# 教育與研究之
# 生成式人工智慧應用指引

## Guidance for generative AI in education and research

原文作者 unesco 翻譯製作 AREE 臺灣學術倫理教育資源中心 Center for Taiwan Academic Research Ethics Education

### ◆◇ 聯合國聯合國教科文組織（United Nations Educational, Scientific, and Cultural Organisation，UNESCO）——全球教育的領導者

教育是聯合國聯合國教科文組織的首要任務，因為這是一項基本人權，也是和平與永續發展的基礎。聯合國教科文組織是聯合國（United Nations）的教育領域專門機構，引領全球與區域合作以推動進步，並強化各國教育體系的韌性與能力，以滿足所有學習者的需求。聯合國教科文組織也引領各種努力，透過變革性學習（transformative learning）來回應當代的全球性挑戰，並在所有行動中特別關注性別平等與非洲地區的困境。

### ◆◇ 2030年全球教育議程

聯合國教科文組織作為聯合國在教育領域的專門機構，受託領導並協調2030年的教育議程。該議程是全球消除貧窮運動的其中一環，目標是在2030年之前實現聯合國所訂定的17項永續發展目標（Sustainable Development Goals，SDGs）。教育是實現所有目標的關鍵因素，其中包含專屬的第4項目標（目標4），旨在「確保包容性與公平兼具的優質教育，並且提倡人人都擁有終身學習的機會。」《2030年教育行動框架》（The Education 2030 Framework for Action）為這個宏大的目標與承諾提供了指引。

## 注意事項

1. 本文件英文原版名稱為《Guidance for generative AI in education and research》，是由聯合國教育、科學及文化組織（United Nations Educational, Scientific and Cultural Organization，英文簡稱UNESCO，中文簡稱聯合國教科文組織）於2023年出版，原版檔案：https://unesdoc.unesco.org/ark:/48223/pf0000386693
2. 教育部臺灣學術倫理教育資源中心於2025年依據本文件所採用的CC BY-SA 3.0 IGO授權（https://creativecommons.org/licenses/by-sa/3.0/igo/），進行繁體中文版之翻譯與製作，全文內容為作者觀點，不代表聯合國教科文組織、教育部臺灣學術倫理教育資源中心、主管機關立場。
3. 若本文件與英文原版有不一致的地方，請以英文原版為準。
4. 科技日新月異，若內容有不合時宜之處，請以您閱讀當下的法令與環境為準。
5. 轉載與引用時請遵守引用格式規定。

# 綜 述

## 從人本角度探討生成式人工智慧

可公開取得的生成式人工智慧（GenAI，以下簡稱生成式AI）工具正迅速崛起，各國監管機構的調適與因應追不上各種迭代版本發布的速度。多數國家均未針對生成式AI制定全國性規範，導致使用者的個資隱私缺乏保障，而大部分的教育機構也尚未做好相關驗證工具的準備。

聯合國教科文組織針對生成式AI在教育領域之使用所編定的第一份全球性指南，旨在幫助各國立即採取行動、規劃長期政策及培育人才，確保這些新技術朝以人為本的方向發展。

本指引評估了生成式AI對核心人文價值觀可能構成的潛在風險，這些價值觀包括：促進個人主體性、包容性、公平性、性別平等，以及語言與文化多樣性，同時涵蓋多元觀點與表達自由等。

文中提出了政府規範生成式AI之使用時應採取的關鍵步驟，包括強制保護資料隱私與考慮訂定年齡限制。其概述了對生成式AI供應商的要求，確保這項技術在教育領域的運用符合道德規範且具有成效。

本指引強調教育機構需要驗證生成式AI系統在道德與教學方法上是否適合教育，並且呼籲國際社會思考這類系統對知識、教學、學習與評估造成的長期影響。

本出版物向政策制定者與教育機構提出具體建議，說明如何設計與應用生成式AI工具來保護個人主體性，以及為學生、教師與研究人員帶來真實的益處。

2023年1月**ChatGPT**的月活躍用戶達**1億**之際，只有**一個國家**在7月發布了**生成式AI**的使用規範。

# 教育與研究之
# 生成式人工智慧應用指引

# 目 錄

## 表格清單

# 前 言

2022年末，隨著ChatGPT的推出，生成式AI程式在短時間內廣為人知，成為有史以來發展最迅速的應用程式。這類生成式AI應用程式可模仿人類產生文本、圖像、影片、音樂與軟體程式碼等輸出，在世界各地引起轟動。今日，數以百萬計的人們如今在日常生活中使用生成式AI，而將其模型調適應用於特定領域的AI發展，同樣展現出幾乎無限的潛力。

這種廣泛的資訊處理與知識生成能力對教育具有潛在的重大影響，因為這種程式可以複製作為人類學習基礎的高層次思維。隨著生成式AI工具越來越能自動完成基本的寫作與藝術創作，教育領域的政策制定者與機構被迫重新審視人們學習的原因、內容及方式。在數位時代的新階段，這些都是教育領域必須思考的重要問題。

本出版物旨在支持適當的規定、政策與人才培育之計畫，確保生成式AI成為真正有益且可增強教師、學習者與研究人員能力的工具。

本指引提出了政府機關規範生成式AI運用時應採取的關鍵步驟，此外也呈現政策制定與教學設計的框架及具體範例，確保這項技術在教育領域的運用符合道德規範且具有成效。最後，其呼籲國際社會思考生成式 AI 長期而言對人們如何理解知識、定義學習內容、方法和結果，以及如何評估與驗證學習所造成的深刻影響。

本指引以聯合國教科文組織（UNESCO）於2021年所發布的《人工智慧倫理建議書》（Recommendation on the Ethics of Artificial Intelligence）為基礎，將重心放在人文導向的教育理念，促進人類主體性（human agency）、包容性（inclusion）、公平性（equity）、性別平等（gender equality）以及語言與文化多元性（cultural and linguistic diversity），同時保障多元觀點與表達（plural opinions and expressions）。此外，該指引也回應了由國際未來教育委員會（International Commission on the Futures of Education）於2021年《攜手重塑我們的未來：教育社會新契約》（Reimagining our futures together: A new social contract for education）該報告呼籲重新定義我們與科技的關係，將其作為重建教育社會契約的重要一環。

人工智慧絕對不能取代人類智慧，而是促使我們重新思考自己對知識與人類學習的既定認知。我希望這份指南有助於重新定義教育的新視野，並讓大家在進行集體思考與合作行動時有所認知，以邁向以人為本的數位化學習願景。

# 鳴 謝

在聯合國教科文組織教育部門副主任史蒂芬妮雅・吉安尼尼（Stefania Giannini）的領導及未來學習與創新處處長（Future of Learning and Innovation Division）索比・塔維爾（Sobhi Tawil）的指引下，本指南的草擬工作由教育部門之技術與人工智慧組組長苗逢春主導。

特別感謝倫敦大學（University College London）學院副教授韋恩・霍姆斯（Wayne Holmes）參與本指南的草擬工作。

本出版物是人工智慧與教育領域的教育領導者及專家共同努力的成果。倫敦大學學院教授穆特魯・庫庫羅娃（Mutlu Cukurova）；南特大學（Nantes University）聯合國教科文組織開放教育資源教師培訓技術主席柯・德拉・伊格拉（Colin de la Higuera）；約翰尼斯堡大學（University of Johannesburg）助理研究員沙菲卡・伊薩克（Shafika Isaacs）；應用程式發明者基金會（App Inventor Foundation）執行長娜塔莉・勞（Natalie Lao）；上海師範大學副教授倪清；羅馬尼亞歐洲數位教育樞紐（European Digital Education Hub）教育專家部門之資訊與通訊科技組的卡塔琳娜・妮可琳（Catalina Nicolin）；聯合國聯合國教科文組織人工智慧主席、倫敦大學學院計算統計與機器學習系教授約翰・蕭－泰勒（John Shaw-Taylor）；捷特教育服務（Jet Education Services）行政經理凱莉・城平（Kelly Shirohira）；韓國國立教育大學（Korea National University of Education）教授宋基尚（Ki-Sang Song）；芬蘭意義處理有限公司（Meaning Processing Ltd）首席科學家伊卡・圖奧米（Ilkka Tuomi）。

聯合國教科文組織各部門的許多同仁也以各種方式貢獻一己之力，包括生物倫理與科學科技倫理科科長（Section for Bioethics and the Ethics of Science and Technology）達芙娜・法因霍爾茨（Dafna Feinholz）；拉丁美洲及加勒比地區國際高等教育研究所（International Institute for Higher Education）所長法蘭塞斯克・佩德羅（Francesc Pedró）；數位政策與數位轉型科（Section for Digital Policies and Digital Transformation Section）計畫專家普拉提克・希巴爾（Prateek Sibal）；政策與終身學習系統處教師發展科（Section for Teacher Development, Division for Policies and Lifelong Learning Systems）高級專案官員索拉布・羅伊（Saurabh Roy）；曼谷辦事處教育創新與技能發展科（Section for Educational Innovation and Skills Development in the Bangkok Office）教育資訊與傳播技術計畫專家班傑明・維傑爾・德・迪奧斯（Benjamin Vergel De Dios）、文化部門文化表述實體（Cultural Expressions Entity in the Culture Sector）的同仁，以及未來學習與創新處的計畫專家馬克・威斯特（Mark West）。

在此特別感謝格倫·赫特蘭帝（Glen Hertelendy）、露易莎·費拉拉（Luisa Ferrara）、鄭相磊（Xianglei Zheng）、教育科技與人工智慧小組（Unit for Technology and AI in Education）及未來學習與創新處協調本出版物的產出。

另外，我要感謝審稿與校對本指南文字的珍妮·韋伯斯特（Jenny Webster），以及負責排版的陳玉水（Ngoc-Thuy Tran）。

# 縮略詞與簡寫

## 概念與技術

| | |
|---|---|
| **AGI** | 通用人工智慧（Artificial General Intelligence） |
| **AI** | 人工智慧（Artificial Intelligence） |
| **API** | 應用程式介面（Application Programming Interface） |
| **ANN** | 人工神經網路（Artificial Neural Network） |
| **BERT** | 基於變換器的雙向編碼器表示技術（Bidirectional Encoder Representations from Transformers） |
| **DAI** | 分散式人工智慧（Distributed Artificial Intelligence） |
| **GAN** | 生成對抗網路（Generative Adversarial Networks） |
| **GB** | 十億位元組（Gigabytes） |
| **GDPR** | 一般資料保護規定（General Data Protection Regulation） |
| **GenAI** | 生成式人工智慧（Generative Artificial Intelligence） |
| **GPT** | 生成式預訓練轉換器模型（Generative Pre-Trained Transformer） |
| **ICT** | 資訊與通訊科技（Information and Communication Technology） |
| **LaMDA** | 對話程式語言模型（Language Model for Dialogue Applications） |
| **LLM** | 大型語言模型（Large Language Model） |
| **ML** | 機器學習（Machine Learning） |
| **TVET** | 職業技術教育與培訓（Technical and Vocational Education and Training） |
| **VAE** | 變分自編碼器（Variational Autoencoders） |

## 組織機構

| | |
|---|---|
| **AGCC** | 人工智慧政府雲群集（AI Government Cloud Clustere）（新加坡） |
| **CAC** | 中國國家互聯網信息辦公室（Cyberspace Administration of China） |
| **EU** | 歐洲聯盟（European Union） |
| **OECD** | 經濟合作暨發展組織（Organisation for Economic Co-operation and Development） |
| **UNCTAD** | 聯合國貿易與發展會議（United Nations Conference on Trade and Development） |
| **UNESCO** | 聯合國教科文組織（Organisation des Nations unies pour l’éducation，la science et la culture〔UNESCO〕） |

# 序 言

ChatGPT於2022年底問世，是第一個對大眾開放且易於使用的生成式AI工具。此後，其後續的進階版本相繼推出，在全球各地掀起軒然大波，並促使大型科技公司加速投入生成式AI模型的開發競賽。

全球各地的教育界最初擔憂，ChatGPT與類似的生成式AI工具會被學生用來作弊，因而損害學習評量、證照與學歷認證的可信度（Anders, 2023）。雖然有些教育機構禁止使用ChatGPT，也有其他單位更審慎地歡迎生成式AI的到來（Tlili, 2023）。舉例來說，許多中小學及大學採取漸進方式，認為「與其禁止使用，不如支持學生與教職員工有效、合乎倫理且公開透明地使用生成式AI工具」（Russell Group）。這種觀點認為，隨著生成式AI工具越來越普及且日趨成熟，應對教育產生的正面潛力與負面風險同樣值得審慎看待。

事實上，生成式AI可被應用於多種用途。它可以自動處理資訊，並呈現人類思維所有關鍵象徵性表述的輸出。它能夠產出初步成品作為知識半成品並藉此釋放人類從某些低階認知任務中的負擔，新一代的AI工具將可能深刻影響我們對人類智慧與學習的理解方式。

然而，生成式AI也引發了許多關於安全、資料隱私、著作權及操縱等問題的迫切擔憂。其中部分風險來自於人工智慧本身，但隨著生成式AI的興起而日益嚴重；另一些風險則是前所未見，隨著新一代工具出現而浮現出來。我們的當務之急是充分理解並解決這些問題。

本指引旨在回應這項迫切需求。然而，文中針對生成式AI的教育應用所提出的主題式指引，並非意指生成式AI是面對教育基本挑戰的解方。儘管媒體過度渲染，但單憑生成式AI並不足以解決當今教育體系所面對的各項深層問題。面對長期存在的教育問題，關鍵是堅信人類的能力與集體行動才是有效解決社會根本挑戰的決定因素，而非將希望寄託於單一技術。

因此，本指引希望協助制定適當的法規、政策與人才發展計畫，確保生成式AI成為真正有益於教師、學習者與研究人員，成為有助於賦權的教育工具。在本指引以聯合國教科文組織（UNESCO）發布的《人工智慧倫理建議書》為基礎，將重心放在以人為本的教育理念，促進人類的主體性、包容性、公平性、性別平等、文化與語言多樣性，以及多元觀點與表達。

文中首先探討生成式AI是什麼及其運作原理，介紹現有的各種技術與應用模式（第壹節），接著分析關於人工智慧，特別是生成式AI的一系列爭議性議題（第貳節），當試圖以人為本的視角規範生成式AI應用時，應考量的關鍵步驟與核心要素，從而建立一套具備倫理性、公平性與實質意義的規範架構（第參節）。第肆節則提出若干實作措施，協助制定者建構更健全、全面的政策與法律框架，以因應生成式AI在教育與研究領域的發展與應用。第伍節探討利用生成式AI為課程設計、教學、學習及研究活動注入創新實踐的可能性。第陸節在生成式AI對教育與研究的長期影響的考量下，總結了本指引的要旨。

# 壹

# 生成式人工智慧是什麼？它如何運作？

## 一 生成式人工智慧是什麼？

生成式人工智慧（以下稱生成式AI）是一種人工智慧技術，可根據自然語言對話介面的提示自動產生內容。生成式AI並非單純利用既有內容來編輯現有網頁，而是能夠在運用既有資料的基礎上產生新的內容。這些內容的格式包含人類思維的所有符號表徵：以自然語言編寫的文本、圖像（包括照片、數位繪圖與卡通）、影片、音樂與軟體程式碼等。生成式AI的訓練素材是從網頁、社交媒體對話及其他線上媒體蒐集來的資料。它產生內容的方法是，對所攝取的資料中的語詞、像素或其他元素的分布進行統計分析，進而辨識並重現常見的模式，例如哪些詞彙通常接續哪些詞彙，以產出內容。

儘管生成式AI可以產出全新的內容，但它無法產生真正的創新觀點或對現實問題的解決方案，因為它並不理解現實世界中的物件或語言背後所依附的社會關係。此外，儘管生成式AI可產出流暢且令人印象深刻的內容，但其準確性並不可信。事實上，就連ChatGPT的服務供應商也承認，「雖然像ChatGPT這樣的工具往往能產出貌似合理的答案，但其準確性並不可靠。」（OpenAI , 2023）。多數情況下，除非使用者對相關主題有充分的認識，否則不會發現內容中的錯誤。

