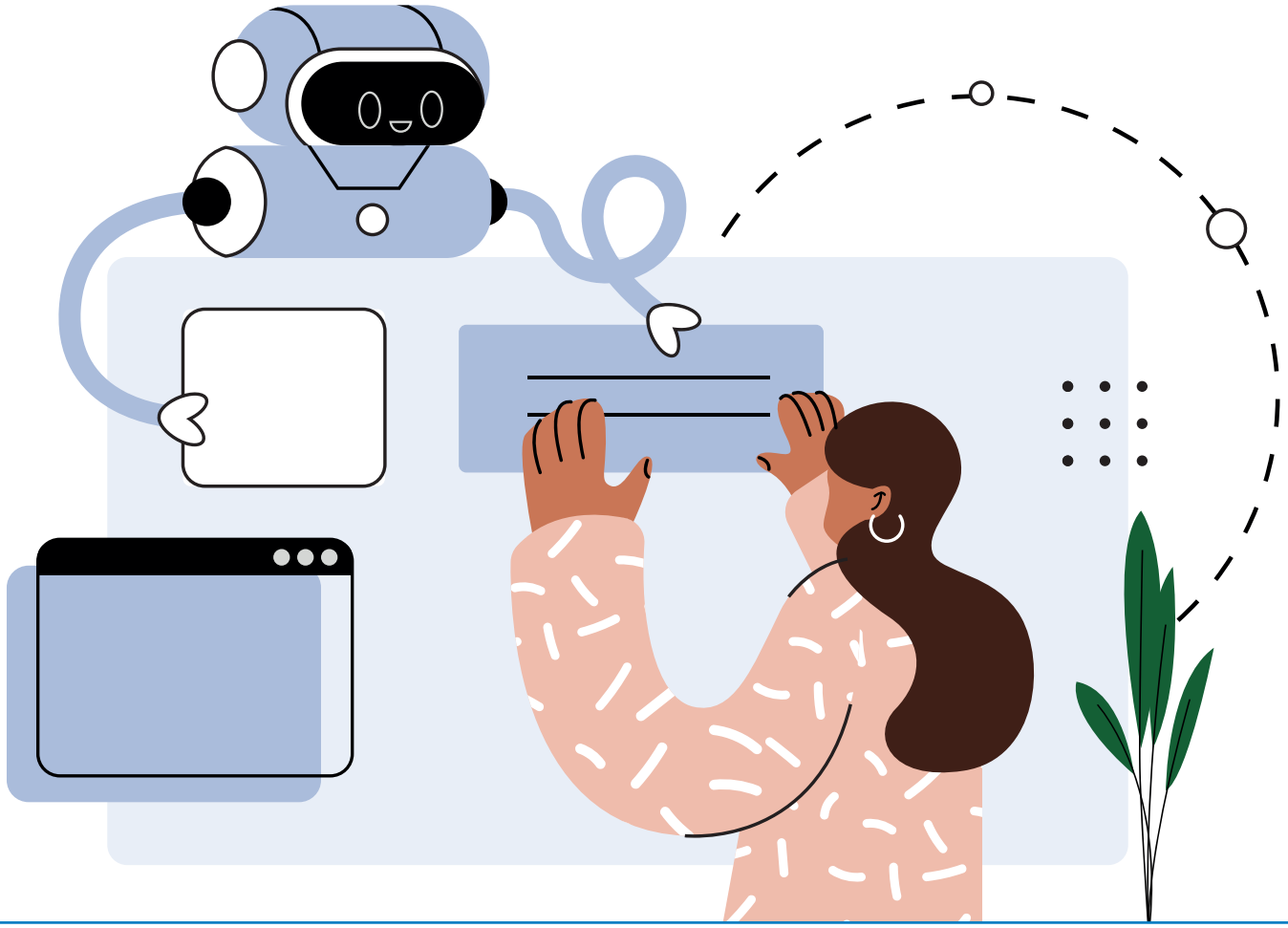
unesco

Hướng dẫn về AI tạo sinh trong giáo dục và nghiên cứu



Giáo dục 2030

UNESCO – lãnh đạo toàn cầu về giáo dục

Giáo dục là ưu tiên hàng đầu của UNESCO vì đó là quyền cơ bản của con người, song cũng là nền tảng của hoà bình và phát triển bền vững. UNESCO là cơ quan chuyên môn của Liên Hợp Quốc về giáo dục, cung cấp định hướng ở cấp độ quốc tế và khu vực nhằm thúc đẩy tiến độ, nâng cao năng lực và sự thích ứng của hệ thống giáo dục các quốc gia nhằm đáp ứng mọi đối tượng người học. Ngoài ra, UNESCO cũng chủ trì các nỗ lực trong việc giải quyết các thách thức đương đại toàn cầu thông qua việc học tập mang tính chuyển hóa, trong đó đặc biệt chú trọng đến bình đẳng giới và Châu Phi trong mọi hoạt động.

unesco

Tổ chức Giáo dục, Khoa học và Văn hóa của Liên Hợp Quốc

Chương trình nghị sự Giáo dục Toàn cầu 2030

Là cơ quan chuyên môn về giáo dục của Liên Hợp Quốc về giáo dục, UNESCO được ủy quyền chủ trì và điều phối thực hiện Chương trình Nghị sự Giáo dục 2030, trong khuôn khổ phong trào quốc tế nhằm xoá bỏ nghèo đói thông qua 17 Mục tiêu Phát triển Bền vững đến năm 2030. Giáo dục được coi là yếu tố thiết yếu để đạt được tất cả các mục tiêu này, và do đó có riêng Mục tiêu 4 hướng tới "***đảm bảo một nền giáo dục hòa nhập, công bằng, chất lượng và thúc đẩy cơ hội HTSĐ cho mọi người.***" Khung Hành động Giáo dục đến năm 2030 đưa ra những định hướng cho việc thực hiện mục tiêu và các cam kết tham vọng này.

Giáo dục 2030

Xuất bản năm 2025 bởi Tổ chức Giáo dục, Khoa học và Văn hoá của Liên Hợp Quốc.
Số 7, place de Fontenoy, 75352 Paris 07 SP, Pháp.

© UNESCO 2025

ISBN 978-92-3-000249-7

DOI https://doi.org/10.54675/KMQF4065

Ấn phẩm được đăng tải trên trang Truy cập Mở theo giấy phép Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (https://creativecommons.org/licenses/by-sa/3.0/igo/). Việc sử dụng ấn phẩm này cần tuân theo các điều khoản sử dụng của Kho Lưu trữ Truy cập mở UNESCO Open Access Repository (https://www.unesco.org/en/open-access/cc-sa).

Những hình ảnh được đánh dấu hoa thị (*) không thuộc giấy phép CC-BY-SA và không được phép sử dụng hoặc sao chép mà không có sự cho phép trước của chủ sở hữu bản quyền.

Nguyên tác: *Guidance for generative AI in education and research*
Xuất bản năm 2023 bởi Tổ chức Giáo dục, Khoa học và Văn hoá của Liên Hợp Quốc

Các thuật ngữ được sử dụng và cách trình bày tài liệu trong suốt ấn phẩm này không hàm ý thể hiện bất kỳ quan điểm nào của UNESCO về tình trạng pháp lý của bất kỳ quốc gia, vùng lãnh thổ, thành phố hay khu vực nào hoặc của các cơ quan chức năng, cũng như về việc phân định biên giới hay ranh giới của các quốc gia, vùng lãnh thổ đó.

Các ý tưởng và quan điểm được trình bày trong ấn phẩm này là của các tác giả; chúng không nhất thiết phản ánh quan điểm của UNESCO và không ràng buộc Tổ chức này.

Người dịch: TS. Ngô Di Lân
Người hiệu đính: Nguyễn Văn Giang
Thiết kế bìa: Olexandra Simkina/Shutterstock.com*

Thiết kế bởi UNESCO

In bởi Công ty TNHH Thương Mại và Dịch vụ Hiệp Nghĩa Việt Nam

In tại Việt Nam

TÓM TẮT

Hướng tới cách tiếp cận lấy con người làm trung tâm trong việc sử dụng AI tạo sinh

Các công cụ AI tạo sinh (GenAI) có thể tiếp cận công khai đang xuất hiện nhanh chóng, và việc phát hành các phiên bản cập nhật hơn đang vượt quá tốc độ thích ứng của các khung pháp lý quốc gia. Sự thiếu vắng các quy định quốc gia về GenAI ở hầu hết các quốc gia khiến quyền riêng tư dữ liệu của người dùng không được bảo vệ và các cơ sở giáo dục phần lớn chưa được chuẩn bị để kiểm chứng các công cụ này.

Hướng dẫn toàn cầu đầu tiên của UNESCO về GenAI trong giáo dục đặt ra mục tiêu hỗ trợ các quốc gia triển khai các biện pháp ngay tức thì, lập kế hoạch chính sách dài hạn và phát triển năng lực con người để đảm bảo một tầm nhìn lấy con người làm trung tâm đối với những công nghệ mới này.

Hướng dẫn này đưa ra đánh giá về những rủi ro tiềm ẩn mà GenAI có thể gây ra đối với các giá trị nhân văn cốt lõi, vốn giúp thúc đẩy quyền tự chủ của con người, sự hòa nhập, công bằng, bình đẳng giới, và đa dạng ngôn ngữ và văn hóa, cũng như các ý kiến và biểu đạt đa dạng.

Hướng dẫn đề xuất các bước chủ đạo để các cơ quan chính phủ điều chỉnh việc sử dụng các công cụ GenAI, bao gồm việc yêu cầu bắt buộc bảo vệ quyền riêng tư dữ liệu và xem xét giới hạn độ tuổi sử dụng. Nó đề ra các yêu cầu đối với các nhà cung cấp GenAI nhằm đảm bảo sử dụng có đạo đức và hiệu quả trong giáo dục.

Hướng dẫn nhấn mạnh sự cần thiết của việc các cơ sở giáo dục kiểm chứng mức độ phù hợp về đạo đức và sư phạm của các hệ thống GenAI trong với giáo dục. Nó kêu gọi cộng đồng quốc tế phản tư về những tác động lâu dài của chúng đối với kiến thức, giảng dạy, học tập và đánh giá.

Trong khi **ChatGPT** đạt **100 triệu người dùng** hoạt động hàng tháng vào tháng 1 năm 2023, chỉ có **một quốc gia** đã ban hành quy định **về AI tạo sinh** tính đến tháng 7 2023

Ấn phẩm này đưa ra các khuyến nghị cụ thể cho các nhà hoạch định chính sách và cơ sở giáo dục về cách đảm bảo các công cụ GenAI được sử dụng để bảo vệ quyền tự chủ của con người và thực sự mang lại lợi ích cho người học, nhà giáo và nhà nghiên cứu.

unesco

"Chiến tranh nảy sinh trong tâm thức của con người, vậy ngay trong tâm thức của con người cần phải dựng lên thành trì của hòa bình"

Hướng dẫn về AI tạo sinh trong giáo dục và nghiên cứu

Lời tựa

© UNESCO

AI tạo sinh (GenAI) đã có màn ra mắt ấn tượng vào cuối năm 2022 với sự ra đời của ChatGPT, ứng dụng có tốc độ tăng trưởng nhanh nhất trong lịch sử. Với khả năng bắt chước các năng lực của con người để tạo ra các sản phẩm như văn bản, hình ảnh, video, âm nhạc và phần mềm, các ứng dụng GenAI này đã gây ra một cơn sốt toàn cầu. Hàng triệu người hiện đang sử dụng GenAI trong cuộc sống hàng ngày và tiềm năng của việc áp dụng các mô hình này cho các ứng dụng AI chuyên biệt dường như là không giới hạn.

Năng lực xử lý thông tin và tạo ra kiến thức sâu rộng cho phép mô phỏng tư duy bậc cao vốn là nên nền tảng của việc học tập của con người sẽ có những hàm ý tiềm tàng to lớn đối với giáo dục. Khi các công cụ GenAI ngày càng có khả năng tự động hóa một số mức độ cơ bản của việc viết và sáng tạo nghệ thuật, chúng buộc các nhà hoạch định chính sách giáo dục và các tổ chức phải xem xét lại lý do tại sao chúng ta học, chúng ta đang học cái gì và như thế nào. Đây hiện là những điểm chủ chốt phải cân nhắc trong giáo dục trong giai đoạn mới này của kỷ nguyên số.

Ấn phẩm này mong muốn hỗ trợ việc lập kế hoạch cho các quy định, chính sách và phát triển năng lực con người phù hợp, nhằm đảm bảo rằng GenAI trở thành một công cụ thực sự mang lại lợi ích và trao quyền cho nhà giáo, người học và nhà nghiên cứu.

Nó đề xuất các bước chính để các cơ quan chính phủ điều chỉnh việc sử dụng AI tạo sinh. Nó cũng đề xuất các khuôn khổ và ví dụ cụ thể cho việc xây dựng chính sách và thiết kế sư phạm cho phép sử dụng công nghệ này một cách có đạo đức và hiệu quả trong giáo dục. Cuối cùng, nó kêu gọi cộng đồng quốc tế xem xét những hàm ý lâu dài sâu sắc của AI tạo sinh đối với cách chúng ta hiểu kiến thức và xác định nội dung học tập, phương pháp và kết quả, cũng như cách chúng ta đánh giá và kiểm chứng quá trình học tập.

Kế thừa *Khuyến nghị về Đạo đức AI (Recommendation on the Ethics of Artificial Intelligence)* năm 2021 của UNESCO, Hướng dẫn này dựa trên nền tảng là cách tiếp cận nhân văn đối với giáo dục, thúc đẩy quyền tự chủ của con người, sự hòa nhập, công bằng, bình đẳng giới, và đa dạng văn hóa và ngôn ngữ, cũng như các ý kiến và biểu đạt đa dạng. Hơn nữa, nó đáp ứng lời kêu gọi của báo cáo năm 2021 của Ủy ban Quốc tế về Tương lai của Giáo dục, "*Cùng nhau tưởng tượng lại tương lai của chúng ta: Một khế ước xã hội mới cho giáo dục*" (Reimagining our futures together: A new social contract for education) để định nghĩa lại mối quan hệ của chúng ta với công nghệ, như một phần không thể thiếu trong nỗ lực của chúng ta để đổi mới khế ước xã hội cho giáo dục.

AI không được soán ngôi trí thông minh của con người. Thay vào đó, nó đòi hỏi chúng ta xem xét lại những hiểu biết đã sẵn có về tri thức và việc học của con người. Tôi hy vọng rằng Hướng dẫn này sẽ giúp chúng ta xác định lại viễn cảnh tương lai cho giáo dục và đặt nền móng cho tư duy và nỗ lực tập thể mà có thể dẫn đến tương lai học tập số lấy con người làm trung tâm cho mọi người.

Stefania Giannini,
Trợ lý Tổng Giám đốc UNESCO về Giáo dục

Lời cảm ơn

Dưới sự lãnh đạo của Stefania Giannini, Trợ lý Tổng Giám đốc về Giáo dục, và sự hướng dẫn của Sobhi Tawil, Giám đốc Bộ phận Tương lai của Học tập và Đổi mới tại UNESCO, việc dự thảo ấn phẩm này được dẫn dắt bởi Fengchun Miao, Trưởng đơn vị Công nghệ và AI trong Giáo dục.

Xin gửi lời cảm ơn đặc biệt tới Wayne Holmes, Phó Giáo sư tại Đại học College London, người đã đồng dự thảo ấn phẩm này.

Ấn phẩm này là kết quả của nỗ lực tập thể của các nhà lãnh đạo giáo dục và chuyên gia trong lĩnh vực AI và giáo dục.

Nó được kết tinh từ những hiểu biết sâu sắc và đóng góp của nhiều chuyên gia bao gồm: Mutlu Cukurova, Giáo sư tại Đại học College London; Colin de la Higuera, Chủ tịch UNESCO về Công nghệ để Bồi dưỡng Giáo viên với Tài nguyên Giáo dục Mở tại Đại học Nantes; Shafika Isaacs, Nghiên cứu viên tại Đại học Johannesburg; Natalie Lao, Giám đốc điều hành của App Inventor Foundation; Qin Ni, Phó Giáo sư tại Đại học Sư phạm Thượng Hải; Catalina Nicolin, Chuyên gia CNTT trong Giáo dục tại Trung tâm Giáo dục Số châu Âu ở Romania; John Shaw-Taylor, Chủ tịch UNESCO về AI và Giáo sư Thống kê Máy tính và Học Máy tại Đại học College London; Kelly Shirohira, Giám đốc điều hành tại Jet Education Services; Ki-Sang Song, Giáo sư tại Đại học Giáo dục Quốc gia Hàn Quốc; và Ilkka Tuomi, Trưởng nhóm khoa học tại Meaning Processing Ltd ở Phần Lan.

Nhiều đồng nghiệp ở UNESCO cũng đã đóng góp theo nhiều cách khác nhau bao gồm: Dafna Feinholz, Trưởng Bộ phận Đạo đức Sinh học và Đạo đức của Khoa học và Công nghệ; Francesc Pedró, Giám đốc Viện Giáo dục Đại học Quốc tế ở Mỹ Latinh và Caribe; Prateek Sibal, Chuyên gia Chương trình, Bộ phận Chính sách Số và Chuyển đổi Số; Saurabh Roy, Cán bộ Dự án Cao cấp tại Bộ phận Phát triển Giáo viên, Bộ phận Chính sách và Hệ thống Học tập Suốt đời; Benjamin Vergel De Dios, Chuyên gia Chương trình về CNTT trong Giáo dục, Bộ phận Đổi mới Giáo dục và Phát triển Kỹ năng tại Văn phòng Bangkok; các đồng nghiệp trong Đơn vị Đa dạng Biểu đạt Văn hóa thuộc Lĩnh vực Văn hóa; và Mark West, Chuyên gia Chương trình, Bộ phận Tương lai của Học tập và Đổi mới.

Xin gửi lời cảm ơn đến Glen Hertelendy, Luisa Ferrara và Xianglei Zheng, Đơn vị Công nghệ và AI trong Giáo dục, Bộ phận Tương lai của Học tập và Đổi mới, vì đã điều phối việc sản xuất ấn phẩm này.

Cũng xin gửi lời cảm ơn tới Jenny Webster vì đã biên tập và hiệu đính văn bản, và Trần Thúy Ngọc vì đã thiết kế bố cục.

