中小學使用「生成式人工智慧」注意事項(教師、行政人員及家長版)

中華民國 113 年 7 月 1 日臺教資(三)字第 1132702614 號函核定

近年來,「生成式人工智慧」(Generative AI)工具和「深偽技術」(Deepfake)等的蓬勃發展，為社會帶來許多新的機會和改變，這些技術改變了我們原本獲取知識、傳播訊息、創造內容的方式，同時也增加了許多使用風險。為了幫助中小學教師、行政人員及家長提升「生成式人工智慧」工具使用素養，以善用相關技術並避免造成誤用或濫用，提供以下六個要點作為使用參考。

一、**理解「生成式人工智慧」工具產生的內容可能有所偏誤：**由於用來訓練生成式人工智慧工具的資料來自於既有的紀錄或經驗，如果這些訓練資料本身帶有成見及錯誤，那麼使用生成式人工智慧工具產出的結果就會有偏差或錯誤,且工具本身無法自行判斷所產出的結果是否正確和合理。所以當我們在使用生成式人工智慧工具時，應該檢視與審查其產生的結果，以確保其正確性和合乎常理。

二、**理解「生成式人工智慧」工具可能會減少訊息的多樣性：**如果用來訓練生成式人工智慧工具的資料受到地域或文化的限制，且不夠多元廣泛，這些工具產生的結果可能僅能呈現單一文化的知識脈絡與觀點，甚至進一步強化原有的偏見。所以當我們在使用生成式人工智慧工具時，應該結合自己的經驗和批判性思維來檢視結果,而不是全盤接受生成的內容。

三、**理解「生成式人工智慧」內容的辨識工具僅能作為初步篩檢：**隨著生成式人工智慧工具的快速發展，有許多可以協助辨識生產內容的工具。所以我們需要了解這些辨識工具有其限制性，只能作為輔助與參考，仍須仰賴常理、直覺和其他證據來協助辨識內容是否來自於生成式人工智慧工具。

四、**察覺「深偽技術」日益逼真會產生不實的內容：**深偽技術是能修改臉部影像的深度仿造技術,原理是使用生成式人工智慧工具創建的虛假內容。這項技術能運用既有圖片、影像或聲音等素材，製造出看似真實的影片和圖像，甚至假新聞。所以當我們在觀看網路內容時，不要輕易相信未經審核的影片或照片，並留意該內容是否為深偽技術合成，和判斷可能的目的與動機。

五、**意識「生成式人工智慧」工具可能洩漏個人與組織的隱私與機密：**部分用來訓練生成式人工智慧工具的資料庫目前在取得、儲存和使用上都還沒有完備的法令、規範及倫理上的監管機制，因此，在使用生成式人工智慧工具時，提供的個人資料、敏感訊息及機密數據，都可能會被收錄

到訓練資料庫中，作為未來回應他人的內容。所以我們在使用生成式人工智慧工具時，應該審慎評估提供的資訊，是否具有機密性、隱私性與敏感性，以保護個人與組織的隱私與機密。

六、**避免過度依賴「生成式人工智慧」工具而侵犯智慧財產權與違反學術倫理：**使用人工智慧工具生成教案、試題、計畫等相關教學內容時，須謹慎檢視內容及用詞是否符合教育現場的使用規範與標準。所以我們需規範使用生成式人工智慧於學業的時機與方式，並提醒學生使用時可能會侵害他人的智慧財產權，以及有違反學術倫理的疑慮，如沒有提供出處將造成概念上的抄襲。

生成式人工智慧技術的發展，為我們的生活帶來了許多便利，並廣泛運用在各種情境中，卻也伴隨著一定的風險和挑戰。在這個數位時代，我們應該保持對資訊來源的高度警覺，不要輕易相信未經證實的訊息，並學會如何辨別虛假資訊。同時，我們要提升自己思辨的能力，批判性地分析和評估生成式人工智慧工具所產生的內容，避免被誤導。遵守相關的道德和法律規範，確保使用生成式人工智慧工具時不違反社會常規與資訊倫理。最後，我們應該加強自己的數位素養能力，才能在享受科技進步帶來高度便利的同時，減少科技帶來的風險，讓負面影響降到最小。

附錄 1

中小學使用生成式人工智慧注意事項(教師、行政人員及家長版)示例

一、**理解「生成式人工智慧」工具產生的內容可能有所偏誤：**例如，當我們要求生成式人工智慧工具建議旅遊行程時，如果系統本身的資料庫中沒有該地區的氣候環境、地理位置、社會人文以及文化限制等資料，提供的內容可能來自各種網路遊記文章的綜合體，結果就有可能是一份不順路、充滿非當季活動，甚至包含了虛構景點的行程。

二、**理解「生成式人工智慧」工具可能會減少訊息的多樣性：**例如，當我們向生成式人工智慧工具詢問法律或文化問題時，這些工具可能會是基於研發者國家的法律和文化習俗產生的答案。好比我們要求人工智慧工具生成一張新娘圖片時，它可能產生一張穿著白紗的西方臉孔女性，而不是根據使用者當地的文化習俗來產出不同膚色或其他婚禮的服飾。

三、**理解「生成式人工智慧」內容的辨識工具僅能作為初步篩檢：**例如，當我們使用生成式人工智慧內容辨識工具時，可以快速比對兩篇文章或多篇文章的相似程度，但這些比較結果僅能作為參考。要判斷文章是否為文字或概念抄襲，或根本是由生成式人工智慧工具創造的內容，都仍需要個人比對資料、詳細閱讀理解後，才能進行判斷。