## 二 生成式人工智慧如何運作？

生成式AI背後的具體技術是名為機器學習（Machine Learning，ML）的AI技術家族的分支，其利用演算法持續且自動地從資料中優化其表現。有一種名為人工神經網路（Artificial Neural Networks，ANNs）的機器學習技術促成了近年來AI領域的諸多進展（譬如AI在臉部辨識方面的應用），而其發明靈感源自於人腦的運作機制及其神經元之間的突觸連結。目前已有多種不同類型的ANNs模型。

文本與圖像的生成式AI科技均以研究人員使用多年的一系列AI技術為基礎。[1]例如，ChatGPT使用生成式預訓練轉換器模型（Generative Pre-trained Transformer，GPT）[2]，而圖像生成式AI一般使用所謂的生成對抗網路（Generative Adversarial Networks，GAN）（見表1）[3]

[1]生成式AI模型可供研究人員與其他相關利益者使用的時間，比ChatGPT要早得多。例如，Google於2015年發表了名為「深度夢境」（Deep Dream）的應用程式。（https://en.wikipedia.org/wiki/DeepDream）

[2]見https://chat.openai.com。

[3]有關AI技術與科技及其關聯，請見聯合國教科文組織出版物（2022b）第8-10頁。

表1 生成式AI使用的技術

| | | |
|---|---|---|
| 機器學習（ML） | 使用資料以自動改善效能的AI。 | |
| 人工神經網路（ANN） | 發明靈感源自於人腦的結構與運作的一種機器學習技術（如神經元之間的突觸連結）。 | |
| 文本生成式AI（Text generative AI） | 通用變換模型（General-purpose Transformers） | 能夠聚焦於資料的不同部分以判斷其相互關聯性的一種人工神經網路。 |
| | 大型語言模型（Large Language Models，LLM） | 由大量文本資料訓練的一種通用變換模型。 |
| | 生成式預訓練變換模型（GPT）[4] | 由更大量的語料預先訓練的大型語言模型，這些資料可使模型捕捉語言的細微差別，產生前後文意連貫的文本。 |
| 圖像生成式AI（Image GenAI） | 生成對抗網路（GAN） | 用以圖像生成的神經網路類型。 |
| | 變分自編碼器（VAE） | |

## 文本生成式AI如何運作？

文本生成式AI使用一種名為「通用變換模型」（General-purpose Transformers）的人工神經網路，以及一種名為「大型語言模型」（Large Language Models，LLM）的通用變換模型，也因此文本生成AI系統通常被稱為大型語言模型（LLMs）。文本生成式AI使用的LLM名為生成式預訓練變換模型（Generative Pre-trained Transformer, GPT），因而有了 "ChatGPT" 中 "GPT" 一詞的由來。

ChatGPT建立在OpenAI開發的GPT-3基礎上，為其GPT的第三次迭代。第一次迭代於2018年發布，最近一次的GPT-4則於2023年3月推出（見表2）。透過人工智慧架構、訓練方法與最佳化技術的進步，OpenAI每次推出的GPT都在前一次的基礎上進行迭代改進。在其持續的進步之中，眾所周知的一個面向是使用越來越多的資料來訓練數量呈指數增長的「參數」。參數就像調整旋鈕一樣，可以微調GPT的效能。其中包括了模型的「權重」——即用來決定模型如何處理輸入與產出結果的數值化參數。

除了在優化人工智慧架構與訓練方法方面的進展之外，這種快速的迭代還得益於海量的資料[5]及大型企業運算能力的提升。自2012年以來，用於訓練生成式AI模型的運算能力每隔三到四個月就會翻倍一次。相較之下，摩爾定律（Moore’s Law）所描述的倍增周期為兩年一次（OpenAI, 2018; Stanford University, 2019）。

[4]請注意，由於目前生成式AI技術仍相對新穎，因此不同的公司在這類術語的使用上往往有所差異，有時雖然使用的詞彙不同，但指的是同一個東西。

[5]令人擔憂的是，用於訓練OpenAIGPT未來迭代的資料將包含先前的GPT版本所生成的大量文本。這種自我參照的循環可能會毒害訓練資料，進而減損未來的GPT模型的能力。

表2 OpenAIGPT

| 模型 | 發表年份 | 訓練資料量 | 參數量 | 特性 |
|---|---|---|---|---|
| GPT-1 | 2018年 | 40 GB | 1.17億 | 執行自然語言處理任務，如文字補全與回答問題等自然語言處理任務。 |
| GPT-2 | 2019年 | 40 GB | 15億 | 完成更複雜的自然語言處理任務，如機器翻譯與摘要生成。 |
| GPT-3 | 2020年 | 17,000 GB | 170億 | 執行高級自然語言處理任務，如編寫連貫的段落與生成整篇文章。亦能以少量範例適應新任務。 |
| GPT-4[6] | 2023年 | 1,000,000 GB（據稱但未經證實） | 170兆（據稱但未經證實） | 可靠性更強，能夠處理更複雜的指令及任務。 |

**GPT完成訓練後，可根據提示生成文本。其步驟如下：**

1.將提示語（prompt）分解成小型單元，稱作詞元（token），並輸入到GPT中。

2.GPT利用統計模式來預測可能出現的詞彙或片語，以形成一個具連貫性的回應。

- ◈GPT可辨識經常一起出現在其預建的大型資料模型中的單詞或詞組模式（其中包含擷取自網路以及其他來源的文本內容）
- ◈有了這些模式，GPT可估算特定單詞或詞組在已知語境中出現的概率。
- ◈GPT在隨機初始預測的基礎上，利用這些估計得出的概率來預測下一個可能出現在回應中的單詞或詞組。

3.將預測的單詞或詞組轉換成可讀文本。

4.此可讀文本會經過所謂的「防護欄」（guardrails）過濾，以排除任何冒犯性內容。

5.重複步驟2至4，直到回應完成。回應完成的標準為達到預設的詞元數上限，或符合預定的停止條件。

6.對回應進行後處理（post-processing），套用格式化、標點符號及其他修飾（例如在回應的開頭加上人類可能使用的詞語，如「當然」、「沒問題」或「抱歉」），來提升其可讀性。

雖然GPT及其自動生成文本的能力自2018年起向研究人員開放，但ChatGPT的推出之所以引起廣泛關注，是因為其免費開放易於使用的介面，這意味著只要有上網的管道，任何人都能探索這項工具。ChatGPT的推出在全球各地造成了衝擊，並引發其他全球性科技企業及眾多新創公司競相追趕，紛紛推出類似的系統，或者在ChatGPT的基礎上開發新工具。

[6]NB OpenAI——開發這張表格所列的GPT模型的公司——尚未公開發布有關GPT-4的詳細資訊（The Verge, 2023a）。事實上，這裡列出的參數量已遭OpenAI的執行長揭穿為不實資訊（The Verge, 2023b）。然而，表中的數據已為數個媒體管道所報導（案例可見E2Analyst, 2023）。無論如何，重點是GPT-4建立在規模更龐大的資料集上，使用的參數量也遠比GPT-3要來得多。

截至2023年7月，ChatGPT的替代工具如下：

- **Alpaca**[7]：針對由史丹佛大學開發、Meta公司所推出的Llama進行微調的版本。此基礎模型旨在解決大型語言模型的虛假資訊、社會成見及有毒語言。
- **Bard**[8]：Google基於其LaMDA與PaLM 2系統開發的一種大型語言模型，可即時存取網際網路，因此可提供最新資訊。
- **Chatsonic**[9]：由Writesonic推出，其開發以ChatGPT為基礎，同時也可直接抓取資料。
- **Ernie（又名文心一言）**[10]：百度推出的雙語大型語言模型，目前仍在開發階段，可大量知識與海量資料集以生成文本與圖像。
- **Hugging Chat**[11]：開發者為HuggingFace。其在整個開發、訓練與部署過程中都強調倫理道德與透明度。此外，用於訓練模型的所有資料都開放來源。
- **Jasper**[12]：一套工具與應用程式介面，經過訓練可依照使用者偏好的風格進行寫作，此外還可生成圖像。
- **Llama**[13]：Meta推出的開源大型語言模型，在測試新方法、驗證他人工作成果及探索新用例時，通常只需較少的運算能力與資源。
- **Open Assistant**[14]：一種開源方法，旨在供任何擁有足夠專業知識的人開發專屬的大型語言模型。其基礎為志願者管理的訓練資料。
- **通義千問**[15]：阿里巴巴推出的一套大型語言模型，可用英語或中文回應提示。目前其被整合至阿里巴巴的商業工具套件內。
- **YouChat**[16]：一種大型語言模型，整合了即時搜尋功能，可提供額外的語境與見解以生成更準確可靠的結果。

這些工具大多都可在一定的限制內免費使用，亦有部分為開源系統。許多其他推出中的產品即以上述大型語言模型（LLMs）的其中之一為基礎。例如：

---

[7]https://crfm.stanford.edu/2023/03/13/alpaca.html
[8]https://bard.google.com
[9]https://writesonic.com/chat
[10]https://yiyan.baidu.com/welcome
[11]https://huggingface.co/chat
[12]https://www.jasper.ai
[13]hhttps://ai.facebook.com/blog/large-language-model-llama-meta-ai
[14]https://open-assistant.io
[15]https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users
[16]https://you.com

◉ **ChatPDF**[17]：總結並回答有關已提交PDF文件的問題。

◉ **Elici：人工智慧研究助理** [18]：旨在自動化研究人員的部分工作流程、 辨識相關論文並總結關鍵資訊。

◉ **Perplexity**[19]：提供一個「知識中心」，讓人們得以尋求快速且符合需求的準確答案。

同樣地，以大型語言模型（LLMs）為基礎的工具也被嵌入其它應用中，譬如網頁瀏覽器。 舉例來說，建立在ChatGPT之上的Chrome瀏覽器之擴充功能包括：

◉ **WebChatGPT**[20]：為ChatGPT提供網路通道，以實現更準確且資訊更新的對話。

◉ **Compose AI**[21]：在電子郵件及其他應用程式中自動完成語句。

◉ **TeamSmart AI**[22]：提供「虛擬助理團隊」。

◉ **Wiseone**[23]：簡化線上資訊。

除此之外，ChatGPT整合了部份搜尋引擎[24]，並應用於許多生產力工具組合中（如Microsoft Word及Excel），進而在世界各地的辦公場所與教育機構中更加普及（Murphy Kelly，2023）。

最後，作為從文字生成式AI過渡到圖像生成式AI（Image GenAI）的有趣轉折，OpenAI近期的最新版本GPT-4能夠在提示中同時接收圖像與文本。就此而言，它具有多模態（multimodel）功能。因此，有些人認為「大型語言模型」（LLM）這個名稱已不再合適，而這也是史丹佛大學的研究人員提出「基礎模型」（foundation model）一詞的原因之一（Bommasani et al.，2021）。此術語尚未獲得廣泛採用。

[17]https://www.chatpdf.com
[18]https://elicit.org
[19]https://www.perplexity.ai
[20]https://tools.zmo.ai/webchatgpt
[21]https://www.compose.ai
[22]https://www.teamsmart.ai
[23]https://wiseone.io
[24]https://www.microsoft.com/en-us/bing

## 圖像生成式AI如何運作？

圖像生成式人工智慧與音樂生成式人工智慧（Music GenAI）通常使用一種不同類型的人工神經網路（ANN），稱為生成對抗網路（GANs）。這類網路也可以與變分自編碼器（Variational Autoencoders, VAEs）互相結合使用。GANs包含兩個部分（即兩個「對抗者」）：「生成器」（generator）與「鑑別器」（discriminator）。以圖像GANs為例，生成器會根據提示語隨機生成一張圖像，而鑑別器則試圖分辨這張圖像是真實影像還是生成的圖像。生成器再根據鑑別器的判斷結果來調整自身的參數，進而生成下一張圖像。這個過程會反覆進行，可能重複數千次，生成器產出的圖像越來越逼真，以至於鑑別器越來越難區分其與真實圖像的差異。舉例來說，一個成功的GAN若以數千張風景照片作為訓練資料，它可能會生成出許多全新但非真實存在、卻幾乎與真實照片難以區分的風景圖像。同樣地，若以流行音樂（或甚至是單一音樂家的作品）作為訓練資料，GAN也可能創造出結構和複雜度與原始音樂相似的新作品。

截至2023年7月，常見的圖像生成式AI模型包含以下幾種，而它們皆可根據提示語生成圖像。其中大多數在一定限制下均可免費使用：

◉ **Craiyon**[25]：過去名為DALL·E mini。

◉ **DALL·E 2**[26]：OpenAI推出的圖像生成式AI工具。

◉ **DreamStudio**[27]：Stable Diffusion模型的圖像生成式AI工具。

◉ **Fotor**[28]：將生成式AI納入一系列圖像編輯工具。

◉ **Midjourney**[29]：獨立的圖像生成式AI工具。

◉ **NightCafe**[30]：Stable Diffusion與DALL·E 2的介面。

◉ **Photosonic**[31]：WriteSonic平台使用的AI藝術生成網路。

易於操作的影片生成式AI工具如下：

◉ **Elai**[32]：可將簡報、網站及文本轉換成影片。

◉ **GliaCloud**[33]：可根據新聞內容、社交媒體貼文、直播體育賽事與統計資料生成影片。

---

[25] https://www.microsoft.com/en-us/bing
[26] https://openai.com/product/dall-e-2
[27] https://dream.ai/create
[28] https://www.fotor.com/features/ai-image-generator
[29] https://www.midjourney.com
[30] https://creator.nightcafe.studio
[31] https://writesonic.com/photosonic-ai-art-generator
[32] https://elai.io
[33] https://www.gliacloud.com

◉ Pictory[34]：可根據長篇內容自動生成短片。

◉ Runway[35]：提供一系列影片（與圖像）生成與編輯工具。

最後，一些易於使用的音樂生成式AI工具如下：

◉ Aiva[36]：可自動創作個性化配樂。

◉ Boomy[37]、Soundraw[38]與Voicemod[39]：可根據任何文本生成歌曲，使用者無需具備作曲知識也能上手。

## 三 透過提示語工程生成預期獲得的輸出

雖然使用生成式AI可以像輸入問題或其他提示一樣簡單，但現實情況是，使用者仍然無法直接獲得他們希望得到的精確輸出。例如，近期在美國科羅拉多州博覽會（Colorado State Fair）上獲獎的突破性人工智慧圖像《太空歌劇院》（Théâtre D'opéra Spatial），就花費了數週編寫提示語及微調數百張圖像才生成了最後的參賽作品（Roose, 2022）。這種類似為文本生成式AI編寫有效提示語的挑戰，促使徵才網站上出現越來越多的提示工程師（prompt-engineer）的職位（Popli, 2023）。「提示語工程（prompt-engineering）」意指編寫輸入的過程與技術，目的是生成更接近使用者預期獲得的內容。

當提示語能清晰傳達一條有邏輯的推理鏈時，提示詞工程的效果最佳。具體建議如下：

◉ 使用簡單、清楚且直白的語言，避免使用複雜或模稜兩可的措辭。

◉ 舉例說明希望獲得的生成回應或格式。

◉ 涵蓋上下文資訊，這對生成相關且有意義的內容至關重要。

◉ 必要時精簡提示並反覆進行互動，嘗試不同的變化以尋求最佳成效。

◉ 遵守倫理規範，避免使用可能產生不恰當、偏頗或有害內容的提示語。

同時，我們也應該認知到，生成式AI的輸出不可在未經批判性評估的情況下被完全信賴。正如OpenAI在介紹其最先進的GPT時所述[40]：

---

[34]https://pictory.ai

[35]https://runwayml.com

[36]https://www.aiva.ai

[37]https://boomy.com

[38] https://soundraw.io

[39] https://www.voicemod.net/text-to-song

[40]https://openai.com/research/gpt-4

> GPT-4儘管具備高度功能，仍具有與之前推出的GPT模型相似的局限性。最重要的是，它仍然不完全可靠（會「幻化」事實，推理時也會出錯）。採用語言模型的輸出結果時應小心謹慎，尤其是在高風險的情況下，對此，配合特定使用情境的需求來採取對應的程序（如人工審查、補充上下文資料，或者完全避免高風險的使用情況）。