Mục lục

Lời tựa **2**

Lời cảm ơn **3**

Danh mục các từ viết tắt **6**

Giới thiệu **7**

1. AI tạo sinh là gì và nó hoạt động như thế nào? **8**

1.1 AI tạo sinh là gì? 8

1.2 AI tạo sinh hoạt động như thế nào? 8

1.2.1 Các mô hình GenAI văn bản hoạt động như thế nào 9

1.2.2 Các mô hình GenAI hình ảnh hoạt động như thế nào 11

1.3 Kỹ thuật soạn câu lệnh để tạo ra các kết quả như mong muốn 11

1.4 EdGPT mới nổi và hàm ý của nó 13

2. Những tranh cãi xung quanh AI tạo sinh và hàm ý của chúng đối với giáo dục **14**

2.1 Làm trầm trọng thêm tình trạng nghèo kỹ thuật số 14

2.2 Vượt quá khả năng thích ứng của các quy định quốc gia 14

2.3 Sử dụng nội dung mà chưa có sự đồng ý 15

2.4 Các mô hình được sử dụng nhưng không giải thích được 15

2.5 Nội dung do AI tạo ra làm ô nhiễm Internet 16

2.6 Thiếu hiểu biết về thế giới thực 16

2.7 Làm giảm mức độ đa dạng quan điểm và suy yếu những tiếng nói thiểu số 17

2.8 Tạo ra deepfake thật hơn 17

3. Quy định việc sử dụng AI tạo sinh trong giáo dục **18**

3.1 Cách tiếp cận AI lấy con người làm trung tâm 18

3.2 Các bước để quản trị GenAI trong giáo dục 18

3.3 Quy định về GenAI: Các yếu tố chính 20

3.3.1 Các cơ quan quản lý chính phủ 20

3.3.2 Nhà cung cấp công cụ GenAI 21

3.3.3 Người dùng là tổ chức 23

3.3.4 Người dùng cá nhân 23

4. Hướng tới một khung chính sách cho việc sử dụng AI tạo sinh trong giáo dục và nghiên cứu **24**

4.1 Thúc đẩy sự hòa nhập, công bằng và đa dạng ngôn ngữ và văn hóa 24

4.2 Bảo vệ quyền tự chủ của con người 24

4.3 Giám sát và kiểm chứng các hệ thống GenAI cho giáo dục 25

4.4 Phát triển năng lực AI bao gồm các kỹ năng liên quan đến GenAI cho người học 25
4.5 Phát triển năng lực cho nhà giáo và nhà nghiên cứu nhằm sử dụng GenAI đúng cách 26
4.6 Thúc đẩy ý kiến đa dạng và biểu đạt đa dạng của ý tưởng 26
4.7 Thử nghiệm các mô hình ứng dụng phù hợp cho từng địa phương và xây dựng cơ sở bằng chứng tích lũy 27
4.8 Xem xét các tác động lâu dài theo cách liên ngành và liên lĩnh vực 27
5. Thúc đẩy sử dụng GenAI theo cách sáng tạo trong giáo dục và nghiên cứu 28
5.1 Chiến lược của tổ chức nhằm thúc đẩy sử dụng GenAI có trách nhiệm và sáng tạo 28
5.2 Cách tiếp cận sư phạm 'lấy con người làm trung tâm và tương tác phù hợp' 29
5.3 Đồng thiết kế việc sử dụng GenAI trong giáo dục và nghiên cứu 29
5.3.1 AI tạo sinh cho nghiên cứu 29
5.3.2 AI tạo sinh để hỗ trợ giảng dạy 30
5.3.3 AI tạo sinh như một huấn luyện viên 1:1 để tự học các kỹ năng nền tảng 31
5.3.4 AI tạo sinh để thúc đẩy học tập dựa trên tra cứu hoặc dự án 33
5.3.5 AI tạo sinh để hỗ trợ người học có nhu cầu đặc biệt 34
6. GenAI và tương lai của giáo dục và nghiên cứu 36
6.1 Các vấn đề đạo đức chưa được khai phá 36
6.2 Bản quyền và sở hữu trí tuệ 36
6.3 Nguồn nội dung và học tập 36
6.4 Câu trả lời đồng nhất so với kết quả đầu ra đa dạng và sáng tạo 37
6.5 Tư duy lại về đánh giá và chuẩn đầu ra 37
6.6 Quy trình tư duy 37
Kết luận 38
Tài liệu tham khảo 39

Danh mục bảng biểu

Bảng 1. Các kỹ thuật được sử dụng trong AI tạo sinh 8
Bảng 2. Mô hình GPT của OpenAI 9
Bảng 3. Các trường hợp sử dụng GenAI đồng thiết kế cho nghiên cứu 30
Bảng 4. Các trường hợp sử dụng GenAI đồng thiết kế nhằm hỗ trợ nhà giáo và giảng dạy 31
Bảng 5. Thử nghiệm các mô hình ứng dụng phù hợp cho từng địa phương và xây dựng cơ sở bằng chứng tích lũy 32
Bảng 6. Các trường hợp sử dụng GenAI đồng thiết kế nhằm thúc đẩy học tập dựa trên tra cứu hoặc dự án 33
Bảng 7. Các trường hợp sử dụng GenAI đồng thiết kế nhằm hỗ trợ người học có nhu cầu đặc biệt 34

Danh mục các từ viết tắt

Các khái niệm và công nghệ

AGI	Trí tuệ nhân tạo tổng quát
AI	Trí tuệ nhân tạo
API	Giao diện lập trình ứng dụng
ANN	Mạng nơ-ron nhân tạo
DAI	Trí tuệ nhân tạo phân tán
GAN	Mạng đối kháng tạo sinh
GB	Gigabyte
GDPR	Quy định chung về bảo vệ dữ liệu
GenAI	Trí tuệ nhân tạo tạo sinh
GPT	Bộ chuyển đổi tạo sinh được huấn luyện trước
ICT	Công nghệ thông tin và truyền thông
LaMDA	Mô hình ngôn ngữ cho các ứng dụng đối thoại
LLM	Mô hình ngôn ngữ lớn
ML	Học máy
VAE	Bộ mã hóa tự động biến phân

Các tổ chức

AGCC	Cụm Điện toán Đám mây AI Chính phủ (Xinh-ga-po)
CAC	Cục Quản lý Không gian mạng Trung Quốc
EU	Liên minh Châu Âu
OECD	Tổ chức Hợp tác và Phát triển Kinh tế
UNCTAD	Hội nghị Liên Hợp Quốc về Thương mại và Phát triển
UNESCO	Tổ chức Giáo dục, Khoa học và Văn hóa Liên Hợp Quốc

Giới thiệu

Việc công bố ChatGPT vào cuối năm 2022, công cụ AI tạo sinh (GenAI) dễ sử dụng đầu tiên được phổ biến rộng rãi cho công chúng,[1] theo sau là các phiên bản tinh vi hơn, đã gây chấn động trên toàn thế giới và thúc đẩy một cuộc đua giữa các công ty công nghệ lớn để định vị mình trong lĩnh vực phát triển mô hình GenAI.[2]

Trên toàn thế giới, mối quan ngại ban đầu trong giáo dục là ChatGPT và các công cụ GenAI tương tự sẽ được người học sử dụng để gian lận trong bài tập của các em, từ đó làm giảm đi giá trị của đánh giá học tập, các chứng nhận và bằng cấp (Anders, 2023). Trong khi một số cơ sở giáo dục đã cấm sử dụng ChatGPT, những cơ sở khác lại thận trọng chào đón sự xuất hiện của GenAI (Tlili, 2023). Chẳng hạn, nhiều trường học và đại học đã chọn cách tiếp cận cởi mở, tin rằng 'thay vì tìm cách cấm sử dụng chúng, người học và nhân viên cần được hỗ trợ trong việc sử dụng các công cụ GenAI một cách hiệu quả, có đạo đức và minh bạch' (Russell Group, 2023). Cách tiếp cận này thừa nhận rằng GenAI được sử dụng rộng rãi, nhiều khả năng chỉ trở nên tinh vi hơn, và có cả một số tác động tiêu cực lẫn tiềm năng tích cực vô song cho giáo dục.

Đúng là GenAI có vô số cách sử dụng tiềm năng. Nó có thể tự động hóa việc xử lý thông tin và trình bày kết quả trên tất cả các biểu đạt ký hiệu chính của tư duy con người. Nó cho phép tạo ra các sản phẩm hoàn chỉnh bằng hoàn thiện các sản phẩm tri thức bán thành phẩm. Bằng cách giải phóng con người khỏi một số loại kỹ năng tư duy bậc thấp, thế hệ công cụ AI mới này có thể có những hàm ý sâu sắc đối với cách chúng ta hiểu về trí thông minh và quá trình học tập của con người.

Tuy nhiên, GenAI cũng làm dấy lên nhiều mối quan ngại trước mắt về các vấn đề như an toàn, quyền riêng tư dữ liệu, bản quyền và nguy cơ chỉnh sửa. Một số trong số này là những rủi ro chung liên quan đến AI mà đã bị GenAI làm trầm trọng thêm, trong khi những rủi ro khác mới xuất hiện với thế hệ công cụ mới nhất này. Giờ đây, việc hiểu đầy đủ và giải quyết từng vấn đề và mối quan ngại này là ưu tiên cấp bách.

Hướng dẫn này được thiết kế để đáp ứng nhu cầu cấp thiết này. Tuy nhiên, một bộ hướng dẫn chuyên đề về GenAI cho giáo dục không nên được xem như một tuyên bố rằng GenAI là giải pháp cho những thách thức cơ bản của giáo dục. Bất chấp sự phóng đại của truyền thông, GenAI một mình không có khả năng giải quyết bất kỳ vấn đề nào mà các hệ thống giáo dục trên thế giới đang phải đối mặt. Trong việc đáp ứng các vấn đề giáo dục tồn tại lâu dài, điều quan trọng là phải kiên định rằng năng lực con người và nỗ lực tập thể, chứ không phải công nghệ, mới là yếu tố quyết định trong các giải pháp hiệu quả cho những thách thức cơ bản mà xã hội đang phải đối mặt.

Do đó, Hướng dẫn này đặt ra mục tiêu hỗ trợ việc dự thảo các quy định, chính sách và chương trình phát triển năng lực con người phù hợp, để đảm bảo rằng GenAI trở thành một công cụ thực sự mang lại lợi ích và trao quyền cho nhà giáo, người học và nhà nghiên cứu. Dựa trên *Khuyến nghị về Đạo đức AI của UNESCO*, Hướng dẫn được xây dựng dựa trên cách tiếp cận lấy con người làm trung tâm, thúc đẩy quyền tự chủ của con người, sự hòa nhập, công bằng, bình đẳng giới, và đa dạng văn hóa và ngôn ngữ, cũng như các ý kiến và biểu đạt đa dạng.

Hướng dẫn trước tiên xem xét GenAI là gì và nó hoạt động như thế nào, và chỉ ra các công nghệ và mô hình đa dạng hiện có (Phần 1), trước khi xác định một loạt các vấn đề đạo đức và chính sách gây tranh cãi xung quanh cả AI nói chung và GenAI nói riêng (Phần 2). Tiếp theo là một cuộc thảo luận về các bước và các yếu tố chính cần được xem xét khi tìm cách quản trị GenAI dựa trên cách tiếp cận lấy con người làm trung tâm - một cách tiếp cận đảm bảo việc sử dụng có đạo đức, an toàn, công bằng và có ý nghĩa (Phần 3). Phần 4 sau đó đề xuất các biện pháp có thể được triển khai để phát triển các khung chính sách mạch lạc, toàn diện để điều chỉnh việc sử dụng GenAI trong giáo dục và nghiên cứu, trong khi Phần 5 xem xét các khả năng sử dụng GenAI một cách sáng tạo trong thiết kế chương trình giảng dạy, giảng dạy, học tập và các hoạt động nghiên cứu. Phần 6 kết thúc Hướng dẫn với những cân nhắc về tác động lâu dài của GenAI đối với giáo dục và nghiên cứu.

1. AI tạo sinh là gì và nó hoạt động như thế nào?

1.1 AI tạo sinh là gì?

AI tạo sinh (GenAI) là một công nghệ trí tuệ nhân tạo (AI) tự động tạo ra nội dung dựa trên các câu lệnh được soạn bằng ngôn ngữ tự nhiên trên giao diện hội thoại. Thay vì chỉ sắp xếp các trang web hiện có, bằng cách dựa vào nội dung sẵn có, GenAI thực sự tạo ra nội dung mới. Nội dung có thể xuất hiện dưới các định dạng bao gồm tất cả các biểu đạt ký hiệu của tư duy con người: văn bản viết bằng ngôn ngữ tự nhiên, hình ảnh (bao gồm ảnh, tranh vẽ số và phim hoạt hình), video, âm nhạc và mã nguồn phần mềm. GenAI được huấn luyện bằng cách sử dụng dữ liệu thu thập từ các trang web, cuộc trò chuyện trên mạng xã hội và các phương tiện truyền thông trực tuyến khác. Nó tạo ra nội dung bằng cách phân tích thống kê phân phối của từ, điểm ảnh hoặc các yếu tố khác trong dữ liệu mà nó đã tiếp nhận, rồi xác định và lặp lại các mẫu phổ biến (ví dụ, những từ nào thường theo sau những từ khác).

Mặc dù GenAI có thể tạo ra nội dung mới, nhưng nó không thể tạo ra ý tưởng mới hoặc giải pháp cho các thách thức trong thế giới thực, vì nó không hiểu các đối tượng thế giới thực hoặc các mối quan hệ xã hội đằng sau ngôn ngữ. Hơn nữa, mặc dù kết quả đầu ra của nó lưu loát và ấn tượng, GenAI không có độ chính xác đáng tin cậy. Thật vậy, ngay cả nhà cung cấp của ChatGPT cũng thừa nhận, 'Mặc dù các công cụ như ChatGPT thường có thể tạo ra câu trả lời nghe có vẻ hợp lý, nhưng không thể tin rằng chúng luôn chính xác.' (OpenAI, 2023). Thông thường thì các lỗi sẽ không được nhận ra trừ khi người dùng có kiến thức vững vàng về chủ đề đang được đề cập.

1.2 AI tạo sinh hoạt động như thế nào?

Các công nghệ cụ thể đằng sau GenAI là một phần của nhóm các công nghệ AI được gọi là học máy (ML), sử dụng các thuật toán để cho phép nó liên tục và tự động cải thiện hiệu suất từ dữ liệu. Loại ML đã dẫn đến nhiều tiến bộ trong AI mà chúng ta đã thấy trong những năm gần đây, chẳng hạn như việc sử dụng AI để nhận dạng khuôn mặt, được gọi là mạng nơ-ron nhân tạo (ANN), được lấy cảm hứng từ cách bộ não con người hoạt động và các kết nối synap giữa các nơ-ron. Có nhiều loại ANN khác nhau.

Cả công nghệ GenAI văn bản và hình ảnh đều dựa trên một tập hợp các công nghệ AI đã có sẵn đối với các nhà nghiên cứu trong nhiều năm. Chẳng hạn, ChatGPT sử dụng một bộ chuyển đổi tạo sinh được huấn luyện trước (GPT), trong khi GenAI hình ảnh thường sử dụng cái được gọi là mạng đối kháng tạo sinh (GAN) (xem **Bảng 1**).[3]

Bảng 1. Các kỹ thuật được sử dụng trong AI tạo sinh

Học máy (ML)		Một loại AI sử dụng dữ liệu để tự động cải thiện hiệu suất của nó.
Mạng nơ-ron nhân tạo (ANN)		Một loại ML được lấy cảm hứng từ cấu trúc và chức năng của bộ não người (ví dụ: các kết nối synap giữa các nơ-ron).
AI tạo sinh văn bản	**Bộ chuyển đổi đa năng**	Một loại ANN có khả năng tập trung vào các phần khác nhau của dữ liệu để xác định cách chúng liên quan với nhau.
	Mô hình ngôn ngữ lớn (LLM)	Một loại bộ chuyển đổi đa năng được huấn luyện trên lượng lớn dữ liệu văn bản.
	Bộ chuyển đổi tạo sinh được huấn luyện trước (GPT)[4]	Một loại LLM được huấn luyện trước trên lượng dữ liệu còn lớn hơn, cho phép mô hình nắm bắt các sắc thái của ngôn ngữ và tạo ra văn bản vừa mạch lạc, vừa ý thức được ngữ cảnh.
AI tạo sinh hình ảnh	**Mạng đối kháng tạo sinh (GAN)**	Các loại mạng nơ-ron được sử dụng để tạo ra hình ảnh.
	Bộ mã hóa tự động biến phân (VAE)	

1.2.1. Các mô hình GenAI văn bản hoạt động như thế nào

GenAI văn bản sử dụng một loại ANN được gọi là bộ chuyển đổi đa năng, và một loại bộ chuyển đổi đa năng được gọi là mô hình ngôn ngữ lớn. Đây là lý do tại sao các hệ thống GenAI văn bản thường được gọi là mô hình ngôn ngữ lớn, hay LLM. Dạng LLM được sử dụng bởi GenAI văn bản còn được biết đến là "bộ chuyển đổi tạo sinh được huấn luyện trước", hay GPT (vậy nên mới có chữ 'GPT' trong 'ChatGPT').

ChatGPT được xây dựng trên GPT-3 do OpenAI phát triển. Đây là phiên bản thứ ba của GPT của họ, phiên bản đầu tiên được ra mắt vào năm 2018 và phiên bản gần đây nhất, GPT-4, vào tháng 3 năm 2023 (xem **Bảng 2**). Mỗi OpenAI GPT liên tục cải thiện so với phiên bản trước thông qua những tiến bộ trong kiến trúc AI, phương pháp huấn luyện và kỹ thuật tối ưu hóa. Một khía cạnh được biết đến rộng rãi của sự tiến bộ không ngừng của công cụ này là việc sử dụng lượng dữ liệu ngày càng lớn để tăng số lượng 'tham số' theo cấp số nhân. Các tham số có thể được coi là các 'núm vặn' có thể được tinh chỉnh để cải thiện hiệu suất của mô hình GPT. Chúng bao gồm 'trọng số' của mô hình, các tham số xác định cách mô hình xử lý đầu vào và tạo ra kết quả đầu ra.

Ngoài những tiến bộ trong việc tối ưu hóa kiến trúc AI và phương pháp huấn luyện, quá trình lặp lại nhanh chóng này cũng được triển khai nhờ lượng dữ liệu khổng lồ và những cải tiến về khả năng tính toán có sẵn cho các công ty lớn. Kể từ năm 2012, khả năng tính toán được sử dụng để huấn luyện các mô hình GenAI đã tăng gấp đôi cứ sau 3-4 tháng. So sánh với Định luật Moore có chu kỳ tăng gấp đôi sau 2 năm (OpenAI, 2018; Đại học Stanford, 2019).

