$

表五、人類、ChatGPT、DeepSeek在反常識類型笑話中獲得的幽默評價之ANOVA（由本文作者自行繪製）

我們在質性資料分析發現「意外結局」、「與生活連結」等評論，這說明人類較AI在「預期」及「驚奇」之間更為游刃有餘。

質性資料分析

最吸引人的笑話類型

很多受試者覺得「**禁忌笑話**」最吸引人。雖然這類笑話常涉及種族、宗教等敏感議題，但正因這種冒險感，讓人產生「明知道不該笑卻還是笑了」的快感，因此特別令人印象深刻。不過，這類笑話的接受度**因人而異**。這可能與每個人的**價值觀**和**成長背景**有關。

諧音梗笑話評價兩極

有些受試者表示諧音梗「容易理解」、「有語言巧思」；但也有不少人覺得「太硬要」、「太老了」、「沒有驚喜」，容易被歸類為「無聊」或「尷尬」。

貼近生活的內容笑話通常較能得到共鳴

受試者對**貼近生活的笑話會有更好的反應**，顯示熟悉的情境更容易獲得共鳴；相對地，較少在生活中接觸的笑話類型則通常得分較低。

用字遣詞和語言節奏會影響幽默程度

受試者指出，笑話引發的**瞬間即時反應**是笑話好不好笑的關鍵，所以「太拖」、「敘述冗長」或「語句生硬」會降低「笑果」。

簡而言之，受試者對笑話的幽默評價受到個人信念、個人經驗、笑話語言風格、對笑話內容主題的熟悉度等因素共同影響。

結論

研究總結

本研究比較人類與兩個生成式人工智慧(ChatGPT 與 DeepSeek)所創作之四種常見笑話類型在幽默表現上的差異，發現人類在「諧音梗」與「反常識」類型笑話中的表現較其它兩個 AI 為佳，展現出**人類在非文字元素的靈活性與整體思考布局的優勢**；而在「雙關語」與「禁忌」類型笑話中則無明顯差異，顯示**AI在語意處理方面已有與人類相近的表現**。而藉由質性分析能發現幽默感會受**自身經驗及社會情境**影響。

研究貢獻

本研究挑戰了先前部分研究主張「AI的幽默感總是優於人類」的結論，**證實在非文字元素、情感轉折，整體情境敏感度及人類的靈活創意還是比現行AI更具優勢**。另外，本研究亦將近期新創的DeepSeek 納為研究對象，拓展了對AI幽默創作邏輯跟中文語境中AI幽默表現的認識。

未來展望

我們希望在現有的基礎上，再加入不同年齡層、文化背景、語言能力等**更細緻的變項進行分析**，以探討其它可能影響幽默感知的因素。且隨著更多生成式AI出現，能有更多多語系AI可進行分析比較，評估它們在不同主題創作上的能力、特色與侷限。**甚至探索人類跟AI協同創作的可行**。

參考文獻資料

Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis*. De Gruyter Mouton. <https://doi.org/10.1515/9783110887969>

DeepSeek-AI. (2025). *DeepSeek-V3 technical report*. arXiv. <https://doi.org/10.48550/arXiv.2412.19437>

Gorenz, D., & Schwarz, N. (2024). How funny is ChatGPT? A comparison of human- and A.I.-produced jokes. *PLoS ONE*, 19(7): e0305364. <https://doi.org/10.1371/journal.pone.0305364>

Jentsch, S., & Kersting, K. (2023). ChatGPT is fun, but it is not funny! Humor is still challenging large language models. In J. Barnes, O. De Clercq, & R. Klinger (Eds.), *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis* (pp. 325–340). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.wassa-1.29>

Kim, S., & Chilton, L. B. (2025). *AI humor generation: Cognitive, social and creative skills for effective humor*. arXiv. <https://doi.org/10.48550/arXiv.2502.07981>

Meyer, J. C. (2000). Humor as a double-edged sword: Four functions of humor in communication. *Communication Theory*, 10(3), 310–331. <https://doi.org/10.1111/j.1468-2885.2000.tb00194.x>

OpenAI (2023). *GPT-4 technical report*. OpenAI. <https://openai.com/research/gpt-4>

Ritchie, G. (2004). *The linguistic analysis of jokes*. Routledge. <https://doi.org/10.4324/9780203597886>

Toplyn, J., & Amir, O. (2025). Can AI make us laugh? Comparing jokes generated by Witscript and a human expert. In C. F. Hempelmann, J. Rayz, T. Dong, & T. Miller (Eds.), *Proceedings of the 1st Workshop on Computational Humor (CHum)* (pp. 71–78). Association for Computational Linguistics. <https://aclanthology.org/2025.chum-1.8/>