

## 加大柏克萊研究顯示：學生使用 AI 加劇分數膨脹

駐休士頓辦事處教育組

過去一個月來，《大西洋月刊》、《哈佛雜誌》以及《自然》雜誌等，紛紛發表文章抨擊高教界日益嚴重的「分數膨脹」現象。過去批評多歸咎於沒原則的教授、被寵壞並要求被當成「行政主管口中所謂顧客」的學生，以及普遍性的學術標準下滑。但加州大學柏克萊分校高等教育研究中心資深研究員 Igor Chirikov 發表的最新研究《人工智慧與分數膨脹》(Artificial Intelligence and Grade Inflation)，則揭示了新的核心關鍵：學生在課外缺乏監督下，使用 AI 撰寫報告或編寫程式對評量造成的實質衝擊。

這項研究採取計量經濟學常用的「雙重差分法」(difference-in-difference)，追蹤了德州一所研究型大學自 2018 至 2025 年間、橫跨 9 大專業領域、涵蓋超過 50 萬人次的修課巨量資料。結果證實，自 2022 年 ChatGPT 問世後，高度依賴課外作業且 AI 易於代勞的課程，其高分群出現爆炸性增長，拿 A 的學生比例足足增加了 30%，尤其在 2023 年該工具普及的第一個完整學期內，拿 A 率就迅速升高了 10%。

Chirikov 透過實證模型，排除了 AI 作為個別輔導提升學生實力，或是學生投機選修特定課程的「分類效應」。他發現，在同樣性質的寫作密集課程中，課外作業計分比重愈高(如從 10% 提高到 40%)的科目，分數飆升愈明顯，證實分數膨脹的主因是學生直接提交 AI 協助的產出作為評分基礎。

### 研究背景

Chirikov 進行這項研究有幾個原因。首先，分數膨脹是一個日益受到關注的議題。哈佛大學主管大學部的 Amanda Claybaugh 院長在 2025 年的一項研究中指出，在 2005 年至 2025 年之間，哈佛學院所頒發的 A 級分比例，已經從 24% 暴增到了 60.2%。

第二，Chirikov 說：「先前的研究大多把分數膨脹看作是『評分標準』的問題，也就是在探討指導教授、系所或學校，在評估學生作業

時是否變得越來越寬容」。相比之下，Chirkov 的研究看到了一種不同的機制：即學生在交作業前，AI 就已經把作業改得很完美了。如果 AI 沒有產生影響，我們預期在 ChatGPT 發表之後，高度借助 AI 的課程和較少借助的課程，其成績分佈應該會和發表前一樣，繼續保持平行發展。

第三，雖然與 AI 沒有直接關係，但 Chirkov 的研究讓他能夠去質疑「在校平均成績 (GPA)」的有效性，因為 GPA 在求職市場上，一直被當作代表學生技術與能力的一種替代指標。Chirkov 說 GPA 非常重要，因為雇主、研究所學程以及學生自己，往往都把 GPA 當作代表學生技術與能力的一種訊號。但是，如果分數上升是因為 AI 提高了交出去的作業品質，而不是因為學生學到更多東西，那麼這個訊號就會變得不再可靠。

他進一步表示：「研究顯示，AI 可能會讓這個問題變得更糟糕，因為它在那些最容易用 AI 來替代學生努力的課程中，把分數灌水得最嚴重。」

《人工智慧與分數膨脹》這項研究，延伸了 Chirkov 所提出的「基於任務導向來研究教育 AI」的研究方法（可參閱稍早發表在《世界大學新聞》的另一篇〈大學教師如何管理 AI：來自 3 萬 1,000 份課程大綱的實證，註 1〉文章）。

## 研究結果

在引入 ChatGPT 之後，成績的增長主要集中在成績分佈的頂端（高分群）。A 級分成績的平均比例從 2022 年的 44% 上升到 2025 年的 48%。在同一時期內，C 級分成績的比例從大約 8% 下降到 7%，而 D/F（不及格）級分的比例則從 3% 下降到 2%。

根據 Chirkov 的因果關係分析，AI 借助程度對 A 級分成績的影響最大。在高度借助 AI 的課程中，獲得 A 的學生比例增加 13%。隨著成績分佈階梯向下延伸，這種影響逐漸變小——獲得至少 A- 的比例增加了 9%，獲得至少 B+ 的比例增加了 5%，再往下的影響則更小。

更具啟示性的是，AI 的效應顯現得非常迅速。在一張名為《依

AI 借助程度劃分的成績效應》的圖表中顯示，在 2023 年(即 ChatGPT 被廣泛使用的第一個完整學期)，獲得 A 級分成績的學生比例就飆升了約 10%，並在 2024 年進一步增長。

### 因果關係的證據基礎

為了進一步驗證因果關係，Chirkov 特別引入「口頭報告佔比」作為「對照實驗」(placebo test)。他指出，目前的 AI 工具雖然非常擅長撰寫論文或編寫程式，但還沒辦法代替學生在教授面前進行口頭報告。如果分數膨脹確實是因為 AI 代勞造成的，那麼在口頭報告比例高的課程中，分數就不應該出現增長的現象。而對照實驗的數據結果顯示，口頭報告較多的課程，其分數完全沒有受到顯著影響。此一結果充分印證了分數膨脹具有高度的「任務特定性」，僅顯著存在於 AI 能輕易替代學生努力的傳統課外紙筆作業。

這項發現深刻動搖了「在校平均成績」作為求職或申請研究所能力指標的有效性。當哈佛大學頒發 A 級分的比例在過去 20 年間已從 24% 暴漲至 60.2%，AI 的介入更讓這份文憑價值加速失真，因為它虛胖了分數，卻未能反映實質學力的提升。Chirkov 強調，這份研究為全球高等教育的評量機制敲響了警鐘，迫使大學端必須全面反思並重構評量結構，減少可被 AI 包裝成「形似學習成果」的課外作業，轉向能真正檢測「學生真實技能」的測驗模式，否則，學校成績將無法傳遞真實資訊，而失去辨別人才資訊的公信力。

註 1：大學教師如何管理 AI：來自 31,000 份課程大綱的實證，來源：

<https://www.universityworldnews.com/post.php?story=2026020317034193>

撰稿人/譯稿人：藍先茜

資料來源：2026 年 5 月 14 日 University World News, Student AI use is fuelling grade inflation - Berkeley study (Retrieved from :

<https://www.universityworldnews.com/post.php?story=20260514074518988>) 。

AI 協作：本文採用 AI 協作，由 Gemini 3 蒐集、摘譯及分析資料，經核對及綜整而成。