Bảng 2. Các mô hình GPT của OpenAI

Mô hình	Năm công bố	Lượng dữ liệu sử dụng để huấn luyện	Số lượng tham số	Đặc điểm
GPT-1	2018	40 GB	117 triệu	Có khả năng triển khai các nhiệm vụ xử lý ngôn ngữ tự nhiên như hoàn thành văn bản và trả lời câu hỏi.
GPT-2	2019	40 GB	1.500 triệu	Có khả năng triển khai các nhiệm vụ xử lý ngôn ngữ tự nhiên phức tạp hơn như dịch máy và tóm tắt.
GPT-3	2020	17.000 GB	175.000 triệu	Có khả năng triển khai các nhiệm vụ xử lý ngôn ngữ tự nhiên nâng cao như viết đoạn văn mạch lạc và tạo ra toàn bộ bài viết. Cũng có khả năng thích ứng với các nhiệm vụ mới chỉ với một vài ví dụ.
GPT-4[6]	2023	1.000.000 GB (được báo cáo nhưng chưa được xác nhận)	170.000.000 triệu (được báo cáo nhưng chưa được xác nhận)	Độ tin cậy được nâng cao và có khả năng xử lý các hướng dẫn phức tạp hơn

Sau khi mô hình GPT đã được huấn luyện, việc soạn thảo văn bản dựa trên câu lệnh sẽ gồm các bước sau:

1. Câu lệnh được chia thành các đơn vị nhỏ hơn (gọi là token) được đưa vào mô hình GPT.
2. Mô hình GPT sử dụng các mẫu thống kê để dự đoán các từ hoặc cụm từ có khả năng tạo thành câu trả lời mạch lạc cho lời nhắc.
   - Mô hình GPT xác định các mẫu từ và cụm từ thường xuất hiện cùng nhau trong mô hình dữ liệu lớn được xây dựng sẵn của nó (bao gồm văn bản được thu thập từ Internet và những nơi khác).
   - Sử dụng các mẫu này, mô hình GPT ước tính xác suất xuất hiện của các từ hoặc cụm từ cụ thể trong một ngữ cảnh nhất định.
   - Bắt đầu với một dự đoán ngẫu nhiên, mô hình GPT sử dụng các xác suất được ước tính này để dự đoán từ hoặc cụm từ tiếp theo có khả năng xuất hiện trong câu trả lời của nó.
3. Các từ hoặc cụm từ được dự đoán được chuyển đổi thành văn bản có thể đọc được.
4. Văn bản có thể đọc được sau đó được đưa qua cái được gọi là 'bộ lọc' để loại bỏ bất kỳ nội dung độc hại nào.
5. Các bước 2 đến 4 được lặp lại cho đến khi hoàn thành câu trả lời. Câu trả lời được coi là hoàn thành khi nó đạt đến giới hạn token tối đa hoặc đáp ứng các tiêu chí dừng được xác định trước.

6. Câu trả lời sẽ được xử lý hậu kỳ để cải thiện khả năng đọc bằng cách áp dụng định dạng, dấu câu và các cải tiến khác (chẳng hạn như bắt đầu câu trả lời bằng những từ mà con người có thể sử dụng, như 'Chắc chắn', 'Dĩ nhiên' hoặc 'Tôi xin lỗi').

Mặc dù các nhà nghiên cứu đã có thể dùng thử mô hình GPT và khả năng tự động tạo văn bản từ năm 2018, điều làm cho việc ra mắt ChatGPT trở nên mới mẻ là khả năng truy cập miễn phí thông qua một giao diện dễ sử dụng, có nghĩa là bất kỳ ai có truy cập Internet đều có thể khám phá công cụ này.

Đến tháng 7 năm 2023, một số lựa chọn thay thế cho ChatGPT bao gồm:

- **Alpaca:**[7] Một phiên bản tinh chỉnh của Llama của Meta, từ Đại học Stanford, nhằm giải quyết thông tin sai lệch của LLM, định kiến xã hội và ngôn ngữ độc hại.
- **Bard:**[8] Một LLM từ Google, dựa trên hệ thống LaMDA và PaLM 2, có quyền truy cập vào Internet theo thời gian thực, có nghĩa là nó có thể cung cấp thông tin cập nhật nhất.
- **Chatsonic:**[9] Được tạo bởi Writesonic, nó xây dựng dựa trên ChatGPT trong khi cũng thu thập dữ liệu trực tiếp từ Google. Do đó, nó có ít khả năng đưa ra câu trả lời với thông tin sai lệch.
- **Ernie** (còn được gọi là **Wenxin Yiyan** 文心一言):[10]
  Một LLM song ngữ từ Baidu, vẫn đang trong quá trình phát triển, tích hợp kiến thức rộng lớn với các bộ dữ liệu khổng lồ để tạo ra văn bản và hình ảnh.
- **Hugging Chat:**[11] Được tạo bởi Hugging Face, một công ty luôn nhấn mạnh đạo đức và tính minh bạch trong suốt quá trình phát triển, huấn luyện và triển khai của họ. Ngoài ra, tất cả dữ liệu được sử dụng để huấn luyện các mô hình của họ đều là mã nguồn mở.
- **Jasper:**[12] Một bộ công cụ và API có thể, ví dụ, được huấn luyện để viết theo phong cách ưa thích của từmg người dùng. Nó cũng có thể tạo ra hình ảnh.
- **Llama:**[13] Một LLM mã nguồn mở từ Meta yêu cầu ít sức mạnh tính toán và tài nguyên hơn để thử nghiệm các phương pháp mới, kiểm tra công việc của người khác và khám phá các mục đích sử dụng mới.
- **Open Assistant:**[14] Một phương pháp dùng mã nguồn mở được thiết kế để cho phép bất kỳ ai có đủ năng lực chuyên môn tự phát triển LLM của riêng họ. Nó được xây dựng dựa trên dữ liệu huấn luyện được tuyển chọn bởi các tình nguyện viên.
- **Tongyi Qianwen** (通义千问):[15] Một LLM từ Alibaba có thể trả lời các câu lệnh bằng tiếng Anh hoặc tiếng Trung. Nó đang được tích hợp vào bộ công cụ kinh doanh của Alibaba.
- **YouChat:**[16] Một LLM kết hợp khả năng tìm kiếm thời gian thực để cung cấp bối cảnh và hiểu biết sâu sắc bổ sung nhằm tạo ra kết quả chính xác và đáng tin cậy hơn.

Hầu hết các công cụ này đều miễn phí (ở một mức độ nhất định), trong khi một số là mã nguồn mở. Nhiều sản phẩm khác đang được ra mắt dựa trên một trong những LLM này. Ví dụ bao gồm:

- **ChatPDF:**[17] Tóm tắt và trả lời câu hỏi về các tài liệu PDF được tải lên.
- **Elicit: Trợ lý Nghiên cứu AI:**[18] Đặt mục tiêu tự động hóa các khâu trong quy trình làm việc của các nhà nghiên cứu, xác định các bài báo khoa học liên quan và tóm tắt thông tin quan trọng.
- **Perplexity:**[19] Cung cấp một 'kho trung chuyển kiến thức' cho những người tìm kiếm câu trả lời nhanh chóng, chính xác được tinh chỉnh theo nhu cầu của họ.

Tương tự, các công cụ dựa trên LLM đang được tích hợp vào các sản phẩm khác, chẳng hạn như trình duyệt web. Chẳng hạn, các tiện ích mở rộng cho trình duyệt Chrome được xây dựng trên ChatGPT bao gồm:

- **WebChatGPT:**[10] Cung cấp cho ChatGPT quyền truy cập Internet để cho phép các cuộc hội thoại trở nên chính xác và cập nhật hơn.
- **Compose AI:**[21] Tự động hoàn thành câu trong email và các nơi khác.
- **TeamSmart AI:**[22] Cung cấp một 'nhóm trợ lý ảo'.
- **Wiseone:**[23] Đơn giản hóa thông tin trực tuyến.

Ngoài ra, ChatGPT đã được tích hợp vào một số công cụ tìm kiếm,[24] và đang được triển khai trên các danh mục lớn các công cụ năng suất (ví dụ: Microsoft Word và Excel), làm cho nó càng dễ tiếp cận hơn trong các văn phòng và cơ sở giáo dục trên toàn thế giới (Murphy Kelly, 2023).

Cuối cùng, như một bước chuyển tiếp thú vị sang GenAI hình ảnh, mô hình GPT mới nhất từ OpenAI, GPT-4, có khả năng xử lý hình ảnh cũng như văn bản trong các câu lệnh của nó. Theo nghĩa này, nó là đa phương thức. Do đó, một số người cho rằng tên 'mô hình ngôn ngữ lớn' (LLM) đang trở nên kém phù hợp hơn, đó là một lý do tại sao các nhà nghiên cứu tại Đại học Stanford đã đề xuất thuật ngữ 'mô hình nền tảng' (Bommasani và cộng sự, 2021). Lựa chọn thay thế này vẫn chưa được áp dụng rộng rãi.

1.2.2. Các mô hình GenAI hình ảnh hoạt động như thế nào

GenAI hình ảnh và GenAI âm nhạc thường sử dụng một loại ANN khác được gọi là mạng đối kháng tạo sinh (GAN) có thể kết hợp với bộ mã hóa tự động biến phân. Một số mô hình GenAI hình ảnh như Dall·E và Stable Diffusion sử dụng Mô hình Khuếch tán, một loại ANN tạo sinh khác. Lấy GAN làm ví dụ để giải thích cách các mô hình GenAI hình ảnh hoạt động: GAN có 2 phần (2 'đối thủ'), 'bộ tạo sinh' và 'bộ phân biệt'. Trong trường hợp GAN hình ảnh, bộ tạo tạo ra một hình ảnh ngẫu nhiên để đáp ứng một lời nhắc, và bộ phân biệt cố gắng phân biệt giữa hình ảnh được tạo ra này và hình ảnh thật. Sau đó, bộ tạo sinh sử dụng kết quả của bộ phân biệt để điều chỉnh các tham số của nó, để tạo ra một hình ảnh khác. Quá trình này được lặp lại, có thể hàng nghìn lần, với bộ tạo sinh tạo ra những hình ảnh ngày càng thực tế hơn mà bộ phân biệt ngày càng khó phân biệt với hình ảnh thật. Chẳng hạn, một GAN thành công được huấn luyện trên một bộ dữ liệu gồm hàng nghìn bức ảnh phong cảnh có thể tạo ra những hình ảnh phong cảnh mới nhưng gần như giống hệt ảnh thật. Trong khi đó, một GAN được huấn luyện trên một bộ dữ liệu âm nhạc phổ biến (hoặc thậm chí âm nhạc của một nghệ sĩ duy nhất) có thể tạo ra những bản nhạc mới tuân theo cấu trúc và độ phức tạp của bản nhạc gốc.

Tính đến **tháng 7 năm 2023**, các mô hình GenAI hình ảnh hiện có bao gồm những mô hình sau đây, trong đó tất cả đều tạo ra hình ảnh từ câu lệnh bằng văn bản. Hầu hết đều miễn phí, với một số giới hạn nhất định:

- **Craiyon:**[25] Trước đây được biết đến với tên DALL·E mini.
- **DALL·E 2:**[26] Công cụ GenAI hình ảnh của
- **DreamStudio:**[27] Công cụ GenAI hình ảnh của Stable Diffusion.
- **Fotor:**[28] Tích hợp GenAI vào một loạt công cụ chỉnh sửa hình ảnh.
- **Midjourney:**[29] Một công cụ GenAI hình ảnh độc lập.
- **NightCafe:**[20] Giao diện cho Stable Diffusion và DALL·E 2.
- **Photosonic:**[31] Trình tạo nghệ thuật AI của WriteSonic.

Ví dụ về **GenAI video** dễ tiếp cận bao gồm::

- **Elai:**[32] Có thể chuyển đổi các bài thuyết trình, trang web và văn bản thành video.
- **GliaCloud:**[33] Có thể tạo video từ nội dung tin tức, bài đăng trên mạng xã hội, sự kiện thể thao trực tiếp và dữ liệu thống kê.
- **Pictory:**[34] Có thể tự động tạo video ngắn từ nội dung dài.
- **Runway:**[35] Cung cấp một loạt công cụ tạo và chỉnh sửa video (và hình ảnh).

Cuối cùng, đây là một số ví dụ về **GenAI âm nhạc** dễ tiếp cận:

- **Aiva:**[36] Có thể tự động tạo nhạc nền được cá nhân hóa.
- **Boomy,**[37] **Soundraw,**[38] **và Voicemod:**[39] Có thể tạo bài hát từ bất kỳ văn bản nào và không đòi hỏi kiến thức về sáng tác âm nhạc.

1.3 Kỹ thuật soạn câu lệnh để tạo ra các kết quả như mong muốn

Mặc dù việc sử dụng GenAI có thể đơn giản như việc gõ một câu hỏi hoặc câu lệnh khác, thực tế là vẫn không đơn giản để người dùng có được chính xác kết quả mà họ muốn. Chẳng hạn, bức ảnh AI đột phá *Théâtre D'opéra Spatial* đã giành giải thưởng tại Hội chợ Tiểu bang Colorado ở Hoa Kỳ, đã mất nhiều tuần để viết câu lệnh và tinh chỉnh hàng trăm bức ảnh để tạo ra bài dự thi cuối cùng (Roose, 2022). Thách thức tương tự trong việc viết câu lệnh hiệu quả cho GenAI văn bản đã dẫn đến việc xuất hiện ngày càng nhiều công việc kĩ năng soạn câu lệnh trên các trang web tuyển dụng (Popli, 2023). 'Kĩ năng soạn câu lệnh' đề cập đến các quy trình và kỹ thuật soạn thảo đầu vào nhằm tạo ra kết quả GenAI sát hơn với ý đồ của người dùng.

Kỹ thuật soạn câu lệnh thành công nhất khi câu lệnh diễn đạt một chuỗi lý luận mạch lạc tập trung vào một vấn đề cụ thể hoặc một chuỗi các suy nghĩ theo thứ tự logic. Các khuyến nghị cụ thể bao gồm:

- Sử dụng ngôn ngữ **đơn giản**, rõ ràng và chân phương, dễ hiểu, tránh cách diễn đạt phức tạp hoặc mơ hồ.
- Đưa ra **ví dụ** để minh họa câu trả lời mong muốn hoặc định dạng của câu trả lời.
- Mô tả **ngữ cảnh**, điều này rất quan trọng để tạo ra các hoàn thành chuẩn xác và có ý nghĩa.
- **Tinh chỉnh** và lặp lại khi cần thiết, thử nghiệm với các biến thể khác nhau.
- Đảm bảo tính **đạo đức**, tránh các câu lệnh có thể tạo ra nội dung không phù hợp, thiên vị hoặc độc hại.

Cũng cần nhận ra ngay lập tức rằng không thể tin tưởng vào kết quả của GenAI mà không có đánh giá mang tính phản biện. Như OpenAI viết về mô hình GPT cao cấp nhất của họ:[40]

> **'Dù năng lực xử lý tốt hơn song GPT-4 vẫn có những hạn chế tương tự như các mô hình GPT trước đó. Quan trọng nhất, nó vẫn chưa hoàn toàn đáng tin cậy (nó 'bịa' ra các sự kiện và mắc lỗi lập luận). Cần phải hết sức thận trọng khi sử dụng câu trả lời của mô hình ngôn ngữ, đặc biệt là trong các bối cảnh quan trọng, với giao thức chính xác (như sự kiểm tra của con người, bổ sung thêm bối cảnh, hoặc tránh sử dụng trong các tình huống quan trọng) phù hợp với nhu cầu của một mục đích sử dụng cụ thể.'**

Hàm ý đối với giáo dục và nghiên cứu:

Mặc dù GenAI có thể giúp các nhà giáo và nhà nghiên cứu tạo ra văn bản và các kết quả hữu ích khác để hỗ trợ công việc của họ, nhưng đây không phải là một quá trình đơn giản. Có thể cần thử nghiệm nhiều lần một câu lệnh trước khi đạt được kết quả như mong muốn. Một lo ngại là những người học trẻ tuổi, vì họ ít chuyên môn hơn nhà giáo, có thể vô tình và dễ dãi chấp nhận các câu trả lời hời hợt, không chính xác hoặc thậm chí xấu độc của GenAI.

Với chất lượng đầu ra của GenAI, các bài kiểm tra nghiêm ngặt và đánh giá hiệu suất bởi người dùng nên được tiến hành trước khi đưa các công cụ này vào sử dụng ở quy mô lớn hoặc trong các tình huống quan trọng. Các bài tập như vậy nên được thiết kế với một thước đo hiệu suất phù hợp nhất với loại nhiệm vụ mà người dùng yêu cầu từ GenAI. Chẳng hạn, đối với việc giải quyết các bài toán, 'độ chính xác' có thể được sử dụng làm thước đo chính để định lượng tần suất một công cụ GenAI tạo ra câu trả lời đúng; đối với việc trả lời các câu hỏi nhạy cảm, thước đo chính để đo lường hiệu suất có thể là 'tỷ lệ trả lời' (tần suất mà GenAI trực tiếp trả lời một câu hỏi); đối với việc viết mã nguồn, thước đo có thể là 'phần trăm mã nguồn được tạo ra có thể thực thi trực tiếp' (liệu mã nguồn được tạo ra có thể được chạy trực tiếp trong môi trường lập trình và vượt qua các bài kiểm tra đơn vị hay không); và đối với lý luận trực quan, thước đo có thể là 'khớp tuyệt đối' (liệu các đối tượng trực quan được tạo ra có khớp tuyệt đối với các sự thật nền tảng hay không) (Chen và cộng sự, 2023).

Tóm lại, ở bề nổi, GenAI rất dễ sử dụng; tuy nhiên, các kết quả phức tạp hơn cần đầu vào của người có kỹ năng và phải được đánh giá kĩ lưỡng trước khi sử dụng.

1.4 EdGPT mới nổi và hàm ý của nó

Do các mô hình GenAI có thể đóng vai trò là cơ sở hoặc điểm khởi đầu để phát triển các mô hình chuyên biệt hoặc dành riêng cho lĩnh vực cụ thể hơn, một số nhà nghiên cứu đã đề xuất rằng mô hình GPT nên được đổi tên thành 'mô hình nền tảng' (Bommasani và cộng sự, 2021). Trong giáo dục, các nhà phát triển và nhà nghiên cứu đã bắt đầu tinh chỉnh một mô hình nền tảng để phát triển 'EdGPT'.[41] Các mô hình EdGPT được huấn luyện với dữ liệu chuyên biệt để phục vụ các mục tiêu giáo dục. Nói cách khác, EdGPT đặt ra mục tiêu tinh chỉnh mô hình đã được đúc kết từ một lượng lớn dữ liệu huấn luyện chung với một lượng nhỏ hơn dữ liệu chất lượng cao, dành riêng cho lĩnh vực giáo dục.