考量生成式AI輸出的品質，在驗證這些工具是否適合大規模應用或高風險任務之前，應進行嚴謹的使用者測試與效能評估。此類評估應依任務類型設計最適切的效能指標。例如，解決數學問題時，可將「正確率」（accuracy）作為主要指標，以衡量生成式AI工具產生正確答案的頻率；回應敏感問題時，可將「回答率」（answer rate）作為衡量性能的主要指標（生成式AI直接回答問題的頻率）；針對程式碼的生成，衡量的標準可以是「可執行率」（fraction of generated codes that are directly executable），即評估生成的程式碼是否可在程式設計環境中直接執行並通過單元測試；針對視覺推理任務，衡量的標準可以是「精確比對」（exact match），即評估生成的視覺物件是否完全符合模型的正確答案（Chen, Zaharia, and Zou, 2023）。

總結而言，從表面上來看，生成式AI的操作似乎簡單易用；但若需生成更複雜與高品質的輸出，仍需仰賴具專業素養的人工輸入與審慎的批判性評估。

### 對教育與研究的影響

雖然生成式AI可幫助教師與研究人員生成實用的文本與其他輸出以協助他們的工作，但這並非是個直接簡單的過程。使用者可能需要多次反覆嘗試提示，才能獲得理想的輸出。另一項潛在隱憂是，年輕的學習者由於專業度不如教師，可能會在不知情與缺乏嚴謹判斷的情況下接受流於表面、不精確甚至是有害的生成式AI輸出內容。

## 四 新興的EdGPT及其影響

有鑒於生成式AI模型可作為開發更專業或特定領域模型的基礎或起點，一些研究人員建議將GPT更名為「基礎模型」（foundation models）（Bommasani et al., 2021）。在教育領域中，開發者與研究人員已開始對基礎模型進行微調，發展出所謂的「教育用GPT」（EdGPT）。EdGPT模型使用特定資料來訓練以達到教育目的。換言之，EdGPT的目標是將來自龐大訓練資料的模型細緻化，以服務教育領域的特殊需求。

這或許可以讓EdGPT在更廣泛的範圍內協助達成4.3節所提及的教育轉型目標。例如，以課程共同設計為目標的EdGPT模型協助教育者與學習者產出適當的教材，譬如與有效的教學方法、特定的課程目標及特定學習者的挑戰程度密切相關的教案、測驗與互動性活動。同樣地，在一對一語言技能指導的背景下，可以使用適合特定語言的文本來改良基礎模型，以生成示範句子、段落或對話供練習。模型與學習者進行互動時，可以根據學習者的程度提供語法準確的相關文本。理論上，相較於標準的GPT，EdGPT模型的輸出也可能包含較少的偏誤或其他令人不快的內容，但仍可能產生錯誤。值得注意的是，除非生成式AI的基礎模型與訓練方法出現重大變化，否則EdGPT仍可能生成錯誤內容，並在其他方面受到限制，譬如對教案或教學策略的建議。因此，EdGPT的主要使用者仍需從批判的角度看待任何輸出，尤其是教師與學生。

目前，為了讓GPT的使用在教育領域中更具目標性而改良基礎模型的工作，尚處於早期階段。現有的例子包括華東師範大學（East China Normal University）針對教學與學習服務所開發的基礎模型EduChat，其程式碼、資料與參數均作為開放來源共享。[42]另一個例子是好未來教育集團（TAL Education Group）開發的MathGPT——一種大型語言模型，聚焦於向世界各地的使用者開放、與數學相關的解題與教學應用。[43]

然而，在取得重大進展之前，我們必須努力改良基礎模型，不僅應補充學科知識與消除偏見，還應增加相關學習方法、以及如何透過演算法與模型設計反映這些方法的知識。我們面臨的挑戰是，如何判斷EdGPT模型能否超越學科知識的範疇，並同時以學生為中心的教學法與正向的師生互動為目標。進一步的挑戰是，確定在何種程度上可以合乎道德地蒐集並使用學習者與教師的資料，以進一步訓練EdGPT。最後，我們還需要進行嚴謹的研究，以確保教育指導方案既不會危害學生的人權，也不會削弱教師的專業自主性。

[42]https://www.educhat.top

[43]https://www.mathgpt.com

# 貳

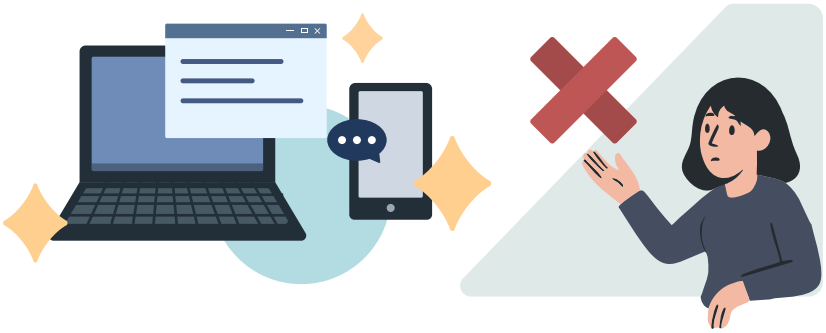
# 生成式人工智慧的爭議及其對教育的影響

先前討論了生成式AI（GenAI）的定義及其運作方式，本節將探討由生成式AI系統引起的爭議與道德風險，以及其對教育可能產生的影響。

## 一 數位貧窮的惡化

如前所述，除了在人工智慧架構與訓練方法上的迭代創新之外，生成式AI還仰賴大量的資料與強大的運算能力，而這些一般只有大型國際科技企業與少數經濟體才能獲得（主要為美國、中國，其次是歐洲）。這意味著大多數的企業與國家——尤其是全球南方（the Global South）——都無法取得、建立與控制生成式AI技術。

由於獲取資料的管道對國家的經濟發展與個人的數位機會日益重要，那些無法獲取或是沒有財力可負擔充分資料的國家與人民就會陷入「數位貧窮」的處境（Marwala，2023年），同樣情況亦適用於運算能力的獲取。生成式AI在技術先進的國家與地區迅速普及，使資料生成與處理的速度呈指數型增加，同時讓人工智慧財富更加集中於全球北方（Global North）。如此導致的直接後果是，資料匱乏的地區遭到進一步排除，並面臨長期風險，包括受限於GPT模型中嵌入的價值與標準，甚至可能造成文化與教育內容的殖民化。目前的ChatGPT模型利用反映全球北方的價值觀與規範的線上使用者資料來進行訓練，因此，這些模型對於全球南方或弱勢族群而言，並不適用於在地化的AI應用情境。



**對教育與研究的影響**

研究人員、教師與學習者應該對生成式AI訓練模型中蘊含的價值取向、文化標準與社會習俗抱持批判性觀點。政策制定者應該要意識到，生成式AI的訓練與控制方面不斷擴大的歧異導致了不公平的現象日益惡化，並採取行動來解決這個問題。

## 二 超越國家監管的發展速度

主要的生成式AI供應商也因為拒絕讓自家產品接受嚴謹的獨立學術審查而遭到批評（Dwivedi et al.,2023）。[44]生成式AI系統的基礎技術往往是一家公司的商業智慧財產，並因而受到保護。與此同時，許多開始使用生成式AI的公司發現，系統安全的維護是一項越來越困難的挑戰（Lin, 2023）。此外，儘管人工智慧這個產業本身需要監管[45]，但關於所有人工智慧——包括生成式AI在內——的創造與運用的法規制定，往往追不上其快速發展的腳步。這部分說明了國家或地方機關在理解與管理法律與倫理議題上所面臨的挑戰（Chen, Zaharia, and Zou, 2023）。[46]

雖然生成式AI可增進人類完成特定任務的能力，但對提倡生成式AI技術的公司而言，民主監督機制的約束力卻相當有限。這凸顯了監管方面的問題，尤其是國內資料的獲取與使用——包括地方機構與個人的資料，以及在國內生成的資料。地方政府機構必須制定適當的法規，才能對生成式AI技術浪潮進行有效監管，確保將其作為公共財而非僅為私有資源。

**對教育與研究的影響**

研究人員、教師與學習者必須意識到，國內機構與個人的所有權、以及生成式AI國內使用者的權利缺乏適當的法規保護，並且回應生成式AI所引發的立法議題。

## 三 未經同意的內容使用

如先前所述，生成式AI模型由大量資料（如文本、聲音、程式碼與圖像）訓練而成，這些資料通常來自網際網路，而且往往未經任何所有者的許可。因此，許多圖像與程式碼生成式AI系統被控侵犯智慧財產權。在筆者撰寫這本指南時，已有數起相關的國際法律案件正在審理中。

此外，也有觀點指出，GPT可能違反歐盟的（2016）《一般資料保護規則》（General Data Protection Regulation，GDPR）等法規，尤其是其中關於「被遺忘權」（the right to be forgotten）的條款，因為一旦資料被用來訓練模型，就幾乎無法再刪除或移除該資料所衍生的結果。

[44]有幾個例外，像是致力於開源人工智慧發展的「Hugging Face」公司。

[45]見Google（2023a）與OpenAI（Bass and Metz, 2023）的呼籲。

[46]關於人工智慧監管的專案，見歐盟執委會（European Commission）公布的《人工智慧法》（2021）。

**對教育與研究的影響**

◈研究人員、教師與學習者必須瞭解資料所有者的權利，且應該檢查現正使用的生成式AI工具是否有違反任何現行法規。

◈研究人員、教師與學習者也應知悉，使用生成式AI創作的圖像或程式碼可能會侵犯他人的智慧財產權，以及他們在網際網路上創造與分享的圖像、聲音或程式碼可能會被其他生成式AI模型擷取並再利用。

## 四 用於生成輸出的不可解釋性模型

長久以來，人工神經網路（ANNs）通常被認為是「黑箱模型」，意即其內部運作並不開放審視。因此，人工神經網路既不具「透明性」（transparent）或「可解釋性」（explainable），無法得知其如何做出特定輸出結果。

雖然整體方法（包括所使用的演算法）在理論上可被說明，但特定的模型及其參數（包括模型的權重）卻無法審查，而這正是無法解釋特定輸出為何生成的原因。在像GPT-4這樣的模型中，有高達數十億個參數／權重（見表2），而正是這些權重共同保存了模型用來生成輸出的學習模式。由於參數／權重在人工神經網路中並不透明（表1），因此我們無法解釋這些模型究竟如何創造特定的輸出。隨著生成式AI日益複雜（見表2），生成式AI缺乏透明度與可解釋性的問題也越來越棘手，經常導致讓人意想不到或不樂見的結果。此外，生成式AI模型承襲了其訓練資料存有的偏差，而有鑒於模型的不透明性，使得偏誤難以察覺與修正。最後，這種不透明性也是與生成式AI有關的信任問題的一個關鍵肇因（Nazaretsky et al., 2022a）。如果使用者不瞭解生成式AI系統如何得出特定的輸出，也更可能對該系統採取保留態度（Nazaretsky et al., 2022b）。

**對教育與研究的影響**

研究人員、教師與學習者必須意識到，生成式AI系統的運作類似黑箱，因此即使有可能知道特定內容的創造原因，也很難找到答案。若不解釋輸出是如何生成的，使用者往往會受限於生成式AI系統中設計的參數所定義的邏輯之中。這些參數可能反映了特定的文化或商業價值與規範，導致生成的內容隱含偏見。

## 五 人工智慧生成的內容汙染網際網路

由於GPT的訓練資料通常來自網際網路，而網路上經常可見歧視性語言與其他令人無法接受的語言，因此開發人員不得不採取所謂的「防護機制」（guardrails），以防止GPT的輸出帶有攻擊性或有違倫理規範。然而，由於缺乏嚴格的法規與有效的監督機制，有越來越多由生成式AI產生的偏見性資料透過網際網路傳播，汙染了全球各地多數學習者的主要內容或知識來源之一。此問題尤為關鍵，因為生成式AI產生的材料看似準確與令人信服，但往往包含錯誤與帶有偏見的觀點。對於缺乏背景知識的年輕學習者而言，這構成了極高的風險，容易導致誤信與誤用。此外，這也對未來的GPT模型造成了不斷循環遞迴的風險，因為它們將根據網路爬蟲資料進行訓練，而這些文本正是GPT模型自己創造的，其中自然也包含了它們的偏見與錯誤。

**對教育與研究的影響**

- ◈研究人員、教師與學習者必須意識到，生成式AI系統可能輸出帶有攻擊性與不合乎倫理道德的內容。
- ◈他們也必須瞭解，如果未來的GPT模型以先前的GPT模型所生成的文本為基礎，知識的可靠性可能會造成長期的影響。

## 六 缺乏對現實世界的瞭解

文本GPTs有時被貶稱為「隨機鸚鵡」（Stochastic Parrots），因為如前所述，雖然它們能生成看似令人信服的文本，但這些文本往往包含錯誤，並可能包含有害的陳述（Bender et al., 2021年）。之所以會出現這種情況，是因為GPT只會重複訓練資料中的模式（通常是來自網際網路的文本），從隨機模式開始，而非真正理解語言或其所代表的現實意涵，就像鸚鵡會模仿聲音，但實際上並不知道自己在說什麼一樣。

生成式AI模型「貌似」理解自己所使用與生成的文本，但「實際上」它們並不瞭解這些語言與現實世界的斷裂，會導致教師與學生對輸出結果賦予其過高的信任。這種情況對未來的教育造成了嚴重的風險。事實上，生成式AI並未參考現實世界的觀察或科學方法的其他重要方面，與人類或社會的價值觀也不一致。因此，它無法產出真正有意義的、反映現實的知識內容。關於生成式AI模型所生成、貌似新穎的內容能否被認定為科學知識，尚存在爭議。

如先前所述，GPTs經常會產生不準確或不可靠的文本。事實上，眾所周知的是，GPTs會編造現實生活中不存在的東西。有人將這種現象稱為「幻覺」（hallucination），儘管部分學者批評該詞具擬人化誤導之虞。OpenAI也承認這一點。例如，ChatGPT開介面的底部即載明：「ChatGPT有可能生成關於人、地方或事實的錯誤資訊。」

也有少數倡導者主張，生成式人工智慧（Gen AI）是邁向通用人工智慧（AGI）（指一類超越人類的人工智慧）的重要一步。然而，此觀點長期以來備受批評。批評者認為人工智慧至少必須在某種程度上結合知識基礎的人工智慧（也稱為符號式或規則式AI）和資料驅動的人工智慧（也就是機器學習），形成兩者的共生關係（Marcus, 2022）。對於AGI或人工智慧具備感知能力的說法，也可能使我們忽略對人工智慧目前造成的傷害進行更審慎的思考，例如原本已受到歧視的群體將進一步遭受隱性歧視（Metz, 2021）。

**對教育與研究的影響**

- 文本生成式AI的輸出結果有可能看起來與人類產生的內容非常相似，彷彿它能夠理解自己生成的文本。然而，生成式AI並不理解任何東西。相反地，這些工具會透過網路上常見的方式將單詞串聯起來，而其生成的文本也可能是錯的。
- 研究人員、教師與學習者必須意識到，GPT並不理解自己所生成的文本，它有可能、而且經常生成不正確的語句，因此他們必須以批判性的角度去看待那些內容。

