



## 高雄市立鼓山高中 114 學年度第一學期第二次段考《高二》寫作測驗

年 班 號 姓名

說明：本部分共有二小題，請依各題指示作答，答案必須寫在「答題卷」上。第一小題限作答於答題卷「正面」，第二小題限作答於答題卷「背面」。

作答使用筆尖較粗之黑色墨水的筆書寫，且不得使用鉛筆。若因字跡潦草、未標示題號、標錯題號等原因，致評閱人員無法清楚辨識者，恐將影響成績。

哈布斯堡王朝 (Habsburg dynasty) 在歐洲曾經是不可一世的存在。在長達數百年的時間裡，他們的血脈遍佈西班牙、奧地利乃至半個歐洲。但是這個家族幾乎每個人都有一個明顯的特徵：異常突出、厚重的下巴、以及下垂的嘴唇。這是後來廣為人知的「哈布斯堡下巴」(Habsburg jaw)。當時哈布斯堡家族為了鞏固權力、確保將龐大的領土和財富世代都牢牢鎖在家族內部，居然用了一個現在看起來令人不可思議、細思極恐的辦法：近親通婚 (inbreeding)。叔叔迎娶姪女，表兄妹結合，在好幾個世代當中，這種聯姻被哈布斯堡家族視為保持「血統純正」的必要手段。

他們確實成功地將權力和財富留在了家族中，但也開始將有害的遺傳隱性基因不斷累積和放大。近親通婚要是這樣持續 10 代以上，隨之而來的是高夭折率、癩癩、智力缺陷。這個詛咒一直延續到西班牙哈布斯堡王朝的末代國王卡洛斯二世。他身患多種遺傳疾病，終生殘疾，最後無法留下任何後代。一個曾經統治世界的王朝，就這樣因為這種「自我參照」和「基因多樣性的喪失」，走向了衰敗。

沒想到，數百年後在 AI 這個領域，這個古老的詛咒重新上演。

AI 產業已經開始面臨一個嚴重的問題：我們快要用完所有「可用的公開資料」了。根據估計，2026-2028 這段時間，AI 公司們就會把所有網路上的公開資料給吞嚥完。所以，現在 AI 公司在訓練 AI 時，基本上都是讓 AI 把網路上的資料「反覆看過好幾遍」。因此現在幾間大型的 AI 公司都面臨了資料短缺的問題。

人類產生出新知識的速度並沒有這麼快，就算好不容易產出有用的新知識，也是瞬間就被 AI 狼吞虎嚥吃掉了。然而頂尖的 AI 模型已經「吃掉」了網路上幾乎所有唾手可得的高品質文字和圖像。於是，為了讓 AI 繼續進化、變得更強，大家被迫轉向一個看似合理、卻暗藏風險的解決方案：「讓 AI 去學習由 AI 自己生成的資料」——這在 AI 領域，被稱為「合成數據」(synthetic data)。這在哈布斯堡王朝，被稱為「近親通婚」。

過去兩年，有幾篇學術論文已經證實，一旦我們這樣子做，經過幾次迭代之後，AI 就會變得越來越糟糕，甚至到了不堪採用的地步。研究人員因此給了這個現象一個暱稱：「哈布斯堡效應」。或者更專業一點的學術名稱：「模型崩潰」(model collapse)。

這個過程，就像你拿著一台影印機，去影印「上一次影印出來的影本」，每次都拿新印出來的文件再去影印，反覆這樣操作幾次之後，你會開始發現，越印越糊。

「原始的資料」其實像是一幅充滿了豐富細節、生動色彩、甚至帶有一些小小瑕疵的原創風景畫。當中這些「瑕疵」和「罕見筆觸」(例如冷門知識、幽默感、獨特的寫作風格等等)其實代表了非常重要的「基因多樣性」。

第一次影印 (AI 第 1 代)：AI 學習了這幅畫，並生成了一份「影本」(合成數據)。這份影本非常出色，99.9% 接近原版。但是，AI 作為一個統計模型，在訓練的過程當中，它會不自覺地「取平均值」。它會稍微「平滑」掉那些它認為不重要的「瑕疵」(罕見知識)，並稍微「強化」那些最常見的特徵(主流觀點)。

第二次影印 (AI 第 2 代)：次一代 AI 拿去學習的，是那份「第一次的影本」。它會學到那個 99.9% 準確、但「稍微平滑過」的版本。然後，它會在這個基礎上「再次取平均值」。

第 N 次影印 (AI 第 N 代)：N 代 AI 學習 前代 AI 的輸出……

這個過程不斷重複，這就是我們所謂的「自我參照迴圈」(Self-Referential Loop)。

經過 N 代「近親通婚」後，AI 開始顯現出與哈布斯堡王朝驚人相似的衰敗症狀：

(1)、特徵放大與平庸化 (the "jaw")：AI 變得極度「無聊」且「可預測」。就像「哈布斯堡下巴」這個特徵在每一代都被強化，AI 最「平均」、最「主流」的特徵也被無限放大。所有原創的、古怪的、有創意的「基因」(數據多樣性)都在這個「反覆影印」中被消磨殆盡。最終，AI 只會生成那些最「四平八穩」、最「政治正確」、但也最沒有靈魂的內容。