Điều này có khả năng cho phép EdGPT có khả năng hỗ trợ việc đạt được các mục tiêu chuyển đổi được liệt kê trong Phần 4.3 tốt hơn. Chẳng hạn, các mô hình EdGPT hỗ trợ thiết kế chương trình giáo dục có thể cho phép các thầy cô và người học soạn các tài liệu giáo dục phù hợp như bài giảng, bài kiểm tra và các hoạt động tương tác gắn kết chặt chẽ với một phương pháp sư phạm hiệu quả và các mục tiêu giáo dục cụ thể và các mức độ thách thức phù hợp cho một số người học cá biệt. Tương tự, trong bối cảnh huấn luyện viên kỹ năng ngôn ngữ 1:1, một mô hình nền tảng được tinh chỉnh với các văn bản phù hợp cho một ngôn ngữ cụ thể có thể được sử dụng để tạo ra các câu, đoạn văn hoặc cuộc hội thoại mẫu phục vụ thực hành. Khi người học tương tác với mô hình, nó có thể trả lời với văn bản phù hợp và chính xác về mặt ngữ pháp ở đúng trình độ của họ. Về mặt lý thuyết, kết quả đầu ra của các mô hình EdGPT cũng có thể chứa ít thành kiến chung hoặc nội dung xấu hơn so với mô hình GPT tiêu chuẩn, nhưng vẫn có thể mắc lỗi. Điều quan trọng cần lưu ý là, trừ khi các mô hình GenAI cơ bản và cách tiếp cận thay đổi đáng kể, EdGPT vẫn có thể mắc lỗi và bộc lộ các hạn chế khác. Do đó, vẫn quan trọng là những người dùng EdGPT chủ đạo, đặc biệt là nhà giáo và người học, cần đánh giá khắt khe mọi kết quả đầu ra.

Hiện tại, việc tinh chỉnh các mô hình nền tảng để sử dụng mô hình GPT một cách có mục tiêu hơn trong giáo dục đang ở giai đoạn đầu. Các ví dụ hiện có bao gồm EduChat, một mô hình nền tảng do Đại học Sư phạm Hoa Đông phát triển để cung cấp dịch vụ cho giảng dạy và học tập, và mã nguồn, dữ liệu cũng như tham số của nó được chia sẻ dưới dạng mã nguồn mở.[42] Một ví dụ khác là MathGPT đang được phát triển bởi Tập đoàn Giáo dục TAL - một LLM tập trung vào việc giải quyết vấn đề và giảng dạy về toán học cho người dùng trên toàn thế giới.[43]

Tuy nhiên, trước khi có thể đạt được tiến bộ đáng kể, điều quan trọng là phải nỗ lực tinh chỉnh các mô hình nền tảng không chỉ thông qua việc bổ sung kiến thức môn học và loại bỏ thiên kiến, mà còn thông qua việc bổ sung kiến thức về các phương pháp học tập phù hợp, và cách thể hiện điều này trong thiết kế của các thuật toán và mô hình. Thách thức nằm ở việc xác định mức độ mà các mô hình EdGPT có thể vượt ra ngoài kiến thức môn học để nhắm đến phương pháp sư phạm lấy người học làm trung tâm và tương tác tích cực giữa nhà giáo và người học. Thách thức tiếp theo là xác định mức độ mà dữ liệu của người học và nhà giáo có thể được thu thập và sử dụng một cách có đạo đức để cung cấp thông tin cho một EdGPT. Cuối cùng, cũng cần có nhiều nghiên cứu để đảm bảo rằng EdGPT không làm giảm bớt quyền con người của người học cũng như không tước quyền của nhà giáo.

2. Những tranh cãi xung quanh AI tạo sinh và hàm ý của chúng đối với giáo dục

Sau khi đã thảo luận về GenAI là gì và nó hoạt động như thế nào, phần này sẽ điểm qua các tranh cãi và rủi ro đạo đức do tất cả các hệ thống GenAI gây ra, bên cạnh một số hàm ý đối với giáo dục.

2.1 Làm trầm trọng thêm tình trạng nghèo kỹ thuật số

Như đã nêu ở trên, GenAI dựa vào lượng dữ liệu khổng lồ và sức mạnh tính toán vượt trội ngoài những đổi mới lặp đi lặp lại trong kiến trúc AI và phương pháp huấn luyện, mà phần lớn chỉ có sẵn cho các công ty công nghệ quốc tế lớn nhất và một số nền kinh tế (chủ yếu là Hoa Kỳ, Cộng hòa Nhân dân Trung Hoa, và ở mức độ thấp hơn là châu Âu). Điều này có nghĩa là khả năng tạo ra và kiểm soát GenAI nằm ngoài tầm với của hầu hết các công ty và hầu hết các quốc gia, đặc biệt là những quốc gia ở Nam bán cầu.

Khi việc tiếp cận dữ liệu trở nên ngày càng cần thiết cho sự phát triển kinh tế của các quốc gia và cho các cơ hội số hóa của các cá nhân, những quốc gia và người dân không có quyền truy cập hoặc không đủ khả năng chi trả cho đủ dữ liệu bị đặt vào tình trạng "nghèo dữ liệu" (Marwala, 2023). Tình hình cũng tương tự đối với việc tiếp cận sức mạnh tính toán. Sự lan tỏa nhanh chóng của GenAI ở các quốc gia và khu vực tiên tiến về công nghệ đã tăng tốc theo cấp số nhân việc tạo ra và xử lý dữ liệu, và đồng thời đã làm tăng cường sự tập trung của sự giàu có AI ở Bắc bán cầu. Hậu quả tức thời là các khu vực nghèo dữ liệu đã bị loại trừ thêm và đặt vào rủi ro lâu dài bị thuộc địa hóa bởi các tiêu chuẩn được nhúng trong các mô hình GPT. Các mô hình ChatGPT hiện tại được huấn luyện trên dữ liệu từ người dùng trực tuyến phản ánh các giá trị và chuẩn mực của Bắc bán cầu, khiến chúng không phù hợp cho các thuật toán AI phù hợp với địa phương ở các cộng đồng nghèo dữ liệu ở nhiều phần của Nam bán cầu hoặc ở các cộng đồng thiệt thòi hơn ở Bắc bán cầu.

Hàm ý đối với giáo dục và nghiên cứu:

Các nhà nghiên cứu, nhà giáo và người học nên có góc nhìn phản biện về các định hướng giá trị, tiêu chuẩn văn hóa và phong tục xã hội cố hữu trong các mô hình huấn luyện GenAI. Các nhà hoạch định chính sách nên nhận thức và hành động để giải quyết sự trầm trọng thêm của bất bình đẳng do khoảng cách ngày càng lớn trong việc huấn luyện và kiểm soát các mô hình GenAI.

2.2 Vượt quá khả năng thích ứng của các quy định quốc gia

Các nhà cung cấp GenAI chiếm ưu thế cũng bị chỉ trích vì không cho phép hệ thống của họ chịu sự đánh giá học thuật độc lập nghiêm ngặt (Dwivedi và cộng sự, 2023).[44] công nghệ nền tảng của GenAI của một công ty thường được bảo vệ như tài sản trí tuệ của doanh nghiệp. Trong khi đó, nhiều công ty bắt đầu sử dụng GenAI đang thấy ngày càng khó khăn để duy trì tính bảo mật của hệ thống của họ (Lin, 2023). Hơn nữa, bất chấp những lời kêu gọi quy định từ chính ngành công nghiệp AI,[45] việc dự thảo luật về việc tạo ra và sử dụng tất cả AI, bao gồm cả GenAI, thường bị tụt hậu so với tốc độ phát triển nhanh chóng. Điều này một phần giải thích những thách thức mà các cơ quan quốc gia hoặc địa phương gặp phải trong việc hiểu và quản trị các vấn đề pháp lý và đạo đức.[46]

Mặc dù GenAI có thể tăng cường năng lực của con người trong việc hoàn thành một số nhiệm vụ nhất định, nhưng có rất ít sự kiểm soát dân chủ đối với các công ty đang thúc đẩy GenAI. Điều này đặt ra câu hỏi về các quy định, đặc biệt là liên quan đến việc truy cập và sử dụng dữ liệu trong nước bao gồm dữ liệu về các tổ chức và cá nhân địa phương cũng như dữ liệu được tạo ra trên lãnh thổ của các quốc gia. Cần có luật pháp thích hợp để các cơ quan chính phủ địa phương có thể giành được một số kiểm soát đối với làn sóng GenAI đang dâng cao để đảm bảo việc quản trị nó như một lợi ích cộng đồng.

Hàm ý đối với giáo dục và nghiên cứu:

Các nhà nghiên cứu, nhà giáo và người học nên nhận thức về sự thiếu các quy định thích hợp để bảo vệ quyền sở hữu của các tổ chức và cá nhân trong nước và quyền của người dùng GenAI trong nước, cũng như ứng phó với các vấn đề pháp lý do GenAI gây ra.

2.3 Sử dụng nội dung mà chưa có sự đồng ý

Như đã lưu ý trước đây, các mô hình GenAI được xây dựng từ lượng lớn dữ liệu (ví dụ: văn bản, âm thanh, mã nguồn và hình ảnh) thường được lấy từ Internet và thường không có sự cho phép của chủ sở hữu. Nhiều hệ thống GenAI hình ảnh và một số hệ thống GenAI mã nguồn do đó đã bị cáo buộc vi phạm quyền sở hữu trí tuệ. Tại thời điểm viết bài này, có một số vụ kiện pháp lý quốc tế đang diễn ra liên quan đến vấn đề này.

Hơn nữa, một số người đã chỉ ra rằng mô hình GPT có thể vi phạm các luật như Quy định Bảo vệ Dữ liệu Chung (GDPR) của Liên minh Châu Âu (2016), đặc biệt là quyền được quên của mọi người, vì hiện tại không thể xóa dữ liệu của ai đó (hoặc kết quả của dữ liệu đó) khỏi một mô hình GPT sau khi nó đã được huấn luyện.

Hàm ý đối với giáo dục và nghiên cứu:

• Các nhà nghiên cứu, nhà giáo và người học cần biết về quyền của chủ sở hữu dữ liệu và nên kiểm tra xem các công cụ GenAI họ đang sử dụng có vi phạm bất kỳ quy định hiện có nào không.

• Các nhà nghiên cứu, nhà giáo và người học cũng nên nhận thức rằng các hình ảnh hoặc các mã nguồn được tạo ra bằng GenAI có thể vi phạm quyền sở hữu trí tuệ của người khác, và rằng hình ảnh, âm thanh hoặc các mã nguồn mà họ tạo ra và chia sẻ trên Internet có thể bị khai thác bởi các GenAI khác.

2.4 Các mô hình được sử dụng nhưng không giải thích được

Từ lâu đã được công nhận rằng các mạng nơ-ron nhân tạo (ANN) thường là "hộp đen"; nghĩa là, hoạt động nội bộ của chúng không mở để kiểm tra. Kết quả là, ANN không "minh bạch" hoặc "có thể giải thích được", và không thể xác định được cách thức xác định kết quả kết quả đầu ra của chúng.

Mặc dù cách tiếp cận tổng thể, bao gồm các thuật toán được sử dụng, thường có thể giải thích được, nhưng các mô hình cụ thể và các tham số của chúng, bao gồm trọng số của mô hình, không thể kiểm tra được, đó là lý do tại sao một kết quả đầu ra cụ thể được tạo ra không thể được giải thích. Có hàng tỷ tham số/trọng số trong một mô hình như GPT-4 (xem **Bảng 2**) và chính các trọng số tập thể nắm giữ các mẫu đã học mà mô hình sử dụng để tạo ra kết quả đầu ra của nó. Vì các tham số/trọng số không minh bạch trong ANN (**Bảng 1**), người ta không thể giải thích chính xác cách một kết quả đầu ra cụ thể được tạo ra bởi các mô hình này.

Sự thiếu minh bạch và khả năng giải thích của GenAI ngày càng trở nên có vấn đề khi GenAI trở nên ngày càng phức tạp hơn (xem **Bảng 2**), thường tạo ra kết quả không mong đợi hoặc không mong muốn. Ngoài ra, các mô hình GenAI kế thừa và duy trì các thiên kiến có trong dữ liệu huấn luyện của chúng, mà do tính chất không minh bạch của các mô hình, rất khó phát hiện và giải quyết. Cuối cùng, sự mờ ám này cũng là nguyên nhân chính gây ra các vấn đề về tin cậy xung quanh GenAI (Nazaretsky và cộng sự, 2022*a*). Nếu người dùng không hiểu cách một hệ thống GenAI đi đến một đầu ra cụ thể, họ ít có khả năng sẵn sàng áp dụng hoặc sử dụng nó (Nazaretsky và cộng sự, 2022*b*).

Hàm ý đối với giáo dục và nghiên cứu:

Các nhà nghiên cứu, nhà giáo và người học nên nhận thức rằng các hệ thống GenAI hoạt động như những hộp đen và do đó rất khó, nếu không muốn nói là không thể, biết tại sao nội dung cụ thể đã được tạo ra. Sự thiếu giải thích về cách tạo ra đầu ra có xu hướng khóa người dùng trong logic được định nghĩa bởi các tham số được thiết kế trong các hệ thống GenAI. Các tham số này có thể phản ánh các giá trị và chuẩn mực văn hóa hoặc thương mại cụ thể mà ngầm làm thiên lệch nội dung được tạo ra.

2.5 Nội dung do AI tạo ra làm ô nhiễm Internet

Vì dữ liệu huấn luyện mô hình GPT thường được lấy từ Internet, mà rất hay bao gồm ngôn ngữ phân biệt đối xử và khó chấp nhận được khác, nên các nhà phát triển đã phải triển khai cái họ gọi là 'rào cản' để ngăn chặn đầu ra của mô hình GPT trở nên phản cảm và/hoặc phi đạo đức. Tuy nhiên, do thiếu các quy định nghiêm ngặt và cơ chế giám sát hiệu quả, nên các tài liệu có sự thiên lệch được tạo ra bởi GenAI ngày càng lan rộng khắp Internet, làm ô nhiễm một trong những nguồn nội dung hoặc kiến thức chính cho hầu hết người học trên toàn thế giới. Điều này đặc biệt quan trọng vì tài liệu do GenAI tạo ra trông có vẻ khá chính xác và thuyết phục, trong khi thường xuyên chứa đựng những sai lầm và ý tưởng thiên vị. Điều này đặt ra một rủi ro cao đối với những người học trẻ tuổi vốn thiếu kiến thức nền tảng vững chắc về chủ đề đang được đề cập. Nó cũng đặt ra một rủi ro khó tránh khỏi cho các mô hình GPT trong tương lai dự kiến được huấn luyện trên văn bản được thu thập từ Internet mà chính các mô hình GPT đã tạo ra mà có đủ những thiên kiến và thiếu sót của chúng.

Hàm ý đối với giáo dục và nghiên cứu:

- Các nhà nghiên cứu, nhà giáo và người học cần nhận thức rằng các hệ thống GenAI có khả năng tạo ra các tài liệu độc hại và phi đạo đức.
- Họ cũng cần biết về các vấn đề lâu dài có thể phát sinh về độ tin cậy của kiến thức khi các mô hình GPT trong tương lai được phát triển dựa trên văn bản mà chính các mô hình GPT trước đó đã tạo ra.

2.6 Thiếu hiểu biết về thế giới thực

Các mô hình GPT văn bản đôi khi bị gọi một cách miệt thị là 'những con vẹt ngẫu hứng' bởi vì, như đã đề cập trước đó, mặc dù chúng có thể tạo ra các văn bản có vẻ thuyết phục, các văn bản đó thường có nhiều lỗi và có thể bao gồm các nội dung xấu độc (Bender et al., 2021). Tất cả điều này xảy ra bởi vì các mô hình GPT chỉ lặp lại các mẫu ngôn ngữ được tìm thấy trong dữ liệu huấn luyện của chúng (thường là văn bản lấy từ Internet), bắt đầu với các mẫu ngẫu nhiên (hay 'stochastic'), và không hiểu ý nghĩa của chúng - giống như một con vẹt có thể bắt chước âm thanh mà không thực sự hiểu những gì nó đang nói.

Việc các mô hình GenAI 'có vẻ' hiểu văn bản mà chúng sử dụng và tạo ra, và 'sự thật' là chúng không hiểu ngôn ngữ và thế giới thực có thể dẫn đến việc nhà giáo và người học tin tưởng quá mức vào đầu ra. Điều này đặt ra những rủi ro nghiêm trọng cho nền giáo dục tương lai. Trên thực tế, GenAI không có các quan sát về thế giới thực hoặc hiểu biết quan trọng về phương pháp khoa học, nó cũng không phù hợp với các giá trị con người hoặc xã hội. Vì những lý do này, nó không thể tạo ra nội dung thực sự mới mẻ về thế giới thực, các sự vật và mối quan hệ của chúng, con người và quan hệ xã hội, quan hệ người-sự vật, hoặc quan hệ người-công nghệ. Việc nội dung có vẻ mới được tạo ra bởi các mô hình GenAI có thể được công nhận là kiến thức khoa học hay không vẫn còn gây tranh cãi.

Như đã đề cập, các mô hình GPT thường xuyên tạo ra văn bản kém chính xác hoặc không đáng tin cậy. Và đa số cũng biết rằng các mô hình GPT bịa ra một số thứ không thực sự tồn tại. Một số người gọi đây là 'ảo giác' (hallucination), mặc dù những người khác chỉ trích việc sử dụng thuật ngữ có tính nhân cách hóa này, cho rằng nó sẽ gây hiểu lầm. Điều này được thừa nhận bởi các công ty làm ra GenAI. Chẳng hạn, phần cuối của giao diện công khai của ChatGPT có ghi: 'ChatGPT có thể tạo ra thông tin không chính xác về con người, địa điểm, hoặc các sự kiện'.[2]

Một số người ủng hộ cũng đã nhận định rằng GenAI đại diện cho một bước tiến đáng kể trong hành trình hướng tới trí tuệ nhân tạo tổng quát (AGI), một thuật ngữ chỉ một lớp AI thông minh hơn con người. Tuy nhiên, điều này đã bị chỉ trích từ lâu, với lập luận rằng AI sẽ không bao giờ tiến tới AGI ít nhất cho đến khi nó bằng cách nào đó kết hợp, trong một mối quan hệ cộng sinh, cả AI dựa trên kiến thức (còn được gọi là AI biểu tượng hoặc dựa trên quy tắc) và AI dựa trên dữ liệu (còn được gọi là học máy) (Marcus, 2022). Các tuyên bố về AGI hoặc ý thức cũng làm chúng ta bị phân tâm khỏi việc suy xét kĩ lưỡng hơn các tác hại hiện tại đang được gây ra bởi AI, chẳng hạn như phân biệt đối xử ngầm đối với các nhóm vốn đã bị phân biệt đối xử (Metz, 2021).

Hàm ý đối với giáo dục và nghiên cứu:

• Kết quả đầu ra của một GenAI văn bản có thể trông giống hệt như của con người, như thể nó hiểu văn bản mà nó tạo ra. Tuy nhiên, GenAI không hiểu bất cứ điều gì. Thay vào đó, những công cụ này chắp nối các từ lại với nhau theo những cách phổ biến trên Internet. Văn bản được tạo ra do đó cũng có thể không chính xác.

• Các nhà nghiên cứu, nhà giáo và người học cần nhận thức rằng một GPT không hiểu văn bản mà nó tạo ra; rằng nó có thể, và thường xuyên, tạo ra các phát biểu không chính xác; và do đó họ cần có cách tiếp cận phản biện đối với mọi thứ mà nó tạo ra.