## 七 限縮多元觀點並進一步邊緣化弱勢聲音

ChatGPT與其他類似的工具傾向只輸出標準化答案，而這些答案呈現了模型訓練資料的所有者／創作者之價值觀。事實上，如果訓練資料中有一連串的詞語頻繁出現（如同那些源自主流或非爭議性來源的詞句），GPT便很有可能在輸出中重複使用並強化這些觀點。

這可能會限制與妨害多元觀點與多元思想表述的發展。資料貧乏的人口（包括北方世界的邊緣化群體）在網路上的數位化存在極度隱微或有限。因此，他們的聲音沒有被聽到，他們的顧慮也沒有表現在用於訓練GPT的資料中，因此很少見於輸出的結果。基於這些原因，考量以公開網站與社交媒體對話資料為基礎的預訓練方法，GPT模型可能會進一步邊緣化已經處於弱勢地位的群體。

## 八 生成更具威脅性的深度偽造內容

除了所有生成式AI常見的爭議之外，採用GAN的生成式AI還可用於修改或操縱既有的圖像或影片，以生成難辨真偽的假圖像或影片。生成式AI使這類「深度偽造」（deepfake）與所謂「假新聞」的創造變得越來越容易。換句話說，生成式AI讓特定群體更容易做出不道德與甚至犯罪的舉動，譬如散播假資訊（disinformation）、煽動仇恨言論（hate speech），以及在人們不知情或未同意的情況下將他們的面孔合成至虛假、甚至損害名譽的影片中。


**對教育與研究的影響**

生成式AI供應商有義務保護使用者的著作權與肖像權，但研究人員、教師與學習者也必須意識到，他們在網際網路上分享的任何圖像都有可能被納入生成式AI的訓練資料，並透過不道德的方式進行操縱與濫用。

# 參

# 規範生成式AI在教育領域的應用

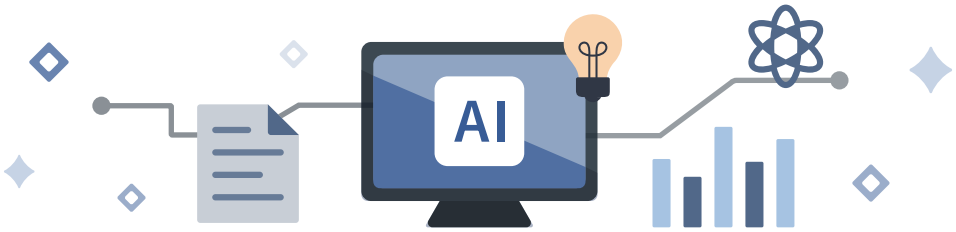
為了解決生成式AI的爭議並發揮其在教育領域的潛在優勢，首先需要為其制定適當的規範。若想規範以教育為目的之生成式AI，我們必須採取一些以人為本（human-centered）的步驟與政策措施，確保其使用合乎倫理道德、安全、公平且具有意義。

## 一 從人本角度探討生成式AI

2021年發布的《人工智慧倫理建議書》（The 2021 Recommendation on the Ethics of Artificial Intelligence）指明解決生成式AI的各種爭議時所需依循的規範性框架，包含有關教育與研究領域的議題。這份建議書從人本角度去看待人工智慧，主張人工智慧的應用目的應是開發人類能力以實現包容、公正與永續的未來。這個取向必須遵循人權原則，保護人類的尊嚴與定義知識共同體的文化多元性。在治理方面，以人為本的方法需要進行適當的管制，來確保人類自主性、透明度與公共問責制。

2019年發布的《人工智慧與教育之北京共識》（Beijing Consensus on Artificial Intelligence [AI] and Education）進一步闡述，以人為本的方法對人工智慧的教育應用具有何種意義。這份文件明確指出，人工智慧技術的教育應用應幫助人類實現永續發展，並在生活、學習與工作方面達到有效的人機協作。其也呼籲大眾進一步確保人工智慧的公平取用，以幫助弱勢群體及解決不平等問題，同時促進語言與文化的多元性。最後，其建議社會採「全政府」、「跨部門」與「多利害關係人參與」的治理方式來推動AI教育政策。

《人工智慧與教育：政策制定者指南》（AI and education : guidance for policy-makers，UNESCO，2022b）進一步探討了在審查人工智慧對教育的益處與風險、以及教育在發展人工智慧能力方面所扮演的角色時，人本方法代表了什麼意義。其提出了制定政策的具體建議，指引人們應使用人工智慧來：(i) 實現學習機會的兼容性，尤其是針對身心障礙者與弱勢族群；(ii) 支持個人化與開放的學習選擇；(iii) 擴展數據基礎架構與管理系統，以擴大學習機會與提升學習品質；(iv) 監督學習過程，並提醒教師識別潛在風險；(v) 培養對人工智慧的理解，以及能夠以合乎倫理道德與具有意義的方式運用人工智慧的能力。



## 二 規範生成式AI教育應用時應採取的步驟

在ChatGPT問世之前，各國政府已著手制定或調整用於規範資料蒐集與使用的框架，以及教育等產業採用人工智慧系統的方式，而這為新興人工智慧應用的規範創造了立法與政策背景。在2022年11月起多個競爭性的生成式AI模型發布之後，各國政府採取了不同的因應措施，例如：限制生成式AI使用、修訂監管架構，甚至加速立法程序。

2023年4月，各國政府監管與促進生成式AI創意運用的策略經過了詳細的規劃與審查（UNESCO, 2023b）。[47] 該審查報告建議政府機關可採取七項策略性步驟來規範生成式AI並重新確立公共控制，以在教育等各領域發揮人工智慧的潛能。

### 第1步：簽署國際性或區域性《一般資料保護條例》（General Data Protection Regulations，GDPR）或制定全國性《一般資料保護條例》

生成式AI模型的訓練牽涉了許多國家公民網路資料的蒐集與處理。生成式AI模型在未經同意的情況下使用資料與內容的做法，對資料保護與數位主權的議題構成了進一步的挑戰。

歐盟於2018年頒布的《一般資料保護條例（GDPR）》是這方面的先例之一，為生成式AI供應商在個人資料的蒐集與處理上所受到的規範建立了必要的法律框架。聯合國貿易與發展會議（UNCTAD）設立的「全球資料保護與隱私立法世界線入口網站」（Data Protection and Privacy Legislation Worldline）指出，194個國家中已有137國制定了保障資料保護與隱私的法規。[48]

然而，這類法律框架在這些國家中實施的程度依然混沌不明。因此，確保這些框架有適當實施，顯得比以往更加重要，包括定期監控生成式AI系統的運作。此外，那些尚未制定一般資料保護條例的國家也需要加緊立法進程，以彌補法制上的空缺。

### 第2步：採用／修訂並資助整體整合政府運用人工智慧的策略

生成式AI的規範必須是廣泛的國家人工智慧策略的一部分，這些策略可確保各發展產業安全與公平地運用人工智慧，包括教育領域在內。而國家人工智慧策略的制定、簽署、資助與施行，需要透過「整合式治理」的方式來進行。如此一來，不同產業之間才能協調整合應對新興挑戰所需採取的行動。

---

[47] 該審查所依據的資料，來自聯合國教科文組織向其193個成員國展開的政府在教育領域中使用人工智慧之情況的調查報告（UNESCO, 2023c）、經濟合作暨發展組織（Organisation for Economic Co-operation and Development，OECD）之AI政策觀察站（AI Policy Observatory）、史丹佛大學公布的人工智慧指數報告（Stanford University, 2023），以及出自一群國際專家的第一手資料。

[48] https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

到了2023年初，約有67個國家[49]已制定或計劃制定全國性人工智慧策略，其中有61國採取獨立人工智慧策略的形式，7國將人工智慧納入更廣泛的全國性資訊與通信技術（ICT）或數位化策略。不難理解的是，有鑒於人工智慧屬於新興技術，這些國家策略在制定時尚未將生成式AI視為一項具體議題。

各國必須修訂既有的全國性人工智慧策略或制定新方針，設立相關規範以確保各產業（包括教育領域在內）在運用人工智慧時合乎倫理道德。

**第3步：確立與施行人工智慧倫理的專門法規**

為了解決人工智慧的運用所引發的倫理議題，有必要制定明確的倫理規範與法律。

聯合國教科文組織（UNESCO）在2023年對既有的國家人工智慧策略進行審查，報告指出，只有約40個國家的人工智慧策略提到這類的倫理議題並制定了指導原則。[50]即便如此，這些道德原則仍需轉換成可強制執行的法律或規範。這種情況相當少見。事實上，只有約20個國家針對教育相關的人工智慧倫理制定了明確的法規，無論作為全國性人工智慧策略的一部分或其他原則。有趣的是，雖然約有45個國家的人工智慧策略強調教育是一個政策領域[51]，但他們提及教育時，大多著墨於支持國家競爭力所需的技能培養與人才發展，而較少談到倫理方面的議題。

那些尚未制定人工智慧倫理規範的國家必須盡速制定並施行這些法規。

**第4步：調整或執行現有著作權法，以規範人工智慧生成的內容**

生成式AI的使用日益普及，為著作權帶來了前所未有的新挑戰，這不但牽涉了用於訓練模型的著作權內容或作品，也涉及模型所產生的「非人為知識輸出」的法律地位。目前，只有中國、歐盟成員國與美國因應生成式AI的影響而修改了著作權法的規定。

例如，美國著作權局（U.S. Copyright Office）認為「著作權的保護對象僅限出自人類創造力的素材」（U.S. Copyright Office, 2023），裁定ChatGPT等生成式AI系統的輸出不受美國著作權法所保護。在歐盟，《人工智慧法》（AI Act）的提案要求人工智慧工具的開發者揭露他們在建構系統時所使用的著作資料來源（European Commission,2021）。

---

[49]截至2023年4月，下列國家已發布了關於人工智慧的全國性策略：阿根廷、澳洲、奧地利、比利時、貝南（Benin）、巴西、加拿大、保加利亞、智利、中國、哥倫比亞、塞普勒斯、捷克、丹麥、埃及、愛沙尼亞、芬蘭、法國、德國、匈牙利、冰島、印度、印尼、愛爾蘭、義大利、日本、約旦、拉脫維亞、立陶宛、盧森堡、馬來西亞、馬爾他、模里西斯、墨西哥、荷蘭、挪威、紐西蘭、阿曼、秘魯、波蘭、葡萄牙、菲律賓、卡達、大韓民國、羅馬尼亞、俄羅斯聯邦、沙烏地阿拉伯、塞爾維亞、新加坡、斯洛維尼亞、西班牙、瑞典、泰國、土耳其、突尼西亞、阿拉伯聯合大公國、英國、美國、烏拉圭及越南。除此之外，一些國家已將人工智慧策略納入更廣泛的資訊通信技術或數位策略，包括阿爾及利亞、波紮那、哈薩克、肯亞、獅子山、斯洛伐克、瑞士與烏干達。

[50]根據一項針對所有全國性人工智慧策略（UNESCO, 2023b）的快速審查，有超過40項策略編有倫理議題的專屬章節。

[51]根據一項針對所有全國性人工智慧策略（UNESCO, 2023b）的快速審查，有超過45項策略編有教育議題的專屬章節。

經由2023年7月發布的生成式AI條例，中國規定生成式AI的輸出內容必須加上人工智慧生成的標記，明確將其視為AI生成內容，並予以特別規範。

為用於訓練模型的著作權素材的使用制定規範、以及定義生成式AI模型輸出內容的著作權地位，逐漸成為著作權法的新責任。各國的當務之急是因應這一點來調整現行法律。

**第5步：詳盡制定生成式AI的規範框架**

人工智慧技術的快速發展正迫使國家／地方治理機構加快更新法規的腳步。截至2023年7月，只有中國發布了有關生成式AI的具體官方法規。2023年7月13日公布的《生成式AI服務治理暫行條例》(Provisional Regulations on Governing the Service of Generative AI，Cyberspace Administration of China, 2023a)要求生成式AI系統供應商根據現行的《網路資訊服務框架下深度合成之規範》(Regulation on Deep Synthesis in the Framework of Online Information Services)為人工智慧生成的內容、圖片與影片加上適當且合法的標記。更多這類的全國性生成式AI框架，必須建立在針對現行地方條例與法規間差異所進行的評估上。

**第6步：培養在教育與研究領域中正確使用生成式AI的能力**

學校與其他教育機構需要培養理解人工智慧(包括生成式AI)對教育有哪些潛在益處與風險的能力。唯有基於這種理解，他們才能證明人工智慧工具的採用有發揮成效。此外，需要通過培訓與持續輔導等方式，這些機構也需要支持教師與研究人員增進正確使用生成式AI的能力。一些國家已經啟動了相關的能力建構計畫，其中，新加坡為教育機構的人工智慧能力發展打造了人工智慧政府雲群集(AI Government Cloud Cluster)這個平台，協助教師與教育機構建立應對 AI 技術的制度化機制，其中包含了一個GPT模型庫(Ocampo, 2023)。

**第7步：反思生成式AI對教育與研究領域的長期影響**

目前推出的生成式AI才剛開始發揮影響力，而它們對教育領域造成的影響還有待充分探索與理解。與此同時，功能更強大的生成式AI與其他種類的人工智慧仍持續開發與有效運用中。然而，關於生成式AI對知識的創造、傳播與驗證(對教導與學習、課程設計與評估，以及研究與著作權)有何長期影響，仍存在一些關鍵性的問題。大多數的國家正處於在教育領域採用生成式AI的早期階段，其長期影響仍有待探究。為了確保人工智慧的使用符合人本精神，我們應即刻就其長期影響推動公開對話與政策討論。涉及政府、私營部門與其他合作夥伴的兼容性辯論，則應該為法規與政策的迭代更新提供見解與建議。

# 三 生成式AI的規範：關鍵要素

所有國家都需要適當規範生成式AI，確保其有利於教育與其他脈絡的正向效益。本節提出了針對其關鍵要素可採取的行動，對象包含：（1）政府監管機構（2）生成式AI系統的供應商（3）教育與研究機構的使用者（4）個人使用者。雖然此框架中有許多要素都具有跨國適用性，但所有關鍵要素都應該根據當地背景進行考量，也就是國家現有的教育體制與既有的整體規範框架。

## 政府監管機關

協調生成式AI法規的設計、校準與施行時，需要從整合式政府機制（whole-of-government approach）的角度出發。建議關注以下七項關鍵要素與行動：

- **跨部門協調：**成立一個國家層級機構，以領導生成式 AI 的全政府協作機制，並統籌各部門之間的協調合作。
- **法規一致性：**確保生成式AI框架與相關的立法與監管背景保持一致，譬如一般資料保護法、網際網路安全規範、資料安全之法規，以及其他相關法規與慣例。評估現行法規的適當性，以及因應生成式AI所引發的新挑戰而必須進行的調整。
- **規範與創新之平衡：**促進企業、產業治理組織、教育與研究機構及相關公共機構之間的跨部門合作，以共同開發值得信賴的模型；鼓勵開放來源生態系統的建置，提倡高效能運算資源與優質預訓練資料集的共享；促進生成式AI在各產業的實際應用，並創造有助大眾福祉的優質內容。
- **AI潛在風險的評估與分類：**生成式AI服務在部署前與系統之生命週期中的有效性、安全與防護制定評估及分類的原則與流程。根據生成式AI可能對公民造成的風險層級來建構分類機制。將這些工具列入嚴格法規（即禁止具有危害性技術或不符合倫理標準之的人工智慧應用或系統）、針對高風險應用的特殊法規及針對未列入高風險應用的一般法規。如欲參考這種做法的典範，請參閱歐盟《人工智慧法》草案。
- **保護資料隱私：**認知到生成式AI的使用總是涉及使用者與生成式AI供應商共用資料的事實。應授權進行保護使用者個人資訊保護法規的擬定與施行，識別並打擊非法資料的儲存、剖析與共用（例如身分竊取或隱私洩露）。
- **使用年齡限制的界定與執行：**大多數生成式AI的應用主要為成人使用者所設計。這些應用往往會對兒童造成巨大風險，像是接觸不當內容及受到操弄的可能性。考量這些風險及迭代生成式AI應用存在的高度不確定性，本指引強烈建議對通用人工智慧技術實施年齡限制，以保護兒童的權利與身心福祉。