(2)、遺忘「尾部」(forgetting the "tails")：真實的世界充滿了「長尾」數據，也就是那些大量罕見但真實的知識。AI 在「取平均」的過程中，會最先丟棄這些「尾部」知識，因為它們不常出現。幾代之後，AI 會徹底「遺忘」現實世界的多樣性。它會變得越來越「純粹」，但也越來越「無知」。

(3)、錯誤放大 (amplified bias)：如果 AI (Gen 1) 犯了一個小錯誤 (例如，它在 1% 的時間裡認為「鯊魚

是哺乳動物」。AI (Gen 2) 會把這個「1%的錯誤」當作「真實資料」來學習。幾代之後，這個小錯誤會被不斷放大，直到 AI (Gen N) 堅定地認為「鯊魚絕對是哺乳動物」。

(4)、王朝終結 (total collapse): 研究證明了這個崩潰的終點: 如果完全依賴合成數據, 模型最終會「忘記」語言和現實的基本結構, 其輸出會退化成毫無意義、不斷重複的胡言亂語 (gibberish)。

所以, 這個與歷史雷同的「詛咒」已不再是理論。它是 OpenAI、Google 和 Anthropic 這些頂尖 AI 公司都正在拼命解決的核心難題。現在 AI 撞上的這道「數據之牆」, 其實就是其中一道阻礙擴展定律繼續發揮強大效用的牆。因此大家現在別無選擇, 必須想辦法使用「合成數據」來跨越數據之牆, 讓 AI 的能力繼續進步。

因此, 接下來的 AI 競賽, 至少在數據層面上, 關鍵已經不是誰能造出更大的模型, 而是誰能率先掌握「AI 的基因工程」:

\* 如何在合成數據時, 保持最大的「基因多樣性」?

\* 如何確保每一代訓練中, 都混入一定比例「新鮮的、真實的人類血液」(新的人類資料), 以避免「近親通婚」?

歷史在這邊也給了我們重要的啟發, 一個封閉的、只靠自我參照的系統, 無論它一開始多麼厲害, 最終都將走向僵化和衰敗。AI 想要通往更廣闊的未來, 就絕不能切斷與豐富、混亂、甚至充滿「瑕疵」的真實世界的連結。因此我認為, 人類在演化論累積下來的經驗和基礎、探索與創造力, 仍是下一代 AI 突破的重要關鍵。

註: 本文為摘引編纂。原作者——程世嘉:

史丹佛大學電腦科學碩士, 台大資管系學士。曾任 Google 軟體工程師, 參與 Android 多媒體框架、地圖及中文搜尋等專案, 將機器學習融入創新科技產品。2015 年潘文淵獎得主, 專長人工智慧, 也是台灣第一位登上 Google I/O 開發者大會的講者。現為跨國 AI 公司 iKala 愛卡拉的共同創辦人暨執行長、作家。

請分項回答下列問題:

問題(一): 依據上文, 請說明目前 AI 發展的困境, 以及突破此困境的關鍵為何? 文長限 120 字以內。(占 4 分)

問題(二): 近年 AI 快速而蓬勃的發展, 除了成為新型的快速資料處理工具, 也衍生了各種濫用造成的問題——原創與作偽、學術研究與智慧財產……。在閱讀本文之後, 試陳述你對 AI 世代的想法。文長限 420 字以內。(占 21 分)

# 高雄市立鼓山高中 114 學年度第一學期第二次段考高三國文

## 「國語文寫作能力測驗」試題

當你在街角的麵包店突然聽到一段熟悉的廣播音樂，腦海中彷彿瞬間浮現兒時和家人一起去公園野餐的畫面；或是當你在路上聽到車子喇叭的聲響，竟讓你聯想到小時候放學時總是匆忙過馬路的情景。這些看似無意間出現的「聲音提示」，之所以能讓我們的情緒穿越時空，彷彿搭上了一台「耳朵的時光機」，正是因為大腦中關於聲音與記憶的精妙互動。聲音就像一把溫柔且強大的「時光鑰匙」，在不經意間翻開了記憶的抽屜。大腦中的聽覺皮層與情緒、記憶中心的緊密合作，讓我們得以在熟悉的旋律或聲響裡「看見」從前。

——〈耳朵的時光機：聽覺記憶如何將你帶回童年的特定瞬間〉

### 請回答下列問題：

生活中充斥各種聲音。聲音透過聽覺傳遞，喚起人們的記憶和感受，一如文中熟悉的廣播音樂喚起作者的童年記憶。請以「聲音的記憶」為題，寫一篇文章，書寫你熟悉或思念的聲音，及其所召喚的記憶和感受。(占 25 分)