2.7 Làm giảm mức độ đa dạng quan điểm và suy yếu những tiếng nói thiểu số

ChatGPT và các công cụ tương tự thường chỉ đưa ra những câu trả lời tiêu chuẩn, dựa trên giả định về các giá trị của chủ sở hữu/người tạo ra dữ liệu được sử dụng để huấn luyện các mô hình. Thế nên nếu một chuỗi từ xuất hiện thường xuyên trong dữ liệu huấn luyện - như trường hợp với các chủ đề phổ biến và không gây tranh cãi và các niềm tin phổ biến - thì nó có khả năng được mô hình GPT lặp lại trong đầu ra của nó.

Điều này có nguy cơ giới hạn và làm suy yếu sự phát triển của các ý kiến đa dạng và các cách biểu đạt đa dạng của ý tưởng. Các nhóm người nghèo dữ liệu, bao gồm các cộng đồng yếu thế ở Bắc bán cầu, có sự hiện diện số tối thiểu hoặc hạn chế trên mạng. Do đó, tiếng nói của họ không được lắng nghe và mối quan tâm của họ không được thể hiện trong dữ liệu được sử dụng để huấn luyện mô hình GPT, nên hiếm khi xuất hiện trong các đầu ra. Vì những lý do này, với phương pháp huấn luyện trước dựa trên dữ liệu từ các trang web Internet và cuộc trò chuyện trên mạng xã hội, các mô hình GPT có thể càng gạt bỏ những người vốn đã thiệt thòi.

Hàm ý đối với giáo dục và nghiên cứu:

• Mặc dù những người phát triển và nhà cung cấp các mô hình GenAI có trách nhiệm chính trong việc liên tục giải quyết những thiên kiến trong bộ dữ liệu và đầu ra của các mô hình này, các nhà nghiên cứu, nhà giáo và người học phía người dùng cần biết rằng đầu ra của GenAI văn bản chỉ đại diện cho các quan điểm phổ biến nhất hoặc nổi trội về thế giới tại thời điểm dữ liệu huấn luyện của nó được tạo ra và một số trong đó là có vấn đề hoặc thiên vị (ví dụ: thể hiện giới theo khuôn mẫu).

• Người học, nhà giáo và nhà nghiên cứu không bao giờ nên chấp nhận thông tin do GenAI cung cấp một cách dễ dãi, phải luôn đánh giá nó một cách khắt khe.

• Các nhà nghiên cứu, nhà giáo và người học cũng phải nhận thức được việc các quan điểm thiểu số có thể bị bỏ qua, bởi vì quan điểm số chắc chắn ít xuất hiện hơn hơn trong dữ liệu huấn luyện.

2.8 Tạo ra deepfake thật hơn

Ngoài những tranh cãi phổ biến đối với tất cả GenAI, GAN GenAI có thể được sử dụng để chỉnh sửa hoặc thay đổi hình ảnh hoặc video hiện có để tạo ra những hình ảnh giả mạo khó phân biệt với hình ảnh thật. GenAI đang khiến việc tạo ra những "deepfake" này và tin giả ngày càng dễ dàng hơn. Nói cách khác, GenAI đang khiến một số tác nhân nhất định dễ dàng hơn trong việc triển khai các hành vi phi đạo đức, vô đạo đức và phạm pháp, chẳng hạn như lan truyền thông tin sai lệch, đưa ra phát ngôn thù địch và đưa khuôn mặt của mọi người vào các đoạn phim hoàn toàn giả mạo và đôi khi chứa các nội dung xấu, trong khi không có sự hiểu biết hoặc đồng ý của họ.

Hàm ý đối với giáo dục và nghiên cứu:

Mặc dù các nhà cung cấp GenAI có nghĩa vụ bảo vệ bản quyền và quyền chân dung của người dùng, các nhà nghiên cứu, nhà giáo và người học cũng cần nhận thức rằng bất kỳ hình ảnh nào họ chia sẻ trên Internet đều có thể được đưa vào dữ liệu huấn luyện GenAI và có thể bị chỉnh sửa và sử dụng theo những cách phi đạo đức.

3. Quy định việc sử dụng AI tạo sinh trong giáo dục

Để giải quyết những tranh cãi xung quanh AI tạo sinh và khai thác những lợi ích tiềm năng của GenAI trong giáo dục, trước tiên cần phải có quy định. Việc quản trị GenAI cho mục đích giáo dục đòi hỏi một số bước và biện pháp chính sách dựa trên cách tiếp cận lấy con người làm trung tâm để đảm bảo việc sử dụng có đạo đức, an toàn, công bằng và có ý nghĩa.

3.1 Cách tiếp cận AI lấy con người làm trung tâm

Khuyến nghị về Đạo đức AI năm 2021 của UNESCO cung cấp khung quy phạm cần thiết để bắt đầu giải quyết nhiều tranh cãi xung quanh AI tạo sinh, bao gồm cả những tranh cãi liên quan đến giáo dục và nghiên cứu. Nó dựa trên cách tiếp cận lấy con người làm trung tâm đối với AI, cho rằng việc sử dụng AI nên phục vụ cho sự phát triển năng lực của con người hướng tới một tương lai hòa nhập, công bằng và bền vững. Cách tiếp cận như vậy phải được định hướng bởi các nguyên tắc cơ bản của quyền con người, và bảo vệ phẩm giá con người và sự đa dạng văn hóa vốn là nền tảng của kho tàng kiến thức chung của nhân loại. Về mặt quản trị, cách tiếp cận lấy con người làm trung tâm đòi hỏi phải có quy định thích hợp có thể đảm bảo quyền tự chủ của con người, tính minh bạch và trách nhiệm giải trình trước người dân.

Đồng thuận Bắc Kinh năm 2019 về Trí tuệ Nhân tạo (AI) và Giáo dục tiếp tục làm rõ ý nghĩa của cách tiếp cận lấy con người làm trung tâm đối với việc sử dụng AI trong bối cảnh giáo dục. Đồng thuận khẳng định rằng việc sử dụng công nghệ AI trong giáo dục phải tăng cường năng lực của con người vì sự phát triển bền vững và hợp tác hiệu quả giữa người và máy trong cuộc sống, học tập và làm việc. Nó cũng kêu gọi thêm các hành động để đảm bảo khả năng tiếp cận AI công bằng nhằm hỗ trợ những người yếu thế và giải quyết bất bình đẳng, đồng thời thúc đẩy sự đa dạng về ngôn ngữ và văn hóa. Đồng thuận đề xuất áp dụng cách tiếp cận toàn chính phủ, liên ngành và đa bên liên quan trong việc lập kế hoạch chính sách về AI trong giáo dục.

Tài liệu AI và Giáo dục: Hướng dẫn cho các nhà hoạch định chính sách (UNESCO, 2022*b*) tiếp tục làm rõ nội hàm của cách tiếp cận lấy con người làm trung tâm khi xem xét lợi ích và rủi ro của AI trong giáo dục và vai trò của giáo dục như một phương tiện để phát triển năng lực sử dụng AI. Nó đề xuất các khuyến nghị cụ thể để xây dựng chính sách nhằm định hướng việc sử dụng AI nhằm (i) thúc đẩy cách tiếp cận hòa nhập các chương trình học tập, đặc biệt là cho các nhóm dễ bị tổn thương như người học khuyết tật; (ii) hỗ trợ các lựa chọn học tập cá nhân hóa và mở; (iii) cải thiện các quy định dựa trên dữ liệu và quản lý để mở rộng tiếp cận và nâng cao chất lượng học tập; (iv) giám sát quá trình học tập và cảnh báo nhà giáo về nguy cơ thất bại; và (v) phát triển hiểu biết và kỹ năng để sử dụng AI một cách có đạo đức và có ý nghĩa.

3.2 Các bước để quản lý GenAI trong giáo dục

Trước khi ChatGPT được công bố, các chính phủ đã xây dựng hoặc điều chỉnh các khuôn khổ để quy định việc thu thập và sử dụng dữ liệu cũng như việc áp dụng các hệ thống AI trên các lĩnh vực bao gồm cả giáo dục, điều này tạo nên một bối cảnh pháp lý và chính sách cho việc quy định các ứng dụng AI mới nổi. Sau khi nhiều mô hình GenAI cạnh tranh kể ra đời kể từ tháng 11 năm 2022, chính phủ các nước đã có những phản ứng chính sách khác nhau - từ việc cấm GenAI đến đánh giá nhu cầu điều chỉnh các khuôn khổ hiện có, cho tới việc khẩn trương xây dựng các quy định mới.

Các chiến lược của các chính phủ để quy định và thúc đẩy sử dụng sáng tạo GenAI đã được sơ đồ hóa và đánh giá vào tháng 4 năm 2023 (UNESCO, 2023*b*).[47] Khảo sát này đề xuất một mô hình 7 bước mà các cơ quan chính phủ có thể triển khai để quản trị AI tạo sinh và giành lại quyền kiểm soát của người dân đối với công nghệ này nhằm tận dụng tiềm năng của nó trên các lĩnh vực, bao gồm cả giáo dục.

Bước 1: Phê chuẩn các quy định bảo vệ dữ liệu chung quốc tế hoặc khu vực hoặc phát triển các quy định quốc gia

Việc huấn luyện các mô hình GenAI từ lâu đã bao gồm việc thu thập và xử lý dữ liệu trực tuyến từ công dân trên nhiều quốc gia. Việc các mô hình GenAI sử dụng

dữ liệu và nội dung mà không có sự đồng ý đang tiếp tục đặt ra thách thức đối với việc bảo vệ dữ liệu.

Các quy định bảo vệ dữ liệu chung, với GDPR của EU được ban hành vào năm 2018 là một trong những ví dụ tiên phong, cung cấp khuôn khổ pháp lý cần thiết để quy định việc thu thập và xử lý dữ liệu cá nhân của các nhà cung cấp GenAI. Theo cổng thông tin Luật Bảo vệ Dữ liệu và Quyền riêng tư Trên toàn thế giới của Hội nghị Liên Hợp Quốc về Thương mại và Phát triển (UNCTAD), 137 trong số 194 quốc gia đã ban hành luật để bảo vệ dữ liệu và quyền riêng tư.[48]

Tuy nhiên, mức độ triển khai các khuôn khổ này ở các quốc gia đó vẫn chưa rõ ràng. Do đó, việc đảm bảo rằng chúng được triển khai đúng cách, bao gồm cả việc giám sát thường xuyên hoạt động của các hệ thống GenAI, ngày càng trở nên quan trọng hơn. Điều cũng cấp thiết không kém là việc các quốc gia chưa có luật bảo vệ dữ liệu chung sớm phát triển chúng.

Bước 2: Ban hành/sửa đổi và tài trợ cho các chiến lược toàn chính phủ về AI

Việc quản trị AI tạo sinh phải là một phần cốt lõi của các chiến lược AI quốc gia tổng quát để có thể đảm bảo việc sử dụng AI an toàn và công bằng trên các lĩnh vực phát triển, bao gồm cả giáo dục. Việc xây dựng, thông qua, tài trợ và triển khai các chiến lược AI quốc gia đòi hỏi cách tiếp cận toàn chính phủ. Chỉ có cách tiếp cận như vậy mới có thể đảm bảo sự phối hợp các hành động liên ngành cần thiết cho các phản ứng tích hợp đối với những thách thức mới nổi.

Đến đầu năm 2023, khoảng 67 quốc gia[49] đã phát triển hoặc lên kế hoạch cho các chiến lược quốc gia về AI, trong đó 61 có chiến lược AI độc lập, và 7 quốc gia có các chương về AI được tích hợp trong các chiến lược CNTT hoặc chiến lược số quốc gia rộng hơn. Do tính mới mẻ của nó, có thể hiểu vì sao không có chiến lược quốc gia nào trong số này đề cập đến AI tạo sinh như một vấn đề cụ thể tại thời điểm xuất bản Hướng dẫn này.

Điều quan trọng là các quốc gia cần sửa đổi các chiến lược AI quốc gia hiện có, hoặc phát triển chúng, đảm bảo có các quy định để điều chỉnh việc sử dụng AI có đạo đức trên các lĩnh vực bao gồm cả giáo dục.

Bước 3: Củng cố và triển khai các quy định cụ thể về đạo đức của AI

Để giải quyết các khía cạnh đạo đức do việc sử dụng AI đặt ra, cần có các quy định cụ thể.

Đánh giá của UNESCO năm 2023 về các chiến lược AI quốc gia hiện có cho thấy việc xác định các vấn đề đạo đức như vậy và việc xây dựng các nguyên tắc hướng dẫn chỉ phổ biến ở khoảng 40 chiến lược quốc gia về AI.[50] Và ngay cả ở đây, các nguyên tắc đạo đức sẽ cần được chuyển thành luật hoặc quy định có thể thực thi được. Tuy nhiên điều này lại không phổ biến và trên thực tế, chỉ có khoảng 20 quốc gia đã đặt ra quy định rõ ràng về đạo đức của AI bao gồm cả những quy định liên quan đến giáo dục, hoặc là một phần của chiến lược AI quốc gia hoặc theo cách khác. Thú vị là, trong khi giáo dục được nhấn mạnh như một lĩnh vực chính sách trên khoảng 45 chiến lược AI quốc gia,[51] giáo dục được nhắc đến ở góc độ kỹ năng AI và phát triển năng lực cần thiết để hỗ trợ khả năng cạnh tranh quốc gia nhiều hơn là về khía cạnh đạo đức.

Các quốc gia chưa có quy định về đạo đức của AI phải khẩn trương xây dựng và triển khai chúng.

Bước 4: Điều chỉnh hoặc thực thi luật bản quyền hiện có để quản lý nội dung do AI tạo ra

Việc sử dụng GenAI ngày càng phổ biến đã đặt ra những thách thức mới về bản quyền, cả về nội dung hoặc tác phẩm có bản quyền mà các mô hình sử dụng để huấn luyện, cũng như tình trạng của các sản phẩm kiến thức mà máy móc tạo ra.

Hiện tại chỉ có Trung Quốc, các nước EU và Hoa Kỳ đã điều chỉnh luật bản quyền để tính đến các tác động của AI tạo sinh. Chẳng hạn, Cục Bản quyền Hoa Kỳ đã quyết định rằng kết quả đầu ra của các hệ thống GenAI như ChatGPT không được bảo vệ theo luật bản quyền Hoa Kỳ, với lập luận rằng "bản quyền chỉ có thể bảo vệ tài liệu là sản phẩm của sự sáng tạo của con người" (Cục Bản quyền Hoa Kỳ, 2023). Trong khi đó ở EU, dự thảo Đạo luật AI của EU yêu cầu các nhà phát triển công cụ AI công khai các tài liệu có bản quyền mà họ đã sử dụng trong việc xây dựng hệ thống của họ (Ủy ban Châu Âu, 2021). Trung Quốc thông qua quy định về GenAI được ban hành vào tháng 7 năm 2023, đã yêu cầu gắn nhãn các kết quả đầu ra của GenAI là nội dung do AI tạo ra, và chỉ công nhận chúng là kết quả đầu ra của quá trình tổng hợp số.

Việc quy định sử dụng các tài liệu có bản quyền trong việc huấn luyện các mô hình GenAI và xác định trạng thái bản quyền của các kết quả đầu ra GenAI đang nổi lên như những trách nhiệm mới của luật bản quyền. Việc các luật hiện hành cần được điều chỉnh để tính đến điều này là hết sức cấp thiết.

Bước 5: Xây dựng các khuôn khổ quy định về AI tạo sinh

Tốc độ phát triển nhanh chóng của các công nghệ AI đang buộc các cơ quan quản lý quốc gia/địa phương phải đẩy nhanh việc đổi mới các quy định của họ. Tính đến tháng 7 năm 2023, chỉ có một quốc gia, Trung Quốc, đã ban hành các quy định chính thức và cụ thể về GenAI. Các Quy định Tạm thời về Quản lý Dịch vụ AI tạo sinh được ban hành vào ngày 13 tháng 7 năm 2023 (Cục Quản lý Không gian mạng Trung Quốc, 2023a) yêu cầu các nhà cung cấp hệ thống GenAI gắn nhãn nội dung, hình ảnh và video do AI tạo ra một cách đúng đắn và hợp pháp theo Quy định về Tổng hợp Sâu trong Khuôn khổ Dịch vụ Thông tin Trực tuyến hiện có của họ. Sẽ cần phát triển thêm các khung pháp quy GenAI quốc gia cụ thể như vậy dựa trên đánh giá về những lỗ hổng trong các quy định và luật địa phương hiện có.

Bước 6: Phát triển năng lực sử dụng GenAI đúng cách trong giáo dục và nghiên cứu

Các trường học và các cơ sở giáo dục khác cần có khả năng thấu hiểu các lợi ích và rủi ro tiềm ẩn của AI, bao gồm cả GenAI, đối với giáo dục. Chỉ dựa trên sự hiểu biết như vậy họ mới có thể kiểm định việc áp dụng các công cụ AI. Hơn nữa, nhà giáo và nhà nghiên cứu cần được hỗ trợ để tăng cường năng lực sử dụng GenAI đúng cách, bao gồm cả thông qua huấn luyện và hướng dẫn liên tục. Một số quốc gia đã triển khai các chương trình phát triển năng lực như vậy, bao gồm Singapore, nước đã cung cấp một nền tảng chuyên dụng cho việc phát triển năng lực AI của các cơ sở giáo dục thông qua Cụm Điện toán Đám mây AI Chính phủ, bao gồm một kho lưu trữ chuyên dụng các mô hình GPT (Ocampo, 2023).

Bước 7: Phản tư về các tác động lâu dài của GenAI đối với giáo dục và nghiên cứu

Tác động của các phiên bản GenAI hiện tại chỉ mới bắt đầu xuất hiện, và ảnh hưởng của chúng đối với giáo dục vẫn chưa được khám phá và hiểu đầy đủ. Trong khi đó, các phiên bản mạnh mẽ hơn của GenAI và các loại AI khác tiếp tục được phát triển và triển khai. Tuy nhiên, những câu hỏi quan trọng vẫn còn đó, xoay quanh các tác động của GenAI đối với việc tạo ra, truyền tải và kiểm tra kiến thức - đối với giảng dạy và học tập, đối với thiết kế chương trình giảng dạy và đánh giá, và đối với nghiên cứu và bản quyền. Hầu hết các quốc gia đang ở giai đoạn đầu của việc áp dụng GenAI trong giáo dục, ngay cả khi những tác động lâu dài vẫn chưa được hiểu rõ. Để đảm bảo việc sử dụng AI lấy con người làm trung tâm, cần khẩn trương tiến hành các cuộc tranh luận công khai và đối thoại chính sách về những tác động lâu dài. Cuộc tranh luận toàn diện với sự tham gia của chính phủ, khu vực tư nhân và các đối tác khác, cần cung cấp cái nhìn sâu sắc và đầu vào cho việc cập nhật liên tục các quy định và chính sách.