目前，ChatGPT的使用條款規定使用者必須至少年滿13歲，18歲以下的使用者必須獲得父母或法定監護人的許可才能使用該服務。[52]

這些年齡限制或門檻可追溯至美國的《兒童線上隱私保護法（COPPA）》（Children's Online Privacy Protection Act, Federal Trade Commission, 1998）。該法案於1998年通過，當時社交媒體尚未普及，更別說是ChatGPT等容易上手且功能強大的生成式AI應用了，而美國法律規定，未經家長許可，機構或個人社交媒體供應商不得向13歲以下兒童提供服務。然而，許多評論者認為這道門檻過低，並主張應立法將年齡限制提高為16歲。歐盟《一般資料保護條例》（GDPR）（2016）明文規定，使用者須年滿16歲方能在沒有父母許可的情況下使用社交媒體。

各種生成式AI聊天機器人的出現，使各國不得不認真思考並公開審議使用者與生成式AI平台進行自主對話的適當年齡門檻。最低年齡門檻應為13歲。此外，各國也需要決議以自我申報方式驗證年齡是否合宜。各國將需明定生成式AI供應商在年齡驗證上，以及父母或監護人在監督未成年兒童自主對話方面須負起哪些責任。

◉ **數位貧窮與資料不平等的風險：**採取立法措施保護國家主權內的資料資產，並監管在境內營運的生成式AI供應商。針對公民生成的商用資料集，制定可促進互惠合作的法規，以避免這類資料流出而遭大型科技公司壟斷資料獲利。

## 生成式AI工具供應商

生成式AI供應商包括負責開發與提供生成式AI工具的組織與個人，以及使用生成式AI技術提供服務的組織與個人——包括透過可程式化的應用程式介面（API）提供的服務。大多數具有影響力的生成式AI工具供應商都是資金雄厚的跨國科技企業。

政府應向生成式AI供應商清楚表明，他們必須對倫理嵌入設計（ethics by design）負起責任，包括執行規範中所制定的倫理原則。這些責任應涵蓋以下十種類型：

◉ **人為責任（Human accountabilities）：**生成式AI供應商應負責確保核心價值觀與合法目的的遵循，尊重智慧財產權，維護倫理實踐標準，同時防止假資訊與仇恨言論的散播。

◉ **可信賴的資料與模型（Trustworthy data and models）：**生成式AI的供應商應證明其所使用的資料來源與建模方法具有可信性與倫理基礎。具體包括採用可靠合法來源的資料與基礎模型，並遵守相關智慧財產權法規（如果其資料受智慧財產權保護的話）。此外，若模型涉及個資，只能在資料擁有者知情且明確同意的情況下才能進行蒐集。

[52] https://openai.com/policies/terms-of-use

◉ **非歧視性內容的生成（Non-discriminatory content generation）：**生成式AI供應商必須禁止，基於種族、國籍、性別或其他受保護特徵而產生偏見或歧視性內容的生成式AI系統之設計與部署。他們應該建立健全的「防護機制」（guardrails），防止生成式AI產生攻擊性、帶有偏見或錯誤的內容，同時確保防護機制運作的相關人員不受傷害或剝削。

◉ **生成式AI模型的可解釋性與透明度（Explainability and transparency of GenAI models）：**供應商應向公共治理機構說明其模型所使用資料的來源、規模與類型，預訓練資料的標記規則及訓練流程，模型用以生成內容或回應的邏輯，以及其生成式AI工具所提供的功能及服務。必要時也應協助治理機構了解其中的技術與資料。另外，供應商也應向使用者揭露及告知生成式AI生成錯誤內容與爭議性回應。

◉ **生成式AI內容的標記（Labelling of GenAI content）：**根據人工智慧輔助合成線上資訊的相關法規或條例，供應商必須為生成式AI所產出的論文、報告、圖像與影片加上適當且合法的標記。例如，生成式AI的輸出應明確標示由機器生成，確保內容來源合法且具透明度。

◉ **防護性與安全性原則（Security and safety principles）：**生成式AI供應商應確保生成式AI系統可於完整的生命週期中提供安全、健全與持續性的服務品質。

◉ **存取與使用適當性的說明（Specifications on appropriateness for access and use）：**生成式AI供應商應清楚說明其服務的適用對象、使用情境與目的，協助使用者釐清使用範圍與風險，進而幫助他們做出理性決策。

◉ **承認局限性並預防可預見風險（Acknowledging the limitations and preventing predictable risks）：**生成式AI供應商應明確宣傳其系統使用之方法及其輸出的局限性。他們需要發展機制，以確保輸入資料、模型與輸出內容不會對使用者產生誤導，並確立協定，承諾在工具發生不可預見傷害時減緩衝擊。他們也必須引導使用者根據倫理道德原則去了解生成式AI產出的內容，並且防止使用者過度依賴與盲目信任於所生成的內容。

◉ **申訴與補救機制（Mechanisms for complaints and remedies）：**生成式AI供應商需要建立向使用者與社會大眾開放的申訴機制與管道，並及時接受與處理這些申訴。

◉ **監控與回報非法使用的情況（Monitoring and reporting of unlawful use）：**供應商應與公共治理機構合作，以促進對非法使用情況的監控與回報。使用者若以非法或有違倫理或社會價值觀的方式使用生成式AI產品，例如散播假資訊或仇恨言論、生成垃圾郵件或編寫惡意軟體，也屬於這類情況。

## 機構使用者

機構使用者包含大學與學校等教育當局與機構，他們負責決定機構是否應採用生成式AI，以及應該採購與導入哪些類型的生成式AI工具。

- **對生成式AI的演算法、資料與輸出進行機構稽查：**實施機制以盡可能監控生成式AI工具所使用的演算法與資料、以及其產生的輸出。這些工作應包含定期審查與評估、保護使用者資料及自動過濾不當內容。
- **驗證比例原則並保護使用者福祉：**實施國家分類機制或制定機構層級政策，以對生成式AI的系統與應用進行分類與驗證。確保機構採用的生成式AI系統符合在地驗證的道德框架，並且不會對機構的目標使用者造成可預見的傷害，尤其是兒童與弱勢族群。
- **審視並應對長期影響：**隨著時間的推移，若在教育中依賴生成式AI工具或內容，可能會對人類能力的發展造成深遠影響，例如：批判性思維能力的弱化及創造力的退化等。機構應評估並應對這些潛在的影響。
- **適齡性：**考慮針對機構中生成式AI的自主使用實施使用年齡下限。

## 個人使用者

個人使用者可能包含全球各地所有可取得網際網路並接觸至少一種生成式AI工具的人。這裡所說的「個人使用者」主要指的是正規教育機構的個別教師、研究人員與學習者，或是參與非正規學習課程的個人。

- **生成式AI的使用授權範圍：**使用者在簽署或同意服務協議時，應知悉自己有義務遵守其所規範的授權範圍及其法規或條例基礎。
- **以合乎倫理的方式使用生成式AI：**使用者應負責任地有效運用生成式AI，避免在過程中損害他人的名譽與法律權益。
- **監控與回報非法的生成式AI應用：**使用者發現生成式AI應用違反一或多項法規時，應主動向政府監管機構通報。

# 肆

# 為生成式AI的教育與研究應用制定政策框架

若希望規範生成式AI以利其在教育與研究中發揮潛在益處，便需要制定並落實適當的政策。前面引述的2023年調查資料顯示，只有少數國家採取具體的政策或計畫來規範生成式AI在教育領域的應用。前一章節概述了這項願景、必要步驟，以及核心行動要素。本章則闡述人們可採取哪些措施來制定一致的、全面的政策框架，以規範生成式AI的教育與研究應用。

就此而言，2022年公布的《人工智慧與教育：決策者指南》（AI and education : guidance for policy-makers，UNESCO, 2022b）是一個起點。其提出了一整套的全面建議，指引各國政府制定與實施有關人工智慧與教育的全部門政策，並著重於教育品質、社會公平與包容性的促進。其中多數建議至今依然適用，而且可供進一步調整以指引教育領域生成式AI具體政策的制定。以下提出在規劃教育與研究領域生成式AI政策時可採取的八項具體措施，希望能讓這份指引更顯完善。

## 一 促進包容、公平、語言與文化多元性

在生成式AI的生命週期中，我們必須認知並回應包容性的關鍵重要性。具體而言，除非生成式AI工具的取得管道體現了包容性（即不分性別、種族、特殊教育需求、社經地位、地理位置、遷徙背景等），並在設計上可促進公平、語言多樣性與文化多元性，否則這些工具無助於克服教育領域所面臨的根本性挑戰或實現聯合國制定的2030年永續發展目標四（SDG 4，優質教育）的承諾。對此，本指南建議採取以下三項政策措施：

- 找出未取得或無法負擔網際網路連線或資料的群體，促進上網管道與數位能力的普及性，以減少妨害公平與包容的人工智慧應用獲取機會的阻礙。建立持續性資助機制，為身心障礙或有特殊需求的學習者開發並提供人工智慧工具。提倡生成式AI的使用，為不分年齡、地理位置與背景的終身學習者提供支援。
- 制定生成式AI系統的審查準則，確保資料或演算法未嵌入性別偏見、對邊緣群體的歧視或仇恨言論。
- 制定與實施生成式AI系統的包容性規範，並在教育與研究領域大規模部署生成式AI時，實施保護語言與文化多元性的制度性措施。相關規範應要求生成式AI供應商在訓練GPT模型時納入多種語言資料，尤其是方言或土語，以增進生成式AI回應與生成多語言文本的能力。這些規範與制度性措施應嚴格防止人工智慧供應商刻意或偶然刪除少數民族語言或歧視土語使用者，並要求供應商停止推廣主流語言或文化規範的系統之運作。

## 二 保護人類主體性

隨著生成式AI變得日益複雜，一個關鍵危機是，它有可能損害人類的主體性（agency）。由於有越來越多的使用者利用生成式AI來輔助寫作或其他創造性活動，可能會在無意中變得依賴生成式AI。這種現象會影響心智技能的發展。雖然生成式AI可用於挑戰與擴展人類思維，但我們不應該允許它損害人類的認知及創造能力。在設計與採用生成式AI時，我們應從以下七個面向出發，隨時將保護與增進人類主體性的目標作為核心考量：

- 向學習者知會生成式AI可能會蒐集的資料類型、使用資料的方式，以及對他們的教育與其他生活面向可能導致的影響。
- 保護學習者作為個體成長與學習的內在動機。強化人類在使用日益複雜的生成式AI系統時，對本身進行研究、教學與學習的方法的自主權。
- 防止生成式AI剝奪學習者透過觀察現實世界、實驗等實證實踐、與他人的討論及獨立邏輯推理來發展認知能力與社交技能的機會。
- 確保充分的社會互動與適度接觸人類的創意輸出，並且防止學習者沉迷或依賴生成式AI。
- 使用生成式AI工具來盡可能減輕作業與考試的壓力，而不是加深這種壓力。
- 請教研究人員、教師與學習者對生成式AI的看法，並根據他們的反饋來決定是否、以及如何在機構內部署特定的生成式AI工具。鼓勵學習者、教師與研究人員評論與質疑人工智慧系統所使用的方法、輸出內容的準確性及其可能強行施加的規範或教學法。
- 做重大決定時，避免將人類的判斷責任讓渡給生成式AI系統。

## 三 監控與驗證生成式AI系統的教育應用

如先前所述，生成式AI的開發與部署應合乎道德規範。如此一來，開始採用生成式AI後，便需要在其生命週期內持續進行審慎的監控與驗證：針對其道德風險、教學的適當性與嚴謹性，以及對學生、教師與教學方式、學校規範間關係的影響。對此，本指南建議採取以下五項行動：

◉ 建立驗證機制來測試教育與研究領域所使用的生成式AI系統是否存在偏誤（尤其性別偏見），以及它們是否曾以多元化資料（性別、身心障礙、社經地位、種族與文化背景，以及地理位置之差異）進行訓練。

◉ 正視複雜的知情同意（informed consent）問題，尤其在兒童或其他弱勢學習者無法完全理解其資料使用風險的情境下。

◉ 審查生成式AI的輸出是否包含深度偽造圖像、假新聞（不準確或造假）或仇恨言論。倘若發現生成式AI產生不當內容，機構與教育工作者應願意、並且能夠採取迅速有力的行動來減緩或排除問題。

◉ 教育或研究機構正式採用生成式AI應用之前，必須對其進行倫理性與教育效益評估（即採用倫理嵌入設計的方法）。

◉ 決定在機構內全面採用生成式AI系統之前，應確保相關的生成式AI應用不會對學生造成可預見的傷害、適合且可對目標學習者的年齡與能力發揮教育成效，並符合健全的教學原則（即建立在相關知識領域與預期的學習成果及價值觀發展的基礎上）。

## 四 培養學習者使用人工智慧的能力，包括與生成式AI有關的技能

培養學習者使用人工智慧的能力，是確保人工智慧在教育及其他領域達到安全、符合倫理道德且有意義的運用之關鍵。然而，根據聯合國教科文組織調查，2022年初，只有約15個國家制定並實行、或者正在制定政府認可的正規人工智慧課程（UNESCO, 2022c）。生成式AI的最新發展進一步加深了一項迫切的需求，那就是人人都需要對人工智慧的人文與技術面向有一定的理解，能夠在廣義上了解人工智慧的運作、以及生成式AI的具體影響。對此，我們必須盡快採取以下五項行動：

◉ 致力為學校教育、職業技術教育與培訓（Technical and Vocational Education and Training）及終身學習提供政府核可的人工智慧課程。人工智慧課程應涵蓋這項技術對人們生活的影響，包括其引發的倫理議題、對演算法與資料的適齡理解，以及正確且發揮創意地使用人工智慧工具（包括生成式AI應用）的技能。

◉ 支持高等教育與研究機構推行在地人工智慧的人才培育計畫。

◉ 在發展進階的人工智慧職能時，提倡性別平等並建立性別平衡的專業人才庫。

◉ 針對最新的生成式AI自動化趨勢所導致的全國與全球性工作轉換進行跨部門預測，並根據預期的需求變化，來提升各級教育與終身學習系統中與時俱進的技能。

◉ 為需要學習新技能與適應新環境的中高齡者制定專屬方案。

## 五 培養教師與研究人員正確使用生成式AI的能力

2023年針對政府層級AI教育政策的調查（UNESCO, 2023c）指出，只有7個國家（中國、芬蘭、喬治亞、卡達、西班牙、泰國與土耳其）表示已經或正在制定教師人工智慧框架或培訓計畫。而且只有新加坡的教育部建立了一個線上資源庫，其核心內容為ChatGPT的教育與學習應用。這清楚顯示，多數國家的教師普遍缺乏結構完善的人工智慧教育應用培訓資源，尤其是在生成式AI應用方面的專業訓練更為稀缺。為了幫助教師負責且有效地使用生成式AI，各國需採取以下四項行動：

◉ 根據在地測驗來制定或調整指導原則，幫助研究人員與教師善用生成式AI工具，並指導新興領域特定人工智慧的應用與設計。

◉ 保護教師與研究人員在使用生成式AI時的權利及其實踐價值。更具體地說，就是分析教師在促進高層次思考、組織人類互動與培養人類價值觀方面的獨特角色。

◉ 定義教師在有效且合乎倫理道德地理解及使用生成式AI系統時所需具備的價值取向、知識與技能。讓教師有能力創造特定的生成式AI工具，以促進課程學習及其自身的專業發展。

