

OpenAI 推出新工具試圖阻止其平台上的作弊行為

駐舊金山辦事處教育組

近日流行的人工智能工具 ChatGPT，可以一鍵就任何主題撰寫論文、詩詞和電子郵件等，2022 年底推出時即在整個教育界造成轟動。該工具的創始公司 OpenAI 於近日發布新程式，宣稱可以幫助審查使用人工智能機器人作弊的行為。

但是，新的檢測工具不僅沒有平息教師的恐懼，反而讓科技和教育界有些失望。OpenAI 稱檢測工具可在 15 分鐘內準確識別由 AI 撰寫的文本，但實際誤報率達 10%。即使是普林斯頓大學學生創建的不同檢測工具，都比其效果略好一些。一些專家擔心該公司的檢測工具可能會導致錯誤地指責學生抄襲。

史丹佛大學 AI 教育學習加速器計畫負責人 Victor Lee 表示「我很驚訝 OpenAI 居然以這樣的水準發布這樣的工具，如此輕率恐對學生有害。」

2022 年底 ChatGPT 發布後，全國各地的學區都驚慌失措。由於該工具能夠在幾秒鐘內撰寫出精確內容，ChatGPT 被許多人視為學生的終極作弊工具。有些人認為，這是對原創寫作的致命打擊。學生們開始繳交 AI 撰寫的作業、公司開始將 ChatGPT 納入著作權協議中。幾週之內，紐約市、洛杉磯和西雅圖等地的學校已經禁止使用它。

OpenAI 希望透過開發一種工具來區分人類撰寫及人工智能撰寫的文本，以防止機器人參與學校作弊和其他類型的人工智能詐騙。據該公司稱，該工具約 26% 可識別出由 AI 撰寫的文本，而約 9% 可能錯誤標記為 AI 撰寫。對於少於 1,000 個文字符號的文本，檢測效果較差。儘管如此，OpenAI 表示其文本檢測器仍比其他檢測工具更加準確。

撰稿人/譯稿人：Elissa Miolene/吳柏勳

資料來源：https://www.kenoshanews.com/ap/lifestyles/openai-launches-new-tool-to-deter-cheating-on-its-own-platform/article_40e7e28d-ded2-5f56-bd27-bc4676d16784.html