3.3 Quy định về GenAI: Các thành tố chính

Tất cả các quốc gia cần quản trị GenAI một cách thích hợp để đảm bảo nó mang lại lợi ích cho sự phát triển trong giáo dục và trong các bối cảnh khác. Phần này đề xuất các hành động xung quanh các yếu tố chính có thể được triển khai bởi: (1) các cơ quan quản lý, (2) các nhà cung cấp các công cụ được hỗ trợ bởi AI, (3) người dùng là tổ chức, và (4) người dùng cá nhân. Mặc dù nhiều yếu tố trong khuôn khổ mang tính xuyên quốc gia, tất cả cũng nên được đặt vào bối cảnh riêng của từng nơi, tức các hệ thống giáo dục cụ thể của từng quốc gia và các khuôn khổ quy định chung đã có sẵn.

3.3.1. Các cơ quan quản lý

Cần có cách tiếp cận toàn chính phủ để phối hợp việc thiết kế, điều chỉnh và triển khai các quy định về GenAI. Bảy yếu tố chính và hành động sau đây được khuyến nghị:

- **Phối hợp liên ngành**: Thành lập một cơ quan quốc gia để dẫn dắt cách tiếp cận toàn chính phủ đối với GenAI và phối hợp hợp tác giữa các lĩnh vực.
- **Thích ứng với luật pháp**: Đảm bảo khung quản lý phù hợp với môi trường lập pháp và quy định liên quan của mỗi quốc gia - với, ví dụ, luật bảo vệ dữ liệu chung, quy định về an ninh Internet, luật về an ninh dữ liệu được tạo ra từ hoặc được sử dụng để phục vụ công dân, và các luật và thông lệ thông thường khác có liên quan. Đánh giá sự phù hợp của các quy định hiện có và bất kỳ sự điều chỉnh cần thiết nào để đáp ứng các vấn đề mới do GenAI đặt ra.

- **Cân bằng giữa kiểm soát GenAI và thúc đẩy đổi mới AI:** Thúc đẩy hợp tác liên ngành giữa các công ty, tổ chức, và các cơ sở giáo dục và nghiên cứu, cũng như các cơ quan công quyền liên quan để cùng nhau phát triển các mô hình đáng tin cậy; khuyến khích xây dựng hệ sinh thái mã nguồn mở để thúc đẩy việc chia sẻ tài nguyên siêu máy tính và bộ dữ liệu huấn luyện trước chất lượng cao; và thúc đẩy ứng dụng thực tế của GenAI trên các lĩnh vực và việc tạo ra nội dung chất lượng cao vì lợi ích cộng đồng.

- **Đánh giá và phân loại các rủi ro tiềm ẩn của AI:** Thiết lập các nguyên tắc và quy trình để đánh giá và phân loại hiệu quả, an toàn và bảo mật của các dịch vụ GenAI, trước khi chúng được triển khai và trong suốt vòng đời của hệ thống. Xem xét các cơ chế phân loại dựa trên các mức độ rủi ro mà GenAI có thể đặt ra cho người dân. Phân loại chúng thành các mức độ: quy định nghiêm ngặt (tức là cấm các ứng dụng hoặc hệ thống được hỗ trợ bởi AI có rủi ro không thể chấp nhận được), các quy định đặc biệt cho các ứng dụng có rủi ro cao, và các quy định chung về các ứng dụng không được coi là rủi ro cao. Xem dự thảo Đạo luật AI của EU để biết thêm ví dụ về cách tiếp cận này.

- **Bảo vệ quyền riêng tư dữ liệu**: Tính đến thực tiễn là việc sử dụng GenAI hầu như luôn đòi hỏi việc người dùng chia sẻ dữ liệu của họ với nhà cung cấp GenAI. Yêu cầu bắt buộc việc soạn thảo và thực thi các luật bảo vệ thông tin cá nhân của người dùng cũng như xác định và chống lại việc lưu trữ, lập hồ sơ và chia sẻ dữ liệu bất hợp pháp.

- **Xác định và thực thi giới hạn tuổi sử dụng GenAI**: Hầu hết các ứng dụng GenAI đều được thiết kế cho người dùng ở độ tuổi trưởng thành. Những ứng dụng này thường tiềm ẩn rủi ro đáng kể đối với trẻ em, bao gồm việc tiếp xúc với nội dung không phù hợp cũng như khả năng bị thao túng. Trước những rủi ro này và do sự bất định đáng kể vẫn tiếp tục xoay quanh nhiều thế hệ ứng dụng GenAI, các hạn chế về tuổi được khuyến nghị mạnh mẽ đối với các công nghệ AI nền tảng để bảo vệ quyền và phúc lợi của trẻ em.

  Hiện tại, điều khoản sử dụng của ChatGPT yêu cầu người dùng phải ít nhất 13 tuổi, và người dùng dưới 18 tuổi phải có sự cho phép của cha mẹ hoặc người giám hộ hợp pháp để sử dụng dịch vụ. Những hạn chế hoặc ngưỡng tuổi này được lấy từ Đạo luật Bảo vệ Quyền riêng tư Trực tuyến của Trẻ em của Hoa Kỳ (Ủy ban Thương mại Liên bang, 1998). Được thông qua vào năm 1998 trước khi sử dụng mạng xã hội rộng rãi và lâu trước khi có các ứng dụng GenAI dễ sử dụng và mạnh như ChatGPT, luật Hoa Kỳ quy định rằng các tổ chức hoặc nhà cung cấp mạng xã hội cá nhân không được phép cung cấp dịch vụ cho trẻ em dưới 13 tuổi mà không có sự cho phép của cha mẹ. Nhiều ý kiến cho rằng ngưỡng này là quá thấp và ủng hộ việc lập pháp để nâng độ tuổi lên 16. GDPR của Liên minh Châu Âu (2016) quy định rằng người dùng phải ít nhất 16 tuổi để sử dụng các dịch vụ mạng xã hội mà không cần sự đồng ý của cha mẹ.

  Sự xuất hiện của các chatbot GenAI khác nhau đòi hỏi các quốc gia phải cẩn thận suy xét - và thảo luận công khai ngưỡng tuổi thích hợp cho các cuộc hội thoại độc lập với các nền tảng GenAI. Ngưỡng tối thiểu nên là 13 tuổi. Các quốc gia cũng sẽ cần quyết định xem việc tự báo cáo tuổi có còn là phương tiện xác minh tuổi phù hợp hay không. Các quốc gia sẽ cần bắt buộc trách nhiệm giải trình của nhà cung cấp GenAI về việc xác minh tuổi và trách nhiệm của cha mẹ hoặc người giám hộ trong việc giám sát các cuộc hội thoại độc lập của trẻ chưa đủ tuổi.

- **Quyền sở hữu dữ liệu quốc gia và nguy cơ nghèo dữ liệu**: Triển khai các biện pháp lập pháp để bảo vệ quyền sở hữu dữ liệu quốc gia và quy định các nhà cung cấp GenAI hoạt động trong biên giới của mình. Đối với các bộ dữ liệu được tạo ra bởi các công dân đang được sử dụng cho mục đích thương mại, thiết lập các quy định để thúc đẩy hợp tác cùng có lợi để danh mục dữ liệu này không bị rút cạn khỏi quốc gia để bị khai thác độc quyền bởi các công ty công nghệ lớn.

3.3.2. Các nhà cung cấp công cụ GenAI

Các nhà cung cấp GenAI bao gồm các tổ chức và cá nhân chịu trách nhiệm phát triển và cung cấp các công cụ GenAI, và/hoặc đang sử dụng công nghệ GenAI để cung cấp dịch vụ bao gồm cả thông qua các giao diện lập trình ứng dụng (API) có thể được lập trình. Hầu hết các nhà cung cấp GenAI có ảnh hưởng là các công ty

có ngân quỹ dồi dào. Cần làm rõ với các nhà cung cấp GenAI rằng họ chịu trách nhiệm về đạo đức ngay từ khâu thiết kế, bao gồm cả việc triển khai các nguyên tắc đạo đức được nêu trong các quy định. Mười loại trách nhiệm giải trình sau đây nên được bao gồm:

- **Trách nhiệm giải trình của con người**: Các nhà cung cấp GenAI nên chịu trách nhiệm đảm bảo tuân thủ các giá trị cốt lõi và mục đích hợp pháp, tôn trọng quyền sở hữu trí tuệ, và duy trì thực hành đạo đức, đồng thời ngăn chặn sự lan truyền của thông tin sai lệch và phát ngôn thù địch.

- **Dữ liệu và mô hình đáng tin cậy**: Các nhà cung cấp GenAI nên được yêu cầu chứng minh tính đáng tin cậy và đạo đức của các nguồn dữ liệu và phương pháp được sử dụng bởi các mô hình và kết quả đầu ra của chúng. Chúng phải được yêu cầu bắt buộc áp dụng dữ liệu và mô hình nền tảng có nguồn gốc hợp pháp có thể chứng minh được, và tuân thủ các luật sở hữu trí tuệ liên quan (ví dụ: nếu dữ liệu được bảo vệ bởi quyền sở hữu trí tuệ). Ngoài ra, khi các mô hình cần sử dụng thông tin cá nhân, việc thu thập thông tin đó chỉ nên xảy ra với sự đồng thuận có hiểu biết và rõ ràng của chủ sở hữu.

- **Tạo nội dung không phân biệt đối xử**: Các nhà cung cấp GenAI phải cấm việc thiết kế và triển khai các hệ thống GenAI tạo ra nội dung thiên vị hoặc phân biệt đối xử dựa trên chủng tộc, quốc tịch, giới tính hoặc các đặc điểm được bảo vệ khác. Họ nên đảm bảo rằng có các 'rào cản' mạnh mẽ để ngăn chặn GenAI tạo ra nội dung phản cảm, thiên vị hoặc sai sự thật, đồng thời đảm bảo rằng những người tham gia vào việc xây dựng các rào cản phải được bảo vệ và không bị khai thác.

- **Khả năng giải thích và tính minh bạch của các mô hình GenAI**: Các nhà cung cấp cần gửi lời giải thích của họ về nguồn gốc, quy mô và loại dữ liệu được sử dụng bởi các mô hình tới các cơ quan quản lý; các nguyên tắc của họ về gắn nhãn dữ liệu trong quá trình huấn luyện trước; các phương pháp hoặc thuật toán mà mô hình của họ sử dụng để tạo nội dung hoặc câu trả lời; và các dịch vụ mà công cụ GenAI của họ đang cung cấp. Khi cần thiết, họ nên hỗ trợ giúp các cơ quan quản lý hiểu hơn công nghệ và dữ liệu. Khuynh hướng của GenAI trong việc tạo ra nội dung kém chính xác và câu trả lời có thể gây tranh cãi nên được làm rõ cho người dùng.

- **Gắn nhãn nội dung GenAI**: Theo luật hoặc quy định liên quan về tổng hợp thông tin trực tuyến được hỗ trợ bởi AI, các nhà cung cấp cần gắn nhãn các bài báo, báo cáo, hình ảnh và video do GenAI tạo ra một cách đúng đắn và hợp pháp. Chẳng hạn, kết quả đầu ra của GenAI nên được gắn nhãn rõ ràng là đã được tạo ra bởi máy.

- **Nguyên tắc an ninh và an toàn**: Các nhà cung cấp GenAI nên đảm bảo dịch vụ an toàn, mạnh mẽ và bền vững trong suốt vòng đời của hệ thống GenAI.

- **Quy định về độ phù hợp cho việc truy cập và sử dụng**: Các nhà cung cấp GenAI nên cung cấp các quy định rõ ràng về đối tượng phù hợp, các kịch bản và mục đích sử dụng dịch vụ của họ và giúp người dùng các công cụ GenAI đưa ra quyết định hợp lý và có trách nhiệm.

- **Thừa nhận hạn chế và ngăn chặn rủi ro có thể dự báo trước:** Các nhà cung cấp GenAI nên quảng cáo rõ ràng về những hạn chế của các phương pháp được sử dụng bởi hệ thống và kết quả đầu ra của chúng. Họ cần phát triển công nghệ để đảm bảo rằng dữ liệu đầu vào, phương pháp và kết quả đầu ra không gây ra bất kỳ tổn hại có thể dự đoán được nào cho người dùng, cùng với các quy trình để giảm thiểu tổn hại không thể dự báo khi chúng xảy ra. Họ cũng phải cung cấp hướng dẫn để giúp người dùng hiểu nội dung do GenAI tạo ra dựa trên các nguyên tắc đạo đức, và ngăn chặn sự phụ thuộc quá mức hay nghiện nội dung tạo sinh.

- **Cơ chế khiếu nại và khắc phục**: Các nhà cung cấp GenAI cần thiết lập các cơ chế và kênh để thu thập khiếu nại từ người dùng và công chúng nói chung, và có các hành động kịp thời để chấp nhận và xử lý những khiếu nại này.

- **Giám sát và báo cáo việc sử dụng bất hợp pháp**: Các nhà cung cấp sẽ hợp tác với các cơ quan quản lý để thúc đẩy giám sát và báo cáo việc sử dụng bất hợp pháp. Điều này bao gồm khi mọi người sử dụng sản phẩm GenAI theo cách bất hợp pháp hoặc vi phạm các giá trị đạo đức hoặc xã hội như lan truyền thông tin

sai lệch hoặc phát ngôn thù địch, tạo thông tin rác hoặc phần mềm độc hại.

3.3.3. Người dùng là tổ chức

Người dùng là tổ chức bao gồm các cơ quan giáo dục và các tổ chức như trường đại học và trường học có trách nhiệm xác định liệu GenAI có nên được áp dụng hay không và loại công cụ GenAI nào nên được mua và triển khai trong tổ chức.

- **Kiểm định thuật toán, dữ liệu và kết quả đầu ra của GenAI trong tổ chức**: Triển khai các cơ chế để giám sát tốt nhất có thể các thuật toán và dữ liệu được sử dụng bởi các công cụ GenAI và các kết quả đầu ra mà chúng tạo ra. Điều này nên bao gồm các cuộc kiểm định và đánh giá thường xuyên, bảo vệ dữ liệu người dùng, và tự động lọc bỏ nội dung không phù hợp.
- **Kiểm chứng tính tương xứng và bảo vệ người dùng**: Triển khai các cơ chế phân loại cấp quốc gia hoặc xây dựng chính sách của tổ chức để phân loại và kiểm chứng các hệ thống và ứng dụng GenAI. Đảm bảo rằng các hệ thống GenAI được tổ chức áp dụng phù hợp với các khung đạo đức được kiểm chứng ở địa phương và không gây ra bất kỳ tổn hại nào có thể dự đoán được cho người dùng mục tiêu của tổ chức, đặc biệt là trẻ em và các nhóm dễ bị tổn thương.
- **Đánh giá và ứng phó với các tác động lâu dài**: Theo thời gian, việc phụ thuộc vào các công cụ hoặc nội dung GenAI trong giáo dục có thể có những ảnh hưởng sâu sắc đến sự phát triển của năng lực con người như kỹ năng tư duy mang tính phản biện và sáng tạo. Những tác động tiềm ẩn này cần được đánh giá và xử lý.
- **Phù hợp với độ tuổi**: Xem xét triển khai các hạn chế về độ tuổi tối thiểu cho việc sử dụng độc lập GenAI trong tổ chức.

3.3.4. Người dùng cá nhân

Người dùng cá nhân có thể bao gồm tất cả những ai có quyền truy cập Internet và ít nhất một loại công cụ GenAI trên toàn cầu. Thuật ngữ 'người dùng cá nhân', như được sử dụng ở đây, chủ yếu đề cập đến các nhà giáo, nhà nghiên cứu và người học cá nhân trong các cơ sở giáo dục chính quy hoặc những người tham gia vào các chương trình học tập không chính quy.

- **Nhận thức về các điều khoản tham chiếu về việc sử dụng GenAI**: Khi ký kết hoặc đồng ý với các thỏa thuận dịch vụ, người dùng nên nhận thức được các nghĩa vụ tuân thủ các Điều khoản tham chiếu được quy định trong thỏa thuận và các luật hoặc quy định đằng sau thỏa thuận.
- **Sử dụng có đạo đức các ứng dụng GenAI**: Người dùng nên sử dụng GenAI một cách có trách nhiệm và tránh khai thác nó theo cách có thể làm tổn hại đến danh tiếng và quyền lợi hợp pháp của người khác.
- **Giám sát và tố cáo các ứng dụng GenAI bất hợp pháp**: Khi phát hiện các ứng dụng GenAI vi phạm một hoặc nhiều quy định, người dùng nên thông báo cho các cơ quan quản lý chính phủ.

4. Hướng tới một khung chính sách cho việc sử dụng AI tạo sinh trong giáo dục và nghiên cứu

Việc quản trị GenAI để khai thác các lợi ích tiềm năng cho giáo dục và nghiên cứu đòi hỏi việc phát triển các chính sách phù hợp. Dữ liệu khảo sát năm 2023 được trích dẫn ở trên cho thấy chỉ có một số ít quốc gia đã áp dụng các chính sách hoặc kế hoạch cụ thể cho việc sử dụng AI trong giáo dục. Phần trước đã phác thảo một tầm nhìn, các bước cần thiết và các yếu tố và hành động chính có thể được triển khai bởi các bên liên quan khác nhau. Phần này cung cấp các biện pháp có thể được triển khai để phát triển các khung chính sách mạch lạc, toàn diện để quy định việc sử dụng GenAI trong giáo dục và nghiên cứu.

Một điểm khởi đầu là tài liệu *AI và giáo dục: Hướng dẫn cho các nhà hoạch định chính sách năm 2022* (UNESCO, 2022*b*). Nó đề xuất một bộ khuyến nghị toàn diện để hướng dẫn các chính phủ trong việc phát triển và triển khai các chính sách trên toàn ngành về AI và giáo dục với trọng tâm là thúc đẩy chất lượng giáo dục, công bằng xã hội và hòa nhập. Hầu hết các khuyến nghị vẫn có thể áp dụng và có thể được điều chỉnh thêm để hướng dẫn việc xây dựng các chính sách cụ thể về GenAI trong giáo dục. Tám biện pháp cụ thể sau đây cho việc lập kế hoạch chính sách về GenAI trong giáo dục và nghiên cứu được đề xuất ở đây để bổ sung cho hướng dẫn hiện có này.