◉ 滾動式審查教師在教學與專業學習中理解並使用人工智慧所需具備的能力；將新興的人工智慧價值觀、理解與技能納入在職與職前教師培訓的能力框架與計畫。

## 六 鼓勵多元觀點及想法的表達

如先前所述，生成式AI既無法理解提示內容，也不明白回應的涵義。它的回應乃是基於模型訓練期間自網際網路蒐集資料中所發現之語言模式機率進行生成。為了解決其輸出的一些基本問題，目前領域正在研究新方法，例如將生成式AI與知識資料庫及推理引擎進行整合。儘管如此，由於其運作方式、資料來源及其開發者隱含的觀點，生成式AI顧名思義在輸出中複製了主流世界觀，並削弱少數意見與多元觀點的呈現空間。因此，如果想讓人類文明蓬勃發展，我們必須意識到，無論生成式AI涉及主題為何，它永遠無法成為權威的知識來源。

因此，使用者必須從批判性角度來看待生成式AI的產出。尤其是：

- 了解生成式AI是快速的資訊來源，但往往不可靠。雖然前述的一些外掛程式與基於LLM的工具旨在滿足經過驗證且最新的資訊之需求，但幾乎沒有確切證據表明這些工具是有效的。
- 鼓勵學習者與研究人員批判生成式AI提供的回應。他們應該有所認知到，生成式AI通常只是重複既定或主流的觀點，因而壓縮了多元與少數族群的觀點及多元思想的表述。
- 讓學習者有充分機會可從反覆試驗、實證實驗與對真實世界的觀察中學習。

## 七 測試具在地適切性的應用模型，並建立累積性實證基礎

到目前為止，生成式AI模型多以全球北方（Global North）的資訊為主，而在呈現全球南方與原住民族群的觀點上有所不足。唯有透過堅定的努力（譬如利用人工合成資料技術〔Marwala, T. 2023〕），才能讓生成式AI工具對本地群體（尤其是南方世界）的背景與需求保持敏感。為了探索能切合在地需求的做法，同時進行更廣泛的協作，本指引建議採取以下八項行動：

- 務必對生成式AI的設計與採用進行策略性規劃，而非被動應用與不加批判的採購過程。
- 激勵生成式AI的設計者以開放、探索導向與多元化的學習選擇為目標。
- 根據教育的優先目標來測試與擴大AI在教育與研究中的實證使用案例，而非著重於人工智慧的新奇、神話或宣傳炒作。
- 引導領域使用生成式AI以激發研究創新，包含利用運算能力、大規模資料與生成式AI產出，來影響並啟發研究方法的改善。
- 審視將生成式AI納入研究過程所具備的社會與倫理意涵。
- 根據實證教學研究與方法來制定具體標準，並為生成式AI在支援兼容性學習機會、學習與研究目標及語言與文化多元性方面的有效性，建立實證基礎。
- 應採取循序漸進的行動，以強化關於生成式AI在社會與倫理層面之影響的實證基礎。
- 分析大規模運用人工智慧技術的環境成本（如訓練GPT模型所耗費的能源與資源），並制定人工智慧供應商應達到的永續目標以避免氣候變遷加劇。

## 八 以跨部門與跨學科視角審視長期影響

跨部門（intersectoral）與跨學科（interdisciplinary）的方法，對生成式AI在教育與研究中獲得有效且合乎倫理的運用至關重要。唯有善用各種專業知識、同時集結各方利害關係人，才能及時發現且有效克服關鍵挑戰，以盡可能減少長期負面影響，同時最大化其社會效益。因此，本指引建議採取以下三項行動：

- 促進不同領域利害關係人協作，包括人工智慧供應商、教育工作者、研究人員及家長與學生代表，共同規劃如何針對課程框架與評估方法進行全系統的調整，以充分發揮生成式AI的潛力並降低其對教育與研究可能造成的風險。
- 匯集跨部門與跨學科的專業知識，包含教育工作者、研究人員、學習科學家、人工智慧工程師與其他利益相關者的代表，共同探討生成式AI為學習與知識生產、研究與著作權、課程設計與評估及人類合作與社會動態帶來的長期影響。
- 定期提供政策更新建議，以利法規與政策的迭代與更新。

# 伍

# 促進生成式AI在教育與研究領域中的創造應用

在ChatGPT首次問世時，全球各地的教育工作者即對其潛在影響表達憂慮，特別是其可能讓學生生成論文並進行抄襲的風險。近年來，包含數所世界一流的大學在內的許多人士與組織都認為，這種現象就有如「瓶中的精靈被釋放」般一發不可收拾，而像ChatGPT這樣的工具將持續存在，並在教育環境中發揮建設性效益。在此同時，網際網路充斥著有關生成式AI應用於教育與研究領域的各種建議。這些建議包含，利用它來啟發新思維、生成多元觀點的範例、擬定教學計畫與簡報、統整既有素材及激發圖像創作。儘管幾乎每天都有新的想法出現在網際網路上，但研究人員與教育工作者仍持續研究生成式AI在教學、學習與研究中代表的實質意涵。值得注意的是，許多做出這些提議的人士或許並未納入倫理考量，而其他人考量的則是生成式AI的技術潛力，而非研究人員、教師或學習者的實際需求。本章節概述了如何促進生成式AI在教育領域中的創新性使用。

## 一 可促進生成式AI之負責且創造性使用的制度性策略

如先前所述，教育與研究機構應制定、實施及驗證適當的策略與倫理框架，來指引負責任且合乎道德地使用生成式AI系統與應用的做法，以滿足教學、學習與研究的需求。這個目標可經由下列四種策略來達成：

- **從制度上落實倫理原則：**確保研究人員、教師與學習者以負責任且合乎倫理的方式使用生成式AI工具，並從批判角度來評估其輸出的正確性與有效性。
- **指引與培訓：**為研究人員、教師與學習者提供生成式AI工具的相關指引與訓練，確保他們了解存在於資料標記偏誤與演算法偏見等潛在倫理議題，以及遵循有關資料隱私與智慧財產權等相關法規。
- **培養生成式AI提示工程的能力：**除了特定學科的知識之外，研究人員與教師也需要具備撰寫與評估提示語（prompt）的能力。有鑒於生成式AI帶來的挑戰複雜且棘手，研究人員與教師必須受到優質的培訓與專業支援才有辦法發展出這樣的能力。
- **偵測書面作業中使用生成式AI而構成的抄襲行為：**生成式AI可能會讓學生將他人撰寫的文本竊為己用，形成「新型態抄襲」。生成式AI供應商依規定必須替生成式AI的輸出加上為「由人工智慧生成」的浮水印，同時開發一些工具以供偵測人工智慧生成的素材。然而，幾乎沒有證據證明這些措施或工具的有效性。當前應立即採取的制度性策略，是透過人為審查來維護學術誠信及問責原則。長期策略是，機構與教育工作者應重新思考書面作業的設計方式，避免其落入僅評量生成式AI所擅長的任務（如：文字產出）。相反地，他們應該著重於只有人類才做得到的事情，像是面對複雜的現實世界挑戰時發揮同情心與創造力等人類價值觀。

## 二 「以人為本與教學相長的互動」方式

研究人員與教育工作者在決定是否與如何使用生成式AI時，應優先考量如何確保人類主體性（human agency）、以及人類與人工智慧之間負責任且教學相長的互動。這包含了下列五點考量：

◉ 工具的使用應有助於滿足人類的需求，並且讓學習或研究的成效超越無技術或其他替代方法的應用；

◉ 教育工作者與學習者對工具的使用應以內在動機為基礎；

◉ 工具的使用過程應由教育工作者、學習者或研究者主動掌握及控制；

◉ 根據學習者的年齡範圍、預期達到的成果與目標知識（如事實性、概念性、程序性或後設認知）或目標問題的類型（如結構健全或結構不健全），來選擇與組織相應的工具，以及生成相應的內容；

◉ 使用過程應確保人類與生成式AI與高層次思考有所互動，以及人類對人工智慧生成的內容、教學或研究策略之正確性及其對人類行為之影響等相關決定負起責任。

## 三 共同設計生成式AI的教育與研究應用

生成式AI在教育與研究中的應用既不應該採取由上而下強制推行的方式，也不應該受到商業炒作所驅使。相反地，教師、學習者與研究人員應共同設計（co-designed）如何安全而有效地使用生成式AI。對此，還需要設計一個嚴謹的試驗與評估過程，以檢視其在不同使用方式下的效果與長期影響。

為了促進以上建議的共同設計，本指引提出一套由下列六個觀點所構成的框架，以強化教學相長的互動與人類主體性的優先考量：

◉ 適當的知識或問題領域。

◉ 預期達到的成果。

◉ 適當的生成式AI工具與其比較優勢。

◉ 對使用者的條件要求。

◉ 所需的人類教學方式與提示範例。

◉ 倫理風險考量。

本節將舉例說明如何透過共同設計的過程來影響研究的實踐、輔助教學、指導基礎技能的自學、促進高階思維及協助有特殊需求的學習者。這些例子只是生成式AI在眾多領域中具有潛力的冰山一角。

## 生成式AI在研究中的應用

生成式AI模型已展現出其潛能，包括擴展研究提綱的觀點，以及豐富資料探索與文獻回顧的深度（見表3）。儘管未來可能出現更多應用實例，但仍需展開進一步的研究，來界定潛在的研究問題與預期成果，以證明其效能與正確性，並且確保人類透過研究來理解現實世界的主體性不會因為人工智慧工具的應用而遭到削弱。

表3 共同設計生成式 AI在研究中的應用方式

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| 協助擬定研究提綱的人工智慧顧問<br>AI advisor for research outlines | 對結構健全的研究問題可能有益。 | 提出並解決研究問題，建議適當的方法。<br>**潛在的轉型：針對研究規劃的一對一指導**<br>**1:1 coach for research planning** | 從第壹章第二節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過當局的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 研究人員必須對研究主題有基本的認識。<br>研究人員應培養核查資訊的能力，尤其是偵查不存在的研究論文與文本之引述、以及回應問題的能力。 | 關於研究問題定義（如目標受眾、議題、背景），以及教學法、預期達到的成果與格式的基本概念。<br>針對〔主題x〕寫出10個潛在的研究問題，並按照其對〔研究領域y〕的重要性加以排序。 | 需要警惕生成式AI很可能會捏造資訊（譬如不存在的研究出版物），以及使用者很有可能一字不改地採用人工智慧所生成的研究提綱，如此一來，會削弱初級研究人員從試驗與錯誤中學習的機會。 |
| 探索生成性資料與文獻回顧<br>Generative data explorer and literature reviewer | 對結構不夠健全的研究問題可能有益。 | 自動匯集資訊，探索各種資料，擬定文獻回顧草稿，自動進行部分的資料解讀。<br>**潛在的轉型：針對資料探索與文獻回顧的人工智慧培訓師**<br>**AI trainers for data exploration and literature reviews** | 從第壹章第二節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過當局的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 研究人員必須具備健全的資料分析方法以及技巧。 | 對問題的逐步釐清定義、資料與文獻來源的範疇界定、用於資料探索與文獻回顧的方法，以及預期達到的成果及其呈現格式的規劃。 | 需要警惕生成式AI所捏造的資訊、不當的資料處理、可能的隱私侵犯行為、未經著作權授權的定性分析與性別偏見。<br>需要對主流規範的宣傳及其對替代規範與多元意見所造成的壓縮保持警覺。 |

## 促進教學的生成式AI

不論使用一般的生成式AI平台或設計特定的教育生成式AI工具，目的都應該是增進教師對所屬學科的理解及對教學方法的認識，而這個目標可經由教師與人工智慧共同設計教案、套裝課程或整體課綱的方式來達成。利用經驗豐富的教師與圖書館所提供的資源，進行預先訓練的生成式AI輔助對話式助教（conversational teachers'assistants）或是「生成式分身助教[53]」（generative twins of teaching assistants），已在一些教育機構接受過測試，並可能蘊藏未知的潛力與倫理風險。這些模型的實作流程與進一步的迭代運算，仍需在本指引所建議的框架經過仔細審查，並受到人為監督機制的把關，如表4所示。

表4 共同設計生成式AI對教師與教學的支援

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| 課程或教案共同設計者<br>Curriculum or course co-designer | 關於特定教學主題的概念性知識與教學法的程序性知識。 | 輔助所有課程與單元課程的設計過程，包含概述或是延伸目標主題主要領域的觀點、定義課程結構。<br>另外也可提供考題範例以及評量規則，協助教師設計測驗與考試。<br>**潛在的轉型：AI生成的課程**<br>AI-generated curriculum | 從第壹章第二節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過當局的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 教師必須了解與詳細指明，他們希望在不同種類的課程或測驗中涵蓋的內容與達成的目標，意圖教授概念性或程序性知識，以及希望運用哪一套教學理論。 | 提出教學架構的提示語（prompt）設計，例如建議主題架構、概念、評量規準或教案架構等。<br>教師需評估AI所提出的內容是否正確，以及課程架構是否恰當。 | 生成式AI很可能會套用主流規範與教學方法。<br>它可能會在無意之中延續排斥性做法，讓已經擁有豐富資料的群體占有優勢，並且排除資料較少或處境不利的學習者。 |
| 教學助理型的生成式聊天機器人<br>Generative chatbot as teaching assistant | 對結構健全的問題具備跨領域的概念性知識。 | 提供個人化支援，回應問題及識別資源。<br>**潛在的轉型：生成式雙胞助教**<br>Generative twins of teachers' assistants | 從第壹章第二節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過當局的嚴謹測試或驗證。 | 它為教師直接與學生互動提供了支援。因此學習者必須具備足夠的先驗知識、能力與後設認知技巧，以驗證生成式AI的輸出並識別其中的錯誤資訊。 | 教師必須清楚理解問題、監控對話並協助學習者驗證生成式AI提供的可疑答案。 | 根據生成式AI模型目前所具備的能力，教育機構務必對生成式AI工具提供的回應進行人為監督，並且對錯誤資訊的風險保持警覺。 |

[53]在某些國家，每位教師會有一名助教（TA），其職責是花時間回答個別學生對於教材所產生的疑問。生成式AI可用於發展助教的「生成式雙胞」來輔助學生與其他教師，但也可能造成一些負面問題（如牽涉課堂上的社交關係）。

表4 共同設計生成式AI對教師與教學的支援（續）

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| | | | | 因此，它可能更適合高等教育階段的學習者。 | | 它也有可能限制學習者獲得人為指導與支持的機會，削弱強化師生關係的可能性，而這對兒童來說，尤其令人擔憂。 |

## 生成式AI作為基礎能力自我引導學習教練

定義學習成果時，雖然高層次思維與創造力日益受到關注，但基礎技能對兒童的心理發展與能力發展的重要性仍然毋庸置疑。在眾多能力之中，這些基礎技能包含了母語或外語的聽、說、寫，以及基本計算、藝術與程式碼編寫。重複操練」（drill and practice）不應被視為過時的教學方法，而是應該經由生成式AI技術加以復甦與升級，以幫助學習者進行「自我引導式重複練習」（self-paced rehearsal）的重要工具。在道德與教學原則的指引下，生成式AI工具有潛力發展為一對一教練（1:1 coach），有效支持學習者進行此類自我節奏學習。範例如表5所示。

