

# 教育評估的難題：教師無法分辨答案來自人工智慧還是學生本人

駐歐盟兼駐比利時代表處教育組

人工智慧 (AI) 有許多聰明之處，其中之一就是它能幫助學生解答試題，然而這可能成為教育評估上的一大問題。根據英國雷丁大學 (University of Reading) 心理學和臨床語言科學系所發表的一份研究顯示，大部分教師幾乎無法區分答案來自真實學生或是人工智慧系統。

這份有關在教育評估中使用人工智慧的研究 (\*註) 顯示，在 94% 比例的情況下，人們都沒有意識到學生的作業答案是來自人工智慧聊天機器人 ChatGPT。在該系所一年級和二年級的心理學相關測驗中，聊天機器人所獲得的平均分數甚至比真正的學生還高，人工智慧出色的表現超乎研究人員的預期。英國雷丁大學 (University of Reading) 主導此研究的教授 Peter Scarfe 表示，「從先前的研究中我們知道，人工智慧可以提供絕佳解答，所以這樣的結果並不令人驚訝。然而我們所不知道的是，如果考試中 100% 的答案都由人工智慧編寫，評分者們是否會發現？」事實證明，教師們確實沒有發現到這一點。

## 還不夠聰明

雖然研究顯示，人工智慧在第一年和第二年測驗試題中的得分高於真實學生，但它並非在所有問題上都有相同表現。「人工智慧在回答該系所最後一年級的某個模組問題時表現得很吃力，因為它需要對資訊進行更深入的整合和分析。」這與當前人工智慧難以處理更抽象推理的觀點是一致的。「然而人工智慧不斷地在進步，所以未來它也將有能力解決更多類似的複雜問題。」

該項研究是迄今為止同類研究中規模最大的研究，研究人員表示該研究為教育敲響了「警鐘」，並建議學校和大學採取行動。「其中幾個解決方案包括校方需要轉向更真實的測試形式，例如，學生必須實際展示他們所學的知識，而不單單只用寫出來的評估方式，除此以外，也可在有更嚴格的監督方式下完成評估。」

## 人工智慧將繼續存在

近來大學及其他學校已開始透過電腦管理更多的考試和測驗，但

有可能會回到紙筆和監考人員走動的舊式考試方式。研究人員表示，這是一個已經發生的問題，「我們在雷丁大學正在做大量的研究，看看未來幾年的考試會是什麼樣子以及如何結合人工智慧以改善學習體驗。人工智慧將繼續存在，學生在畢業後也將在工作中繼續使用它。」

雷丁大學教授 Etienne Roesch 表示，「身為教育領域的一份子，我們需要就期望學生如何在學習中使用人工智慧達成一致的意見，這也適用於在生活的其他方面使用人工智慧，以避免社會出現誠信及信任危機。」他進一步解釋，「研究強調了我們身為資訊生產者和消費者的責任。我們必須要加倍努力維護學術誠信。」

雷丁大學負責研究學生體驗的教授 Elizabeth McCrum 補充，「人工智慧很明顯地正在改變我們生活中的許多層面，包括如何教學和評估學生」她強調，在這所英國大學中，他們並不反對使用人工智慧，而是希望將其融入教育，為學生提供更好的學習環境，使其能為未來的工作做更好的準備。「解決方案包括讓評估的模式更符合學生所需的技能，包括人工智慧的運用。我們已對所有課程進行了詳細研究，使其能夠應用人工智慧。我相信，我們現在已有能力向現在和未來的學生們傳授有關人工智慧的知識，並從人工智慧的快速發展中獲益。」

\*註

雷丁大學研究人員創造了 33 名虛構學生，並使用人工智慧聊天機器人 ChatGPT 產出了針對該系心理學學位考試的答案。結果表示，在第一年級第二年的測試中，虛構學生的平均成績比真正的學生高出一半。使用人工智慧工具產生的文章「幾乎無法檢測」。

撰稿人/譯稿人：陳蓓蕾

資料來源：2024 年 7 月 1 日，scientias 網站，<https://scientias.nl/probleem-in-het-onderwijs-docenten-weten-niet-meer-of-een-ai-of-een-echte-student-toets-maakt/>，2024 年 6 月 26 日，BBC，<https://www.bbc.co.uk/news/articles/cqqqln0eg65o>