4.1 Thúc đẩy sự hòa nhập, công bằng, và đa dạng ngôn ngữ và văn hóa

Tầm quan trọng có tính quyết định của sự hòa nhập phải được công nhận và đảm bảo trong suốt vòng đời của GenAI. Cụ thể hơn, các công cụ GenAI sẽ không giúp giải quyết những thách thức cơ bản trong giáo dục hoặc việc đạt được các cam kết MTPTBV 4 trừ khi các công cụ đó được tiếp cận một cách hòa nhập (bất kể giới tính, dân tộc, nhu cầu giáo dục đặc biệt, tình trạng kinh tế-xã hội, vị trí địa lý, tình trạng di tản, v.v.), và nếu chúng không được thiết kế để thúc đẩy sự công bằng, đa dạng ngôn ngữ và văn hóa. Để đạt được điều này, 3 biện pháp chính sách sau đây được khuyến nghị:

- Xác định những người không có hoặc không thể chi trả cho kết nối Internet hoặc dữ liệu, và có hành động để thúc đẩy kết nối toàn cầu và năng lực số để giảm bớt các rào cản đối với việc tiếp cận công bằng và hòa nhập các ứng dụng AI. Thiết lập các cơ chế tài trợ bền vững cho việc phát triển và cung cấp các công cụ hỗ trợ AI cho người học khuyết tật hoặc có nhu cầu đặc biệt. Thúc đẩy việc sử dụng GenAI để hỗ trợ người học suốt đời ở mọi lứa tuổi, địa điểm và hoàn cảnh
- Phát triển tiêu chí để kiểm chứng các hệ thống GenAI nhằm đảm bảo không có sự thiên vị giới tính, phân biệt đối xử chống lại các nhóm yếu thế, hoặc phát ngôn thù địch sẵn có trong dữ liệu hoặc thuật toán.
- Phát triển và triển khai các thông số kỹ thuật bao trùm cho các hệ thống GenAI và triển khai các biện pháp thể chế để bảo vệ sự đa dạng ngôn ngữ và văn hóa khi triển khai GenAI trong giáo dục và nghiên cứu ở quy mô lớn. Các thông số kỹ thuật liên quan nên yêu cầu các nhà cung cấp GenAI đưa dữ liệu bằng nhiều ngôn ngữ, đặc biệt là các ngôn ngữ địa phương hoặc bản địa, vào việc huấn luyện các mô hình GPT để cải thiện năng lực của GenAI trong việc trả lời và tạo ra văn bản đa ngôn ngữ. Các thông số kỹ thuật và biện pháp thể chế cần nghiêm khắc ngăn chặn các nhà cung cấp AI khỏi bất kỳ hành vi cố ý hoặc vô ý loại bỏ các ngôn ngữ thiểu số hoặc phân biệt đối xử đối với người nói các ngôn ngữ bản địa, và yêu cầu các nhà cung cấp ngừng các hệ thống thúc đẩy các ngôn ngữ hoặc chuẩn mực văn hóa nổi trội.

4.2 Bảo vệ quyền tự chủ của con người

Khi GenAI ngày càng trở nên tinh vi, một mối nguy hiểm chính là tiềm năng làm suy yếu quyền tự chủ của con người. Khi ngày càng nhiều cá nhân sử dụng GenAI để hỗ trợ viết lách hoặc các hoạt động sáng tạo khác, họ có thể vô tình trở nên phụ thuộc vào nó. Điều này có thể ảnh hưởng đến việc phát triển các kỹ năng tư duy. Mặc dù GenAI có thể được sử dụng để thách thức và mở rộng năng lực tư duy của con người, nhưng không nên cho phép nó chiếm đoạt tư duy của con người. Việc bảo vệ và tăng cường quyền tự chủ của con người

luôn nên là những cân nhắc cốt yếu khi thiết kế và áp dụng GenAI từ 7 góc độ sau:

- Thông báo cho người học về các loại dữ liệu mà GenAI có thể thu thập từ họ, cách sử dụng những dữ liệu này, và tác động mà nó có thể gây ra đối với giáo dục và cuộc sống của họ nói chung.
- Duy trì động lực nội tại của người học để phát triển và học hỏi với tư cách cá nhân. Củng cố quyền tự chủ của con người đối với cách tiếp cận nghiên cứu, giảng dạy và học tập của chính họ trong bối cảnh sử dụng các hệ thống GenAI ngày càng tinh vi.
- Ngăn chặn việc sử dụng GenAI khi nó có thể lấy mất cơ hội của người học trong việc phát triển năng lực nhận thức và kỹ năng xã hội thông qua quan sát thế giới thực, các thực hành thực nghiệm như thí nghiệm, thảo luận với người khác, và tư duy logic độc lập.
- Sử dụng các công cụ GenAI để giảm thiểu áp lực của bài tập về nhà và các kỳ thi, thay vì khiến nó thêm trầm trọng.
- Tham khảo ý kiến của các nhà nghiên cứu, nhà giáo và người học về quan điểm của họ về GenAI và sử dụng phản hồi để quyết định liệu có nên và cách thức triển khai các công cụ GenAI ở quy mô tổ chức. Khuyến khích người học, nhà giáo và nhà nghiên cứu phê bình và đặt câu hỏi về các phương pháp đằng sau các hệ thống AI, độ chính xác của nội dung kết quả đầu ra, và các chuẩn mực hoặc phương pháp sư phạm mà chúng có thể tạo ra.
- Ngăn chặn việc nhượng lại trách nhiệm giải trình của con người cho các hệ thống GenAI khi đưa ra các quyết định trọng yếu.

4.3 Giám sát và kiểm chứng các hệ thống GenAI cho giáo dục

Như đã nêu, việc phát triển và triển khai GenAI cần có tính đạo đức ngay từ khâu thiết kế. Sau đó, một khi GenAI đã được sử dụng, và trong suốt vòng đời của nó, cần phải được giám sát và kiểm chứng cẩn thận - về các rủi ro đạo đức, tính phù hợp và chính xác về mặt sư phạm, và tác động của nó đối với người học, nhà giáo và mối quan hệ trong lớp học/trường học. Về mặt này, 5 hành động sau đây được khuyến nghị:

- Xây dựng cơ chế kiểm định để kiểm tra xem các hệ thống GenAI được sử dụng trong giáo dục và nghiên cứu có không có thiên kiến, đặc biệt là thiên kiến giới, và liệu chúng có được huấn luyện trên dữ liệu đại diện cho sự đa dạng hay không (về giới tính, khuyết tật, tình trạng xã hội và kinh tế, thành phần dân tộc và nền tảng văn hóa, và vị trí địa lý).
- Giải quyết vấn đề phức tạp của sự đồng thuận dựa trên hiểu biết, đặc biệt là trong các bối cảnh mà trẻ em hoặc người học dễ bị tổn thương khác không có khả năng đưa ra sự đồng ý thực sự dựa trên hiểu biết.
- Kiểm tra xem đầu ra của GenAI có bao gồm hình ảnh deepfake, tin giả (không chính xác hoặc sai sự thật), hoặc phát ngôn thù địch hay không. Nếu phát hiện GenAI đang tạo ra nội dung không phù hợp, các tổ chức và nhà giáo dục nên sẵn sàng và có khả năng triển khai hành động nhanh chóng và mạnh mẽ để giảm thiểu hoặc loại bỏ vấn đề.
- Triển khai kiểm tra nghiêm ngặt về mặt đạo đức đối với các ứng dụng GenAI trước khi chúng được chính thức áp dụng trong các cơ sở giáo dục hoặc nghiên cứu (tức là áp dụng cách tiếp cận đạo đức ngay từ khâu thiết kế).
- Trước khi đưa ra quyết định về việc áp dụng cho tổ chức, đảm bảo rằng các ứng dụng GenAI đang được xem xét không gây hại có thể dự báo được cho người học, có hiệu quả về mặt giáo dục và phù hợp với độ tuổi và năng lực của đối tượng người học mục tiêu, và phù hợp với các nguyên tắc sư phạm hợp lý (tức là dựa trên các lĩnh vực kiến thức liên quan và chuẩn đầu ra được mong đợi và xây dựng hệ giá trị).

4.4 Phát triển năng lực AI bao gồm các kỹ năng liên quan đến GenAI cho người học

Việc phát triển năng lực AI cho người học là chìa khóa để sử dụng AI một cách an toàn, có đạo đức và có ý nghĩa trong giáo dục và hơn thế nữa. Tuy nhiên, theo dữ liệu của UNESCO, chỉ có khoảng 15 quốc gia đã phát triển và triển khai, hoặc đang trong quá trình phát triển, chương trình giảng dạy AI được chính phủ thông qua trong trường học vào đầu năm 2022 (UNESCO, 2022*c*). Những phát triển mới nhất của GenAI đã càng củng cố thêm nhu cầu cấp thiết cho mọi người đạt được mức độ hiểu biết phù hợp về cả khía cạnh con người và công nghệ của AI, hiểu cách nó hoạt động nói chung, cũng

như tác động cụ thể của GenAI. Để làm được điều này, 5 hành động sau đây hiện đang cấp thiết cần thiết:

- Cam kết cung cấp chương trình giảng dạy AI được chính phủ chấp thuận cho giáo dục phổ thông, trong giáo dục và huấn luyện kỹ thuật và nghề nghiệp, cũng như cho học tập suốt đời. Chương trình giảng dạy AI nên bao gồm tác động của AI đến cuộc sống của chúng ta, bao gồm cả các vấn đề đạo đức mà nó đặt ra, cũng như hiểu biết phù hợp với lứa tuổi về thuật toán và dữ liệu, và kỹ năng sử dụng đúng và sáng tạo các công cụ AI bao gồm cả các ứng dụng GenAI;
- Hỗ trợ các cơ sở giáo dục đại học và các cơ quan nghiên cứu tăng cường các chương trình nhằm phát huy tiềm năng AI sở tại
- Thúc đẩy bình đẳng giới trong việc phát triển năng lực AI nâng cao và tạo ra một nhóm chuyên gia có độ cân bằng về giới tính;
- Xây dựng các dự báo liên ngành về sự chuyển dịch việc làm quốc gia và toàn cầu gây ra bởi sự tự động hóa GenAI mới nhất, và tăng cường các kỹ năng có thể thích ứng với tương lai ở tất cả các cấp của hệ thống giáo dục và học tập suốt đời dựa trên những thay đổi dự kiến về nhu cầu; và
- Cung cấp các chương trình đặc biệt cho người lao động lớn tuổi và công dân có thể cần học kỹ năng mới để thích ứng với môi trường mới.

4.5 Phát triển năng lực cho nhà giáo và nhà nghiên cứu để sử dụng GenAI đúng cách

Theo dữ liệu khảo sát năm 2023 về việc sử dụng AI của chính phủ cho giáo dục (UNESCO, 2023*c*), chỉ có khoảng 7 quốc gia (Trung Quốc, Phần Lan, Gru-di-a, Ca-ta, Tây Ban Nha, Thái Lan và Thổ Nhĩ Kỳ) báo cáo rằng họ đã phát triển hoặc đang phát triển các khung hoặc chương trình huấn luyện kĩ năng AI nhà giáo. Chỉ có Bộ Giáo dục Xinh-ga-po báo cáo đã xây dựng một kho lưu trữ trực tuyến tập trung vào việc sử dụng ChatGPT trong giảng dạy và học tập. Điều này cho thấy rõ ràng rằng nhà giáo ở hầu hết các quốc gia không tiếp cận được các chương trình giảng dạy về việc sử dụng AI trong giáo dục được thiết kế bài bản, chưa kể đến việc sử dụng GenAI.

Để chuẩn bị cho nhà giáo sử dụng GenAI một cách có trách nhiệm và hiệu quả, các quốc gia cần triển khai 4 hành động sau:

- Xây dựng hoặc điều chỉnh hướng dẫn dựa trên các bài kiểm tra địa phương để giúp các nhà nghiên cứu và nhà giáo định hướng các công cụ GenAI phổ biến, và chỉ đạo việc thiết kế các ứng dụng AI chuyên biệt mới.
- Bảo vệ quyền của nhà giáo và nhà nghiên cứu và giá trị của các công việc của họ khi sử dụng GenAI. Cụ thể hơn, phân tích vai trò đặc thù của nhà giáo trong việc thúc đẩy tư duy bậc cao, tổ chức tương tác giữa người với người, và nuôi dưỡng các giá trị nhân văn.
- Xác định định hướng giá trị, kiến thức và kỹ năng mà nhà giáo cần có để hiểu và sử dụng các hệ thống GenAI một cách hiệu quả và có đạo đức. Cho phép nhà giáo tạo ra các công cụ cụ thể dựa trên GenAI để đẩy mạnh học tập trong lớp học và trong phát triển chuyên môn của chính họ.
- Liên tục xem xét các năng lực cần thiết để nhà giáo hiểu và sử dụng AI cho việc giảng dạy, học tập và phát triển chuyên môn của họ, và tích hợp các bộ giá trị, hiểu biết và kỹ năng mới nổi về AI vào các khung năng lực và chương trình huấn luyện nhà giáo đang công tác và nhà giáo mới ra trường.

4.6 Thúc đẩy ý kiến đa dạng và biểu đạt đa dạng của ý tưởng

Như đã lưu ý trước đây, GenAI không hiểu cả câu lệnh lẫn câu trả lời. Thay vào đó, câu trả lời của nó dựa trên xác suất của các mẫu ngôn ngữ được tìm thấy trong dữ liệu (từ Internet) mà nó đã tiếp thu khi mô hình của nó được huấn luyện. Để giải quyết một số vấn đề cơ bản của kết quả đầu ra của GenAI, các phương pháp mới hiện đang được nghiên cứu như kết nối GenAI với các cơ sở dữ liệu kiến thức và công cụ lập luận. Tuy nhiên, do cách thức hoạt động, nguồn tài liệu và quan điểm ngầm định của các nhà phát triển, GenAI hiển nhiên chỉ thể hiện thế giới quan nổi trội trong kết quả đầu ra của nó và hạn chế các ý kiến thiểu số và đa dạng khác. Do đó, nếu nền văn minh nhân loại muốn phát triển, điều cần thiết là chúng ta phải nhận ra rằng GenAI không bao giờ có thể là nguồn kiến thức có thẩm quyền về bất kỳ chủ đề nào mà nó tham gia.

Kết quả là, người dùng cần xem xét kết quả đầu ra của GenAI một cách mang tính phản biện. Cụ thể là:

- Hiểu vai trò của GenAI như một nguồn thông tin nhanh chóng nhưng thường không đáng

tin cậy. Mặc dù một số ứng dụng bổ sung (plugin) và công cụ dựa trên LLM được đề cập trước đó được thiết kế để hỗ trợ nhu cầu truy cập thông tin đã được kiểm chứng và cập nhật, nhưng hiện có rất ít bằng chứng vững chắc cho thấy những công cụ này hiệu quả.

- Khuyến khích người học và nhà nghiên cứu phê bình các câu trả lời do GenAI cung cấp. Nhận ra rằng GenAI thường chỉ lặp lại các ý kiến đã được thiết lập hoặc tiêu chuẩn, do đó làm suy yếu các ý kiến đa dạng và thiểu số cũng như cách biểu đạt đa dạng của ý tưởng.
- Cho người học đủ cơ hội để học hỏi từ quá trình thử nghiệm và mắc lỗi, thí nghiệm thực nghiệm, cũng như việc quan sát thế giới thực.

4.7 Thử nghiệm các mô hình ứng dụng phù hợp cho từng địa phương và xây dựng cơ sở bằng chứng tích lũy

Các mô hình GenAI cho đến nay bị chi phối bởi thông tin từ Bắc bán cầu và không đại diện đầy đủ cho tiếng nói từ Nam bán cầu và các cộng đồng bản địa. Chỉ bằng những nỗ lực quyết tâm, ví dụ như khai thác dữ liệu tổng hợp (Marwala, 2023), các công cụ GenAI mới có thể trở nên nhạy cảm hơn với bối cảnh và nhu cầu của các cộng đồng bản địa, đặc biệt là những cộng đồng ở Nam bán cầu. Để khám phá các cách tiếp cận phù hợp với nhu cầu địa phương, đồng thời hợp tác rộng rãi hơn, 8 hành động sau đây được khuyến nghị:

- Đảm bảo việc thiết kế và áp dụng GenAI được lên kế hoạch chiến lược thay vì thúc đẩy cho quá trình mua sắm thụ động và dễ dãi.
- Khuyến khích các nhà thiết kế GenAI nhắm đến các tùy chọn học tập mở, có tính khai phá và có nhiều lựa chọn đa dạng.
- Thử nghiệm và mở rộng quy mô các trường hợp sử dụng dựa trên bằng chứng về việc áp dụng AI trong giáo dục và nghiên cứu phù hợp với các ưu tiên giáo dục, thay vì dựa vào sự mới lạ, huyền thoại hoặc sự phóng đại.
- Hướng dẫn việc sử dụng GenAI để kích hoạt đổi mới trong nghiên cứu, bao gồm cả việc tận dụng năng lực tính toán, dữ liệu quy mô lớn, và kết quả đầu ra của GenAI để tạo cơ sở và truyền cảm hứng cho việc cải thiện các phương pháp nghiên cứu.
- Xem xét các tác động xã hội và đạo đức của việc kết hợp GenAI vào quá trình nghiên cứu.
- Thiết lập các tiêu chí cụ thể dựa trên nghiên cứu và phương pháp sư phạm đã được kiểm chứng và xây dựng cơ sở bằng chứng về hiệu quả của GenAI trong việc hỗ trợ cung cấp cơ hội học tập hòa nhập, đáp ứng mục tiêu học tập và nghiên cứu, và thúc đẩy đa dạng ngôn ngữ và văn hóa.
- Triển khai các bước lặp đi lặp lại để củng cố bằng chứng về tác động xã hội và đạo đức của GenAI.
- Phân tích chi phí môi trường của việc tận dụng công nghệ AI ở quy mô lớn (ví dụ: năng lượng và tài nguyên cần thiết để huấn luyện các mô hình GPT), và phát triển các mục tiêu bền vững mà các nhà cung cấp AI phải đáp ứng trong nỗ lực tránh góp phần làm trầm trọng thêm biến đổi khí hậu.

4.8 Xem xét các tác động lâu dài theo cách liên ngành và liên lĩnh vực

Cách tiếp cận liên ngành và liên lĩnh vực là cần thiết để sử dụng GenAI một cách hiệu quả và có đạo đức trong giáo dục và nghiên cứu. Chỉ thông qua khai thác kiến thức chuyên môn đa dạng, đồng thời tập hợp nhiều bên liên quan, mới có thể xác định kịp thời và giải quyết hiệu quả các thách thức chính để giảm thiểu tác động tiêu cực lâu dài trong khi tận dụng được lợi ích liên tục có tính tích lũy. Do đó, 3 hành động sau đây được khuyến nghị:

- Hợp tác với các nhà cung cấp AI, nhà giáo dục, nhà nghiên cứu, và đại diện của phụ huynh và người học để lập kế hoạch điều chỉnh toàn hệ thống trong các khung chương trình giảng dạy và phương pháp đánh giá, để tận dụng đầy đủ tiềm năng và giảm thiểu rủi ro của GenAI đối với giáo dục và nghiên cứu.
- Tập hợp chuyên môn liên ngành và liên lĩnh vực bao gồm các nhà giáo dục, nhà nghiên cứu, nhà khoa học học tập, kỹ sư AI, và đại diện của các bên liên quan khác để xem xét các tác động lâu dài của GenAI đối với việc học tập và tạo ra kiến thức, nghiên cứu và bản quyền, chương trình giảng dạy và đánh giá, và sự hợp tác của con người và động lực xã hội.
- Cung cấp tư vấn kịp thời để cho quá trình cập nhật liên tục các quy định và chính sách.