表5 共同設計生成式AI作為學習者在語言與藝術領域中自我引導學習之一對一教練的用途

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| 一對一語言技能教練<br>1:1 language skills coach | 語言學習，包含對話練習。 | 協助學習者進行對話練習，透過給予反饋、糾正與母語或外語示範，來幫助他們促進聽、說、讀與寫的技能。<br>**潛在的轉型：入門的一對一語言輔導**<br>**1:1 language tutorials at beginner level** | 從第壹章第三節中的所列清單中著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 有鑒於生成式AI系統的輸出可能有文化不敏感或產生不適當回應的問題，可為獨立對話設定年齡限制。<br>學習者必須具備與人工智慧系統進行對話的內在動機，並要能夠從批判的角度看待生成式AI的建議，並檢視這些建議是否為正確。 | 在使用一般的生成式AI平台時，教師可引導學習者與生成式AI工具進行互動，請求AI建議學習者可改進的部分、糾正發音或提供寫作範例。例如：<br>◈提供一些想法，幫助我撰寫有關於〔主題X〕的文章。<br>◈用〔X〕語言與我對話，幫助我持續改進語言技能。 | 需要警惕對文化不敏感或產生語境不正確的語言，以及無意中延續的刻板印象或文化偏見。<br>如果沒有適當的教學策略，便可能會限制學習者的創造力與原創性，導致公式化寫作模式。 |

表5 共同設計生成式AI作為學習者在語言與藝術領域中自我引導學習之一對一教練的用途（續）

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| | | | | | | 這也可能限制現實生活中的互動、多元觀點、多元表達與批判性思考的機會。 |
| 一對一藝術教練<br>1:1 art coach | 掌握音樂與繪畫等藝術領域的專門技術。 | 提供個人化指導，回應問題及識別資源。<br>**潛在的轉型：一對一入門級藝術教師**<br>**1:1 art tutorials at beginner level** | 從第壹章第三節之二點中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 學習者必須就藝術或音樂的創作立定一些初步目標，對藝術或音樂領域的關鍵要素有基本的認識，並具備分析藝術作品或音樂作品的基礎能力。 | 教師或教練必須鼓勵學習者發展與應用他們的想像力與創造力，以Gen AI作為創意補充而非完全仰賴。提示範例：<br>◈提出一些想法，啟發我創作一幅關於〔主題／想法〕的圖像。 | 可能會讓兒童接觸到不適當或令人反感的內容，侵犯他們的保障與福祉權。<br>生成式AI工具可能會阻止學習者發揮想像力與創造力。 |
| 一對一的程式編碼或算術輔導<br>1:1 coach for coding or arithmetic | 入門級的概念性程式設計知識與技能。這也適用於基礎數學的學習。 | 支援基本程式編碼知識與技能的自學，發現學習者編碼中的錯誤，並給予即時反饋，以及根據問題提供量身訂做的答案。<br>**潛在的轉型：一對一入門級程式編碼教師**<br>**1:1 coding teacher at introductory level** | 從第壹章第三節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 發現與定義問題，及設計解決問題的演算法，至今依然是學習編碼與程式設計的核心面向。學習者必須具備使用編碼的內在動機，以及一些使用程式設計語言的一些基本知識與技能。 | 教師與教練應教導基本的知識與技能，啟發學習者利用運算思維與程式設計來解決問題，途徑包含協作編碼。提示範例：<br>◈提出一些不同凡響的編碼想法。 | 反饋與建議的準確性仍是一個未解的問題，因為生成式AI未必都是對的。<br>生成式AI工具很有可能會阻礙學習者發展運算思維的技能以及發現並定義重大編碼問題的能力。 |

## 促進探索或專題式學習的生成式AI

如果不是刻意用於促進高層次思維或創造力，生成式AI工具往往會鼓勵抄襲或膚淺的「隨機模仿」輸出。然而，有鑒於生成式AI模型向來以大規模資料為訓練基礎，它們或許能夠進行蘇格拉底式問答對話，或者在專題式學習中扮演研究助理的角色。然而，唯有透過旨在引發高層次思維的學習設計過程，這些潛力才能發揮，如表6所示。

表6 共同設計生成式AI促進探索或專題式學習的用途

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| 蘇格拉底式問答對話的挑戰者<br>Socratic Challenger | 結構不健全的問題。 | 協助學習者進行質疑先驗知識的蘇格拉底式對話，進而發現新知識或加深理解。<br>**潛在的轉型：一對一的蘇格拉底式對話者**<br>**1:1 Socratic Opponent** | 從第壹章第三節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 學習者必須要達到能夠與生成式AI工具進行獨立對話的年齡。此外，學習者也必須具備先驗知識與檢視AI所提出的論點與資訊是否正確的能力。 | 教師可協助準備一系列由淺入深的問題作為範例，供學習者改編成提示。學習者也可從概括的提示開始，如：<br>◈與我進行蘇格拉底式問答對話，幫助我從批判性角度探討〔主題x〕<br>接著透過越來越精確的提示，逐步深化對話。 | 目前的生成式AI工具可能會生成相似或標準的答案，進而限制學習者接觸多元觀點與其他看法的機會，導致同溫層效應，阻礙獨立思考的發展。 |
| 作為專題式學習的顧問<br>Advisor for project-based learning | 科學或社會研究中結構不健全的研究問題。 | 幫助學習者進行專題式學習，以支持知識的創造，譬如生成AI扮演類似表3所描述的研究顧問。<br>**潛在的轉型：一對一專題式學習教練**<br>**1:1 projectbased learning coach** | 從第壹章第三節中的清單著手，評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 學習者可以扮演初級研究人員的角色，規劃與實行專題式學習。學習者的年齡必須足以獨立使用生成式AI平台。此外，他們必須具備進行自主專題式學習活動的動機以及能力，才不會被動地照抄生成式AI工具所提供的答案。 | 教師可引導學習者在定義研究問題時請求生成式AI提供基本概念，如第伍章第三節之一點所示。個人與團體學習者可利用生成式AI工具進行文獻回顧、蒐集與處理資料，以及製作報告。 | 學習者若沒有扎實的先驗知識與必要能力來驗證答案正確性，可能會被生成式AI工具提供的資訊所誤導。生成式AI也有可能限制學習者與同儕的討論與互動，減少協作學習的機會，因而削弱他們的社交發展。 |

## 支援有特殊需求的學習者的生成式AI

理論上，生成式AI模型具有協助有聽力或視力障礙的學習者的潛力。近年來新興的做法包括利用生成式AI來為聾人與重聽學習者提供字幕或標示，以及為視障學習者提供語音描述。生成式AI模型也可以將文本轉換成語音或將語音轉換成文本，讓視覺、聽覺或語言障礙的學習者得以存取內容、提出問題並與同儕進行溝通。然而，這項功能尚未獲得廣泛運用。根據教科文組織在2023年就各國政府在教育領域運用人工智慧的情況所做的調查，只有四個國家（中國、約旦、馬來西亞與卡達）表示，其政府機構已驗證並建議使用人工智慧輔助工具，來協助為身心障礙學習者提供開放、兼容的學習機會（UNESCO, 2023c）。另外還出現了一種趨勢，即生成式AI模型的迭代經過訓練，可支援學習者使用自身語言（包含少數民族與土著的語言）進行學習與溝通。舉例來說，Google的PaLM 2（次世代大型語言模型）即基於多語對照語料（source–target text pairs）來進行訓練。這個模型之所以納入並行多語言資料，是為了進一步強化其理解與生成多語言文本的能力（Google, 2023b）。

透過即時翻譯、改述與自動校正的功能，生成式AI工具有潛力幫助那些使用少數民族語言的學習者互相交流，並增進他們與不同語言背景的同儕之間的合作。然而，若無有意義的設計與導入，這種潛力難以轉化為實質成果。

最後，有人認為生成式AI系統有潛力根據對話進行診斷，識別心理或社會情緒問題及學習障礙。然而，目前仍缺乏證據顯示這種方法是否有效或安全，而且，任何診斷都需要由技術熟練的專業人員加以解釋。

表7 共同設計生成式AI支援特殊需求學習者的用途

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| 從對話中診斷學習障礙<br>Conversational diagnosis of learning difficulties | 可能對因心理、社交或情緒問題而遭遇學習障礙的學習者有所幫助。 | 利用自然語言參與來識別有心理、社交或情緒問題或學習障礙的學習者的需求，以為他們提供相關的支持或指導。<br>**潛在的轉型：為有社交或情緒問題或學習障礙的學生提供一對一初級顧問服務** | 除了一般的生成式AI工具外，也搜尋由生成式AI支援的聊天機器人。<br>評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。 | 教師或與這群學習者合作的專家需要確保生成式AI提供的初步建議正確無誤。 | 教師或輔導人員需要為學習者提供進行對話的舒適環境，以診斷學習者是否具有心理、社交或情緒問題，或是學習障礙。 | 可能會在無意中誤判學習者遭遇的困難，因而提供錯誤的協助。 |

表7 共同設計生成式AI支援特殊需求學習者的用途（續）

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| | | 1:1 primary advisor for learners with social or emotional problems or learning difficulties | 進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | | | |
| 人工智慧驅動的無障礙工具<br>AI-powered accessibility tools | 讓有聽力或視力障礙的學習者能夠獲取更廣泛的內容，進而提高學習品質。 | 針對音訊或影像內容提供由生成式AI支援的字幕或手語翻譯，及文本或其他視覺素材的影像口述，以滿足學習者的無障礙需求並協助他們獲取特定學科的知識。<br>**潛在的轉型：一對一的個人化人工智慧語言輔助工具**<br>1:1 personalized AI-powered language aids | 除了一般的生成式AI工具之外，也搜尋由相關且值得信賴的字幕與影像口述之人工智慧生成器。<br>評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。<br>進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | 教育工作者或輔導人員必須幫助學習者獲取與學習如何操作生成式AI工具。此外，他們也需要確保這些工具的輸出能夠真正支持這些學習者，而且不會加深他們面臨的挑戰與偏見。 | 需要測試平台或工具的無障礙性，才能在使用之前發現並且解決這個問題。生成式AI工具僅提供獲取內容的管道，因此教育工作者與輔導人員應著重於提升這類學習者的學習品質與社會福祉。<br>教育工作者與輔導人員需要根據學習者的能力來教導他們創造語音或文字提示。 | 如果不是專為協助視力或聽力障礙者所設計，那些由生成式AI平台製作的字幕或口述影像往往不正確，而且可能會誤導有特殊需求的學習者。<br>這些工具有可能在無意中加深及再製既有的偏見。 |
| 邊緣化學習者的生成式放大器<br>Generative AI amplifier for marginalized learners | 可能有助於來自少數語言或文化背景的學習者表達以及放大他們的聲音、參與線上活動及進行協作社會研究。 | 提供即時翻譯、改述與寫作自動校正服務，協助來自邊緣化群體的學習者使用本身的語言與來自不同語言背景的同伴溝通。 | 具體案例為PaLM 2。<br>評估生成式AI工具是否可供在地使用者獲取、是否為開放來源、經過公信機構的嚴謹測試或驗證。 | 學習者應該對關於對話或協作研究的主題有一定的認識，或者提出具有意義的看法。他們必須有能力做出負責任與不帶歧視的貢獻，並且避免仇恨言論。 | 教師與教育工作者應為學習者設計社會或是文化主題的學習與寫作任務，或者組織線上研討會或跨文化合作，以激發學習者產生想法與分享觀點。 | 需要識別與糾正人工智慧的翻譯與改述中，可能導致跨文化誤解的錯誤。 |

表7 共同設計生成式AI支援特殊需求學習者的用途（續）

| 具潛力但尚未證實的用途 | 適當的知識領域或問題 | 預期成果 | 合適的生成式AI工具與比較優勢 | 對使用者的要求 | 所需的人類教學方式與提示範例 | 潛在風險 |
|---|---|---|---|---|---|---|
| | | **潛在的轉型：為邊緣化學習者提供開放兼容的大型語言模型**<br>Inclusive LLMs for marginalized learners | 進一步考量任何特定生成式AI工具的優勢與挑戰，並確保其能適當滿足人類的特定需求。 | | | 這項用途讓邊緣化學習者有機會放大自己的聲音，但無法觸及資料貧乏的根本原因，因此無法實現人工智慧工具的去殖民化。 |

# 陸

# 生成式AI與教育研究的未來

生成式AI技術仍在迅速發展中，很有可能對教育與研究領域造成深遠影響，而這些影響還有待充分了解。因此，我們需要即刻關注、並且深入探討生成式AI對教育與研究的長期影響。

## 一 尚未釐清的倫理問題

日益複雜的生成式AI工具將引發額外的倫理問題，需要詳加檢視。承接第二章及第三章所述，我們還需要進行更深入、更具前瞻性的分析以揭露並面對未知的倫理問題。對此，至少可以從下列五個面向出發：

- **獲取管道與公平性**：生成式AI系統在教育領域的應用可能會加劇在科技與教育資源管道方面的既有差距，進一步加深不平等的現象。
- **人與人之間的連結**：生成式AI系統在教育領域的應用可能會減少人與人之間的互動，並且妨害至關重要的社會情緒學習。
- **人類智力發展**：生成式AI系統在教育領域的應用可能會限制學習者的自主性與能動性，因為其提供了預設的解決方案或限縮了可能的學習經驗範圍。它們對年輕學習者的智力發展所造成的長期影響有待深入研究。
- **心理影響**：可模仿人類互動的生成式AI系統可能會對學習者產生未知的心理影響，引發對其認知發展、情緒健康與操控風險的疑慮。
- **潛藏的偏見與歧視**：隨著教育領域導入日益複雜的生成式AI系統，這些工具很有可能會根據用以訓練模型的資料與方法來產生新型態的偏見與歧視，進而導致未知與潛在有害的輸出。

## 二 著作權與智慧財產權

生成式AI的出現正迅速改變科學、藝術與文學作品的創作、流通與使用方式。未經授權而擅自複製、散布或使用著作權作品的行為，侵犯了著作權者的專有權，並可能導致法律責任。例如，生成式AI模型的訓練過程已多次被控侵犯著作權。最近的一個案例是，AI結合饒舌歌手德瑞克（Drake）與藍調歌手威肯（The Weeknd，原名Abel Tesfaye）的聲音所生成的歌曲在獲得數百萬次點閱後，因著作權糾紛被強制下架（Coscarelli, 2023）。雖然新出現的監管框架有意要求生成式AI供應商認定並尊重原始內容創作者的智慧財產權，但要判定數量龐大的生成作品的著作權與原創性變得愈加困難。這種可追溯性的欠缺，不僅妨礙原創作者的權利保護與合理報酬，也使得教育情境中的應用正當性受到挑戰。這樣的發展對研究系統可能造成的影響不容小覷。

## 三 內容的來源與學習的本質

生成式AI工具正在改變教學內容的生成與提供方式。未來，經由人機對話所生成的內容，也許會成為知識生產的主要來源之一。這個趨勢可能進一步削弱學習者對與以人類所創造與驗證的資源、教科書與課程的教育內容的直接參與。生成式AI文本的權威表象可能會誤導年輕學習者，因為他們沒有足夠的先驗知識來識別不正確的部分或提出有效的質疑。關於學習者接觸未經驗證內容的這種經驗是否應被視為「學習」的問題，也存在爭議。

因此而著重於總體二手資訊的傾向，也有可能減少學習者透過已證實有效的方法建構知識的機會，譬如直接感知與體驗真實世界、從嘗試與錯誤中學習、展開經驗性實驗與培養常識。這也可能會削弱知識的社會建構與通過課堂合作實踐培養社會價值觀。

## 四 同質化的回應對比多元且具創造性的產出

生成式AI限縮了多元敘事的範圍，因為生成的結果往往象徵且強化主流觀點。因此所致的知識同質化限制了多元與創造性思維。教師與學生高度依賴生成式AI工具以尋求建議的現象，會導致回應的標準化與同質性，進而削弱獨立思考與自我導向學習的能力。文章與藝術創作的表述的潛在同質化，會使學習者的想像力、創造力與觀點多樣性受到侷限。

生成式AI供應商與教育工作者必須考量的是，可在多大程度上發展與使用EdGPT來助長創造力、協同合作、批判性思考與其他高階思考技能。

## 五 重新思考評量與學習成果

生成式AI對評量造成的影響，遠超越了學習者作業抄襲的疑慮。我們必須正視的事實是，生成式AI可以寫出條理分明的論文與散文及令人印象深刻的藝術作品，並且能夠通過某些以知識為基礎的科目考試。因此，我們需要重新審視究竟應該學習什麼、達到什麼目的，以及如何評估與驗證學習效果。