四、**察覺「深偽技術」日益逼真會產生不實的內容：**例如，網路上常有知名人士發表演說或鼓勵投資的影片，面對這些內容，我們必須謹慎且小心求證知識的內容和來源。在深偽技術蓬勃發展的網路環境中，這些影片可能未取得影片主角的同意，或在他們根本不知情的狀況下，被深偽技術整合他們的臉(聲音)到一些完全虛假或有損名譽的影像作品中。

五、**意識「生成式人工智慧」工具可能洩漏個人與組織的隱私與機密：**例如，當我們不清楚生成式人工智慧工具的原理及規範，以公司個人資料文件或商業機密為題材向這些工具詢問解答，個人或公司文件、機密程式碼便有可能被收錄到這些工具的訓練資料庫中，而當其他的使用者再度詢問類似問題時，生成式人工智慧工具以收錄的資料庫回答問題，就有機會造成個人隱私或公司機密外洩，形成資安漏洞。

六、**避免過度依賴「生成式人工智慧」工具而侵犯智慧財產權與違反學術倫理：**例如，當我們使用人工智慧工具產生計畫書時，若有些用詞或字句非一般使用的習慣，應進行修正；或用於出題時，應檢視題目的合宜性及答案的正確性。

附錄 2

生成式人工智慧內容辨識工具與特色說明彙整表（舉例）

AI 生成內容辨識工具	工具特色說明	辨識文字/圖片/影音	資料來源
Content at Scale	Content at Scale 的研發團隊認為 ChatGPT、Claude 和 Gemini 等生成式人工智慧工具所產製的文本會留下某些特殊用詞或語法結構的痕跡，因此，研發團隊使用了大量的部落格文章、維基百科條目、論文及多個大型語言模型（LLM）所產出的文章進行學習強化訓練，藉以偵測文本內容出有多少機率是由生成式人工智慧所產生的。	文字	https://contentatscale.ai/ai-content-detector/
CopyLeaks	官方宣稱擁有 99%以上的辨識準確率和 0.2%誤判率，目前可檢測 ChatGPT、Gemini 和 Claude 等 30 種以上語言的人工智慧內容，且支援瀏覽器擴充及整合到網站或學習管理系統（LMS）中。	文字	https://copyleaks.com
GPT Detector	主要功能為評估多少比例的內容可能是由 GPT3、GPT4 或 ChatGPT 生成的。一般使用者皆可免費使用，但為避免濫用，有設定每日使用額度。另有主動聲明會對受檢資料加密，且不會進行儲存。	文字	https://x.writefull.com/gpt-detector
GPTZero	GPTZero 會在文件、段落和句子等層級中分析人工智慧生成的文本內容，並提供可能為生成式內容的機率，以及對檢測結果的信心程度數據。亦可對一系列的文本檔案進行辨識，最多同時比對 50 個檔案，總容量不超過 15MB，且每份文本最多僅能擷取 50,000 個字元。	文字	https://gptzero.stoplight.io
Hive Moderation	Hive Moderation 可以辨識生成人工智慧生成的文本、圖片與影片及音訊，並可以外掛在 Chrome 擴充功能。辨識結果會以百分比呈現，並依據檢測結果標示出可能包含人工智慧生成內容之範圍。	文字、圖片、影片	https://hivemoderation.com/ai-generated-content-detection
Hugging Face	Hugging Face 是一間人工智慧機器學習工具的開發商，早在 ChatGP 尚未問世的 2019 年，就已架設出人工智慧內容辨識網站，使用者只需要輸入約 50 字，它就能給出此段內容是否由人工智慧生成的概率。	文字	https://huggingface.co/learn/nlp-course/zh-TW/chapter4/2
Scribbr	Scribbr 適用於 ChatGPT、GPT4、Gemini 生成內容的辨識，透過檢測文本內容的具體特徵（例如句子結構或長度、單字選	文字	https://www.scribbr.com/ai-detector/

AI 生成內容辨識工具	工具特色說明	辨識文字/圖片/影音	資料來源
	擇和可預測性）來驗證文字的原創性和真實性。此工具介面簡易，無須建立帳戶即可免費執行無限次的文本辨識。為求準確度，建議提供的文本內容應在 25 至 500 字之間。		
Turnitin	Turnitin 為目前全球使用率最高的線上偵測剽竊系統，主要是可以上傳個人論文檔案，並於幾分鐘之內自動計算出與本文有相似文字(片段)的百分比率，挑出該段內容及可能的原始出處，可提供研究人員自行偵測其文章的原創性，並有助於提升論文的可信度。	文字	https://www.turnitin.tw/
Winston ai	它可以檢測使用 ChatGPT、GPT4、Gemini 和其他生成式工具的內容，官方提供的準確率高達 99.98%。目前已支援多國語言，例如：英文、法文、德文、西班牙文等；支援光學字元辨識（OCR）技術，可以從圖片或掃描檔案（包括手寫的內容），來擷取文字。	文字、圖片	https://gowinston.ai
Writer	一次可以檢查是否為 AI 生成的文字量範圍最少為 60 個字，最多為 5,000 個字，如果一個人使用非常相似的單字序列書寫，那麼他們的書寫也可以觸發偵測器。該判斷工具不會 100% 準確，但可以幫助指示某些東西是人工智慧產生的可能性。	文字	https://writer.com/ai-content-detector/
快刀	此系統是結合文本中的語言特徵進行分析，以識別文章是否由生成式人工智慧生成，可辨識由 GPT4、GPT3、GPT2、BERT、Jasper 等人工智慧語言工具所生成的文本，還可以檢測文章是否使用了翻譯軟體。	文字	https://ai.ppvs.org/

備註：

1. 以上由工具名稱字母排序。
2. 除了 Hugging Face 以外，其他辨識工具主要都需要收費(會有免費版的)。
3. 生成式人工智慧內容辨識工具預計將持續研發，本表僅提供目前常見的版本供參考。