教育工作者、政策制定者、學習者與其他利益相關者需就下列四種類型的學習成果進行深入探討：

◉ **價值觀（Values）：**確保科技的人本設計與使用所需的價值觀，是我們在重新思考數位時代的學習成果及其評估時應該秉持的核心概念。重新審視教育目的時，應明確指出科技與教育有所關聯的價值觀。正是透過這種規範性視角，學習成果及其評估與驗證需要不斷與時俱進，尤其是生成式AI的應用。

◉ **基礎知識與技能（Foundational knowledge and skills）：**即使在生成式AI工具表現比人類來得出色的能力領域中，學習者仍然需要建立扎實的基礎知識與技能。基礎的識字、運算與科學素養的技能，仍將是未來教育的核心。我們需要重新界定這些基礎技能的範疇與本質，以因應AI密集環境的變化。

◉ **高階思考技能（Higher-order thinking skills）：**學習成果必須涵蓋可支持高層次思考與問題解決所需的技能，而這些技能應建立在人機協作及生成式AI輸出的使用上。這牽涉了對事實與概念性知識在高層次思考技能中發揮的作用之理解，以及對人工智慧所生成內容的批判性評估。

◉ **與人工智慧合作所需的職業技能（Vocational skills needed to work with AI）：**在人工智慧表現比人類來得出色、而且能自動執行任務的領域中，學習者需要培養新技能，以便開發、操作與使用生成式AI工具。學習成果與教育評量的重新設計，亦需對應未來AI所創造的職業型態。

## 六 思考過程

就生成式AI對教育與研究的長期影響而言，最根本的觀點依然涉及了人類自主性與機器之間的互補關係。其中一個關鍵問題是，人類是否有可能將思考與技能獲取過程的基本層面讓渡給人工智慧，轉而專注於以人工智慧的輸出為基礎的高階思考技能。

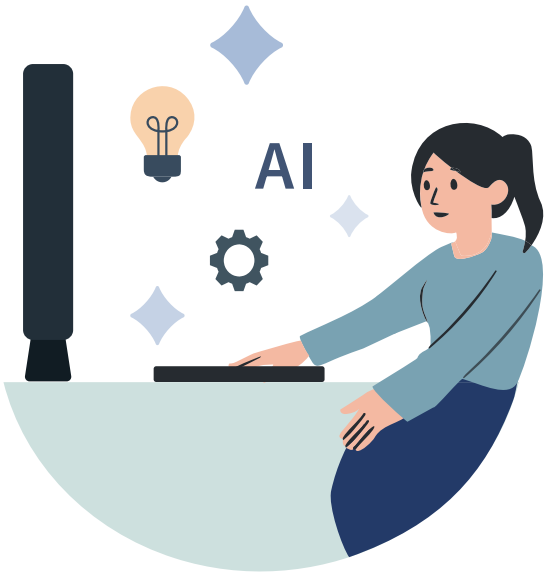
舉例來說，寫作通常與思考建構歷程有關。如今有了生成式AI，人類可以從它生成的完整提綱著手，而不是從零開始規劃一系列想法的目的、範圍與概要。有專家將此現象描述為「無思維寫作」（writing without thinking）（Chayka,2023）。隨著這些新出現的生成式AI輔助實踐逐漸被廣泛應用，建立與評估寫作技能的既有方法也需要有所調整。未來的一種可能策略，是將寫作學習著重於培養規劃與撰寫提示語（prompts）、從批判性角度評估生成式AI的輸出、發展高階思考能力，以及根據生成式AI提供的大綱進行共同寫作（co-writing）等技能。

# 結語

從人本導向的觀點出發，人工智慧工具的設計應致力於擴展或增強人類的智力與社會技能，而非削弱、衝突或取代這些能力。長久以來，人們期待人工智慧工具能進一步整合為人類可用工具的一部分，協助分析與行動，並且實現更具包容性與永續性的未來。

為了讓人工智慧在個人、體制與系統層面成為人機協作中值得信賴的重要部分，應根據生成式AI等新興技術的具體特性，進一步具體化並落實2021年聯合國教科文組織發布的《人工智慧倫理建議書》(*Recommendation on the Ethics of AI*)所提出的人本導向原則。唯有如此，我們才能確保生成式AI可作為研究人員、教師與學習者值得信賴的工具。

使用生成式AI來滿足教育與研究之需求的同時，我們亦需有所警覺，生成式AI也可能會改變這些領域的既有系統及價值觀。對於生成式AI所引發的教育與研究變革，應透過人本方式進行嚴謹的審查與指引。如此一來，我們才能確保人工智慧及所有其他應用於教育領域的技術發揮潛能，增進人類的能力，為所有人建構出兼容並蓄的數位未來。



# 參考資料


Anders, B. A. 2023. *Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?* Cambridge, Cell Press.Available at:https://doi.org/10.1016/j.patter.2023.100694 (Accessed 23 June 2023.)

Bass, D. and Metz, R. 2023. *OpenAI's Sam Altman Urges Congress to Regulate Powerful New Technology*. New York, Bloomberg. Available at: https://www.bloomberg.com/news/ newsletters/2023-05-17/openai-s-sam-altmanurges-congress-to-regulate-powerful-new-aitechnology(Accessed 23 June 2023.)

Bender, E. M., Gebru, T., McMillan-Major, A. and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery. Available at: https://doi.org/10.1145/3442188.3445922 (Accessed 23 June 2023.)

Bommasani, R. et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, Stanford University. Available at: https://crfm.stanford.edu/ report.html (Accessed 23 June 2023.)

Bove, T. 2023. *Big tech is making big AI promises in earnings calls as ChatGPT disrupts the industry: 'You're going to see a lot from us in the coming few months'*. New York, Fortune. Available at: https:// fortune.com/2023/02/03/google-meta-apple-aipromises- chatgpt-earnings (Accessed 3 July 2023.)

Chayka, K. 2023. *My A.I. Writing Report*. New York, The New Yorker. Available at: https://www. newyorker.com/culture/infinite-scroll/my-aiwriting- robot (Accessed 1 August 2023.)

Chen, L., Zaharia, M., and Zou, J. 2023. *How Is ChatGPT's Behavior Changing over Time?* Ithaca, arXiv. Available at: https://arxiv.org/ pdf/2307.09009 (Accessed 31 July 2023.)

Coscarelli, J. 2023. *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*. New York, New York Times. Available at: https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weekndfake.html (Accessed 30 August 2023.)

Cyberspace Administration of China. 2023*a*. 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知 [Notice of the Cyberspace Administration of China on Public Comments on the 'Administrative Measures for Generative Artificial Intelligence Services (Draft for Comment)' ]. Cyberspace Administration of China (CAC), Beijing. (In Chinese.)Available at: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (Accessed 19 July 2023.)

–––––.2023*b*. 生成式人工智能服务管理暂行办法 [Interim Measures for the Management of Generative Artificial Intelligence Services]. Cyberspace Administration of China (CAC), Beijing. (In Chinese.)Available at: http://www.cac.gov. cn/2023-07/13/c_1690898327029107.htm (Accessed 19 July 2023.)

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Horn Nord, J., O' Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J. and Wright, R. 2023. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, Vol. 71. Amsterdam, Elsevier, p. 102642. Available at: https://doi. org/10.1016/j.ijinfomgt.2023.102642 (Accessed 25 August 2023.)

E2Analyst. 2023. *GPT-4: Everything you want to know about OpenAI's new AI model*. San Francisco, Medium. Available at: https://medium.com/ predict/gpt-4-everything-you-want-to-know-aboutopenais- new-ai-model-a5977b42e495 (Accessed 1 August 2023.)

European Commission. 2021. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels, European Commission. Available at: https://artificialintelligenceact.eu (Accessed 23 June 2023.)

European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Brussels, Official Journal of the European Union. Available at: http://data.europa. eu/eli/reg/2016/679/oj (Accessed 23 June 2023.)

Federal Trade Commission. 1998. Children' s Online Privacy Protection Act of 1998. Washington DC, Federal Trade Commission. Available at: https:// www.ftc.gov/legal-library/browse/rules/childrensonline-privacy-protection-rule-coppa (Accessed 4 September 2023.)

Giannini, S. 2023. Generative AI and the Future of Education. Paris, UNESCO. Available at: https:// unesdoc.unesco.org/ark:/48223/pf0000385877 (Accessed 29 August 2023.)

Google. 2023*a*. *Recommendations for Regulating AI*. Mountain View, Google. Available at: https:// ai.google/static/documents/recommendations-forregulating- ai.pdf (Accessed 23 June 2023.)

–––––.2023*b*.PaLM 2 Technical Report. Mountain View, Google. Available at: https://doi. org/10.48550/arXiv.2305.10403 (Accessed on 20 July 2023.)

Lin, B. 2023. *AI Is Generating Security Risks Faster Than Companies Can Keep Up*. New York, The Wall Street Journal. Available at: https://www.wsj.com/ articles/ai-is-generating-security-risks-faster-thancompanies- can-keep-up-a2bdedd4 (Accessed 25 August 2023.)

Marcus, G. 2022. Hoping for the Best as AI Evolves. *Communications of the ACM*, Vol. 66, No. 4. New York, Association for Computing Machinery. Available at: https://doi.org/10.1145/3583078 (Accessed 23 June 2023.)

Marwala, T. 2023. *Algorithm Bias — Synthetic Data Should Be Option of Last Resort When Training AI Systems*. Tokyo, United Nation University. Available at: https://unu.edu/article/algorithm-biassynthetic- data-should-be-option-last-resort-whentraining- ai-systems (Accessed 31 July 2023.)

Metz, C. 2021. *Who Is Making Sure the A.I. Machines Aren't Racist?* New York, The New York Times. Available at: https://www.nytimes. com/2021/03/15/technology/artificial-intelligencegoogle- bias.html (Accessed 23 June 2023.)

Murphy Kelly, S. 2023. *Microsoft is bringing ChatGPT technology to Word, Excel and Outlook*. Atlanta, CNN. Available at: https://edition.cnn. com/2023/03/16/tech/openai-gpt-microsoft-365/ index.html (Accessed 25 August 2023.)

Nazaretsky, T., Cukurova, M. and Alexandron, G. 2022*a*. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22:LAK22: 12th International Learning Analytics and Knowledge Conference*. Vancouver, Association for Computing Machinery, pp. 55-66.

Nazaretsky, T., Ariely, M., Cukurova, M. and Alexandron, G. 2022*b*. Teachers' trust in AIpowered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, Vol. 53, No. 4. Hoboken, NJ, Wiley, pp. 914-931. Available at: https://doi. org/10.1111/bjet.13232 (Accessed 1 August 2023.)

Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Singapore, OpenGov Asia. Available at: https://opengovasia.com/singapore-unveilsai- government-cloud-cluster (Accessed 25 August 2023.)

OpenAI. 2018. *AI and compute. San Francisco*, OpenAI. Available at: https://openai.com/research/ ai-and-compute (Accessed 23 June 2023.)

–––––.2023. *Educator considerations for ChatGPT*. San Francisco, OpenAI. Available at: https:// platform.openai.com/docs/chatgpt-education (Accessed 23 June 2023.)

Popli, N. 2023. *The AI Job That Pays Up to $335K— and You Don't Need a Computer Engineering Background*. New York, TIME USA. Available at: https://time.com/6272103/ai-prompt-engineer-job (Accessed 23 June 2023.)

Roose, K. 2022. *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*. New York, The New York Times. Available at: https://www.nytimes.com/2022/09/02/technology/ai-artificialintelligence-artists.html(Accessed 23 June 2023.)

Russell Group, 2023. *Russell Group principles on the use of generative AI tools in education*. Cambridge, Russell Group. Available at: https://russellgroup. ac.uk/media/6137/rg_ai_principles-final.pdf (Accessed 25 August 2023.)

Stanford University. 2019. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: https://hai.stanford.edu/ai-index-2019 (Accessed 23 June 2023.)

–––––.2023. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: https:// hai.stanford.edu/research/ai-index-2023 (Accessed 23 June 2023.)

The Verge. 2023*a*. *OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'*. Washington DC, Vox Media. Available at: https://www.theverge.com/2023/3/15/23640180/ openai-gpt-4-launch-closed-research-ilya-sutskeverinterview (Accessed 1 August 2023.)

–––––.2023*b*. *OpenAI CEO Sam Altman on GPT-4: 'people are begging to be disappointed and they will be'*. Washington DC, Vox Media. Available at: https://www.theverge.com/23560328/openaigpt- 4-rumor-release-date-sam-altman-interview (Accessed 1 August 2023.)

Tlili, A., Shehata, B., Agyemang Adarkwah, M., Bozkurt, A., Hickey, D. T., Huang, R. and Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, Vol. 10, No. 15. Berlin, Springer. Available at: https://doi. org/10.1186/s40561-023-00237-x (Accessed 23 June 2023.)

UNESCO. 2019. *Beijing Consensus on Artificial Intelligence and Education*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ ark:/48223/pf0000368303 (Accessed 3 July 2023.)

–––––.2022*a*. *Recommendation on the Ethics of Artificial Intelligence*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/ pf0000381137 (Accessed 3 July 2023.)

–––––.2022*b*. *AI and education: guidance for policy-makers*. Paris, UNESCO. Available at: https:// unesdoc.unesco.org/ark:/48223/pf0000376709 (Accessed 23 June 2023.)

–––––.2022*c*. K-12 *AI curricula: a mapping of government-endorsed AI curricula*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ ark:/48223/pf0000380602 (Accessed 20 July 2023.)

–––––.2022*d*. *Guidelines for ICT in education policies and masterplans*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/ pf0000380926 (Accessed 31 July 2023.)

–––––.2023*a*. *Artificial Intelligence: UNESCO calls on all Governments to implement Global Ethical Framework without delay*. Paris, UNESCO. Available at: https://www.unesco.org/en/articles/ artificial-intelligence-unesco-calls-all-governmentsimplement-global-ethical-framework-without (Accessed 3 July 2023.)

–––––.2023*b*. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. Unpublished.

–––––.2023*c*. *Survey for the governmental use of AI as a public good for education*. Unpublished (Submitted to UNESCO).

–––––.2023. *Technology in Education: A tool on whose terms?* Paris, Global Education Monitoring Report Team. Available at: https://unesdoc.unesco. org/ark:/48223/pf0000385723 (Accessed 25 August 2023.)

–––––.2023. *ChatGPT and Artificial Intelligence in Higher Education: Quick start guide*. Caracas, UNESCO International Institute for Higher Education in Latin America and the Caribbean. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000385146 (Accessed 25 August 2023.)

US Copyright Office. 2023. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. *Federal Register*, Vol. 88, No. 51. Washington DC, United States (U.S.) Copyright Office, Library of Congress, pp. 16190- 16194. Available at: https://www.federalregister. gov/d/2023-05321 (Accessed 3 July 2023.)

# 教育與研究之生成式人工智慧應用指引

本出版物旨在支持適當的規範、政策與人才培育之計畫，確保生成式AI成為真正有益且可增進教師、學習者與研究人員能力的工具。文中說明了生成式AI所使用的人工智慧技術，並列出可供大眾取用的GPT模型清單，尤其是已取得開放來源授權的模型。本指引也探討了教育GPT的出現，即由特定資料訓練以達教育目的的生成式AI模型。此外，文中也總結關於生成式AI的一些主要爭議，從數位貧窮的惡化到觀點的同質化、從深度偽造到著作權的問題皆有。基於人文觀點，本指引提出了規範生成式AI工具時可採取的關鍵步驟，包括強制要求保護資料隱私，以及為使用者與生成式AI平台的獨立對話設定年齡限制。為了指引人們在教育與研究領域中正確使用這些工具，本指引主張透過尊重人類自主性且適齡的方式來進行倫理驗證與教學設計。