5. Thúc đẩy sử dụng GenAI theo cách sáng tạo trong giáo dục và nghiên cứu

Khi ChatGPT lần đầu tiên ra mắt, các nhà giáo dục trên toàn thế giới đã bày tỏ lo ngại về khả năng viết bài luận của nó và những cách nó có thể giúp người học gian lận. Gần đây hơn, nhiều người và nhiều tổ chức, bao gồm cả một số trường đại học hàng đầu thế giới, đã lập luận rằng 'thần đèn đã được giải thoát' và các công cụ như ChatGPT sẽ tồn tại và có thể được sử dụng hiệu quả trong môi trường giáo dục. Trong khi đó, Internet hiện đang tràn ngập các đề xuất sử dụng GenAI trong giáo dục và nghiên cứu. Những đề xuất này bao gồm sử dụng nó để truyền cảm hứng cho các ý tưởng mới, tạo ra ví dụ đa chiều, phát triển kế hoạch bài giảng và bài thuyết trình, tóm tắt tài liệu hiện có, và kích thích việc tạo ra hình ảnh. Mặc dù các ý tưởng mới xuất hiện trên Internet gần như mỗi ngày, các nhà nghiên cứu và nhà giáo dục vẫn đang tìm hiểu chính xác GenAI có hàm ý gì đối với việc giảng dạy, học tập và nghiên cứu. Cụ thể hơn, những người đã đưa ra nhiều đề xuất sử dụng có thể chưa cân nhắc đúng mức các nguyên tắc đạo đức, trong khi những người khác bị thôi thúc bởi tiềm năng công nghệ của GenAI hơn là nhu cầu của nhà nghiên cứu, nhà giáo hoặc người học. Phần này sẽ chỉ ra một số cách thức nhằm thúc đẩy việc sử dụng sáng tạo GenAI trong giáo dục.

5.1 Chiến lược của tổ chức nhằm thúc đẩy sử dụng GenAI có trách nhiệm và sáng tạo

Như đã nêu trên, các tổ chức giáo dục và nghiên cứu nên phát triển, triển khai và kiểm chứng các chiến lược và khung đạo đức phù hợp để hướng dẫn việc sử dụng có trách nhiệm và đạo đức các hệ thống và ứng dụng GenAI để đáp ứng nhu cầu giảng dạy, học tập và nghiên cứu. Điều này có thể đạt được thông qua 4 chiến lược sau:

- **Triển khai các nguyên tắc đạo đức trong tổ chức**: Đảm bảo rằng các nhà nghiên cứu, nhà giáo và người học sử dụng các công cụ GenAI một cách có trách nhiệm và đạo đức, và tiếp cận một cách cẩn trọng về độ chính xác và tính hợp lệ của các kết quả đầu ra.
- **Hướng dẫn và huấn luyện**: Cung cấp hướng dẫn và huấn luyện cho các nhà nghiên cứu, nhà giáo và người học về các công cụ GenAI để đảm bảo rằng họ hiểu các vấn đề đạo đức như thiên kiến trong gắn nhãn dữ liệu và thuật toán, và rằng họ tuân thủ các quy định thích hợp về quyền riêng tư dữ liệu và sở hữu trí tuệ.
- **Xây dựng năng lực kỹ thuật soạn câu lệnh GenAI**: Ngoài kiến thức chuyên môn về môn học, các nhà nghiên cứu và nhà giáo cũng sẽ cần chuyên môn trong việc thiết kế và đánh giá kĩ lưỡng các câu lệnh được tạo ra bởi GenAI. Với những thách thức phức tạp do GenAI đặt ra, các nhà nghiên cứu và nhà giáo phải nhận được sự hỗ trợ và huấn luyện chất lượng cao để làm điều này.
- **Phát hiện đạo văn dựa trên GenAI trong bài tập viết:** GenAI có thể cho phép người học tạo ra các văn bản như thể là của các em tự viết, một loại "đạo văn" mới. Các nhà cung cấp GenAI được yêu cầu gắn nhãn kết quả đầu ra của họ bằng dấu watermark "được tạo bởi AI", trong khi các công cụ đang được phát triển để xác định tài liệu đã được sản xuất bởi AI. Tuy nhiên, có ít bằng chứng cho thấy các biện pháp hoặc công cụ này phát huy hiệu quả. Chiến lược tức thời của tổ chức là duy trì tính chính trực học thuật và tăng cường trách nhiệm giải trình thông qua các nỗ lực phát hiện sát sao bởi con người. Chiến lược dài hạn là để các tổ chức và nhà giáo dục tư duy lại về thiết kế của các bài tập viết để chúng không được sử dụng để đánh giá các nhiệm vụ mà các công cụ GenAI có thể làm tốt hơn người học. Thay vào đó, chúng nên giải quyết những gì con người có thể làm mà GenAI và các công cụ AI khác không thể làm, bao gồm áp dụng các giá trị của con người như lòng trắc ẩn và sáng tạo vào các thách thức phức tạp trong thế giới thực.

5.2 Cách tiếp cận sư phạm 'lấy con người làm trung tâm và tương tác phù hợp'

Các nhà nghiên cứu và nhà giáo dục nên ưu tiên quyền tự chủ của con người và tương tác có trách nhiệm, phù hợp về mặt sư phạm giữa con người và các công cụ AI khi quyết định có nên và cách sử dụng GenAI. Điều này bao gồm 5 cân nhắc sau:

- việc sử dụng (các) công cụ nên đóng góp vào nhu cầu của con người và làm cho việc học tập hoặc nghiên cứu hiệu quả hơn so với cách tiếp cận không sử dụng công nghệ hoặc cách tiếp cận thay thế khác;
- việc sử dụng (các) công cụ của người làm giáo dục và người học cần dựa trên động lực nội tại của họ;
- quá trình sử dụng (các) công cụ nên được kiểm soát bởi nhà giáo, người học hoặc nhà nghiên cứu là con người;
- việc lựa chọn và tổ chức (các) công cụ và nội dung chúng tạo ra nên tương xứng, dựa trên phạm vi độ tuổi của người học, kết quả mong đợi, và loại kiến thức mục tiêu (ví dụ: thực tế, khái niệm, thủ tục, hoặc siêu nhận thức) hoặc vấn đề mục tiêu (ví dụ: có cấu trúc tốt hoặc cấu trúc kém);
- các quy trình sử dụng nên đảm bảo sự tham gia tương tác của con người với GenAI và tư duy bậc cao, cũng như trách nhiệm giải trình của con người đối với các quyết định liên quan đến độ chính xác của nội dung do AI tạo ra, chiến lược giảng dạy hoặc nghiên cứu, và tác động của chúng đối với hành vi của con người.

5.3 Đồng thiết kế việc sử dụng GenAI trong giáo dục và nghiên cứu

Việc sử dụng GenAI trong giáo dục và nghiên cứu không nên bị áp đặt theo cách tiếp cận từ trên xuống hoặc bị thúc đẩy bởi sự phóng đại nhằm mục đích thương mại. Thay vào đó, việc sử dụng an toàn và hiệu quả nên được đồng thiết kế bởi nhà giáo, người học và nhà nghiên cứu. Nó cũng cần một quá trình thử nghiệm và kiểm tra kĩ lưỡng nhằm đánh giá mức độ hiệu quả và tác động lâu dài của các phương thức sử dụng khác nhau.

Để thúc đẩy việc đồng thiết kế như được khuyến nghị, Hướng dẫn này đề xuất một khung gồm 6 quan điểm dưới đây để củng cố tương tác phù hợp về mặt sư phạm và ưu tiên quyền tự chủ của con người:

- các lĩnh vực kiến thức hoặc vấn đề phù hợp;
- các kết quả mong đợi;
- các công cụ GenAI phù hợp và lợi thế so sánh;
- các yêu cầu đối với người dùng;
- các yêu cầu phương pháp sư phạm của con người và ví dụ về các câu lệnh; và
- nguy cơ về đạo đức.

Phần này cung cấp ví dụ về cách quá trình đồng thiết kế trong việc sử dụng GenAI có thể thông báo cho các thực hành nghiên cứu, hỗ trợ trong giảng dạy, cung cấp huấn luyện cho việc tự học các kỹ năng nền tảng, thúc đẩy tư duy bậc cao và hỗ trợ người học có nhu cầu đặc biệt. Những ví dụ này chỉ đại diện cho phần nổi của tảng băng chìm về số lượng ngày càng tăng các lĩnh vực mà GenAI có thể có tác động.

5.3.1 AI tạo sinh cho nghiên cứu

Các mô hình GenAI đã chứng minh tiềm năng của chúng trong việc mở rộng quan điểm về đề cương nghiên cứu và làm phong phú thêm việc khám phá dữ liệu cũng như tìm hiểu lịch sử nghiên cứu (xem **Bảng 3**). Mặc dù có thể xuất hiện nhiều trường hợp sử dụng hơn, nghiên cứu mới cần thiết để xác định lĩnh vực tiềm năng của các vấn đề nghiên cứu và kết quả mong đợi, để chứng minh hiệu quả và độ chính xác, và để đảm bảo rằng quyền tự chủ của con người trong việc hiểu thế giới thực thông qua nghiên cứu sẽ không bị ảnh hưởng bởi việc sử dụng các công cụ AI.

Bảng 3. Các trường hợp sử dụng GenAI đồng thiết kế cho nghiên cứu

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Cố vấn AI cho đề cương nghiên cứu	Có thể hữu ích trong các lĩnh vực vấn đề nghiên cứu có cấu trúc tốt.	Phát triển và trả lời các câu hỏi nghiên cứu, đề xuất phương pháp luận phù hợp. **Tiềm năng chuyển đổi: Huấn luyện viên 1:1 cho việc lập kế hoạch nghiên cứu**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Nhà nghiên cứu phải có hiểu biết cơ bản về (các) chủ đề. Nhà nghiên cứu nên phát triển khả năng xác minh thông tin, và đặc biệt có khả năng phát hiện trích dẫn các bài báo nghiên cứu không tồn tại.	Ý tưởng cơ bản cho việc xác định vấn đề nghiên cứu (ví dụ: đối tượng mục tiêu, vấn đề, bối cảnh), cũng như phương pháp luận, kết quả mong đợi và định dạng. Ví dụ về lời nhắc: *Viết 10 câu hỏi nghiên cứu tiềm năng cho [chủ đề x] và xếp hạng chúng theo tầm quan trọng đối với [lĩnh vực nghiên cứu y].*	Cần cảnh giác với rủi ro cao từ việc GenAI tạo ra thông tin sai lệch (chẳng hạn như các ấn phẩm nghiên cứu không tồn tại), hay việc người dùng bị cám dỗ sao chép và dán đề cương nghiên cứu do AI tạo ra, điều này có thể giảm bớt cơ hội học hỏi từ thử nghiệm và mắc lỗi của các nhà nghiên cứu trẻ.
Khám phá dữ liệu tạo sinh và lịch sử nghiên cứu	Có thể hữu ích trong các lĩnh vực vấn đề nghiên cứu có cấu trúc lỏng lẻo.	Tự động thu thập thông tin, khám phá một loạt dữ liệu rộng lớn, đề xuất bản thảo đánh giá tài liệu, và tự động hóa một phần của việc diễn giải dữ liệu. **Tiềm năng chuyển đổi: Huấn luyện viên AI cho việc khai phá dữ liệu và tìm hiểu lịch sử nghiên cứu**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Các nhà nghiên cứu phải có kiến thức vững chắc về phương pháp luận và kỹ thuật phân tích dữ liệu.	Các định nghĩa tiến bộ về vấn đề, phạm vi dữ liệu và nguồn tài liệu, phương pháp được sử dụng cho việc khám phá dữ liệu và đánh giá tài liệu, và kết quả mong đợi cũng như định dạng của chúng.	Cần cảnh giác với thông tin do GenAI tạo ra, việc xử lý dữ liệu không đúng cách, khả năng vi phạm quyền riêng tư, lập hồ sơ trái phép và định kiến giới. Cần cảnh giác với việc lan truyền các chuẩn mực nổi trội và mối đe dọa của chúng đối với các chuẩn mực thay thế và các ý kiến đa dạng.

5.3.2 AI tạo sinh để hỗ trợ giảng dạy

Việc sử dụng cả nền tảng GenAI chung và các công cụ GenAI giáo dục chuyên dụng nên được thiết kế để nâng cao hiểu biết của nhà giáo về môn học của họ cũng như kiến thức về phương pháp giảng dạy, bao gồm cả việc nhà giáo-AI cùng thiết kế bài giảng, gói khóa học, hoặc toàn bộ chương trình giáo dục. Các trợ giảng được hỗ trợ bởi GenAI hoặc 'bản sao tạo sinh của trợ giảng'[53] được huấn luyện trước dựa trên dữ liệu từ nhà giáo có kinh nghiệm và thư viện, đã được thử nghiệm trong một số cơ sở giáo dục và có thể chứa tiềm năng chưa được khám phá cũng như rủi ro đạo đức chưa được biết đến. Các quy trình ứng dụng thực tế và các bản cập nhật tiếp theo của các mô hình này vẫn cần được kiểm tra cẩn thận thông qua khung được khuyến nghị trong Hướng dẫn này và được bảo vệ bởi sự giám sát của con người như được minh họa trong **Bảng 4**.

Bảng 4. Các trường hợp sử dụng GenAI đồng thiết kế nhằm hỗ trợ nhà giáo và giảng dạy

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Đồng thiết kế khoá học hoặc chương trình giáo dục	Kiến thức khái niệm về một số chủ đề giảng dạy nhất định và kiến thức thủ tục về phương pháp giảng dạy.	Hỗ trợ quá trình thiết kế chương trình giảng dạy và bài học, bao gồm phác thảo hoặc mở rộng quan điểm về các lĩnh vực chính của chủ đề mục tiêu và xác định cấu trúc chương trình giáo dục. Nó cũng có thể giúp nhà giáo chuẩn bị các bài kiểm tra và kỳ thi bằng cách đưa ra ví dụ về câu hỏi và tiêu chí đánh giá. **Tiềm năng chuyển đổi: Chương trình giảng dạy do AI tạo ra**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Các nhà giáo phải hiểu và chỉ định cẩn thận những gì họ muốn chương trình giáo dục, khóa học, bài học hoặc bài kiểm tra bao gồm và đạt được, liệu họ muốn giải quyết kiến thức thủ tục hay khái niệm, và lý thuyết giảng dạy nào họ muốn áp dụng.	Câu hỏi cho GenAI về việc đề xuất cấu trúc và ví dụ về kiến thức thực tế về (các) chủ đề, đề xuất phương pháp và quy trình giảng dạy cho các chủ đề hoặc vấn đề, hoặc tạo gói khóa học hoặc kế hoạch bài giảng dựa trên (các) chủ đề và định dạng Các nhà thiết kế chương trình giáo dục là con người cần xác minh kiến thức thực tế và kiểm tra tính phù hợp của các gói khóa học được đề xuất.	Rủi ro GenAI áp đặt các chuẩn mực và phương pháp sư phạm chiếm ưu thế là cao. Nó có thể vô tình duy trì các thực hành loại trừ có lợi cho các nhóm giàu dữ liệu và củng cố bất bình đẳng trong tiếp cận cơ hội giáo dục phù hợp và chất lượng cao, gây bất lợi cho các nhóm nghèo dữ liệu.
Chatbot tạo sinh trong vai trò trợ giảng	Kiến thức khái niệm trên nhiều lĩnh vực trong các vấn đề có cấu trúc tốt.	Cung cấp hỗ trợ cá nhân hóa, trả lời câu hỏi và xác định tài nguyên. **Tiềm năng chuyển đổi: Bản sao tạo sinh của trợ giảng**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Nó hỗ trợ nhà giáo nhưng nhắm trực tiếp đến người học, vì vậy điều này đòi hỏi người học phải có đủ kiến thức trước, khả năng và kỹ năng meta-nhận thức để xác minh kết quả đầu ra của GenAI và nhận thấy thông tin sai lệch. Do đó, nó có thể phù hợp hơn cho người học ở bậc giáo dục đại học.	Yêu cầu nhà giáo hiểu rõ vấn đề, giám sát cuộc trò chuyện và giúp người học xác minh các câu trả lời đáng ngờ do GenAI cung cấp.	Dựa trên khả năng hiện tại của các mô hình GenAI, các cơ sở giáo dục cần đảm bảo sự giám sát của con người đối với các câu trả lời do các công cụ GenAI cung cấp, cảnh giác với rủi ro của thông tin sai lệch. Nó cũng có thể hạn chế việc người học tiếp cận hướng dẫn và hỗ trợ của con người, cản trở sự phát triển mối quan hệ bền chặt giữa nhà giáo và người học, một điều này đặc biệt đáng lo ngại đối với trẻ em.

5.3.3 AI tạo sinh như một huấn luyện viên 1:1 cho việc tự học các kỹ năng nền tảng

Mặc dù tư duy bậc cao và sáng tạo đã thu hút sự chú ý ngày càng tăng khi xác định chuẩn đầu ra, không thể phủ nhận tầm quan trọng của các kỹ năng nền tảng trong sự phát triển tâm lý và tiến trình năng lực của trẻ em. Trong một phổ lớn các khả năng, những kỹ năng nền tảng này bao gồm lắng nghe, phát âm và viết tiếng mẹ đẻ hoặc ngoại ngữ, cũng như kỹ năng số học cơ bản, nghệ thuật và lập trình. 'Luyện tập và thực hành' không nên được coi là một phương pháp sư phạm lỗi thời; thay vào đó, nó nên được làm mới và nâng cấp với các công nghệ GenAI để thúc đẩy việc tự luyện tập các kỹ năng nền tảng của người học. Nếu được định hướng bởi các nguyên tắc đạo đức và sư phạm, các công cụ GenAI có tiềm năng trở thành huấn luyện viên 1:1 cho việc thực hành tự học như vậy, như được minh họa trong **Bảng 5**.

Bảng 5. Các trường hợp sử dụng GenAI đồng thiết kế như một huấn luyện viên 1:1 cho việc tự học các kỹ năng nền tảng trong ngôn ngữ và nghệ thuật

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Huấn luyện viên kỹ năng ngôn ngữ 1:1	Học ngôn ngữ, bao gồm thực hành hội thoại.	Thu hút người học vào thực hành hội thoại để giúp họ cải thiện kỹ năng nghe, nói và viết bằng cách đưa ra câu trả lời, sửa lỗi và mô hình hóa tiếng mẹ đẻ hoặc ngoại ngữ. Giúp người học cải thiện kỹ năng viết của họ. **Tiềm năng chuyển đổi: Kèm cặp ngôn ngữ 1:1 ở cấp độ người mới bắt đầu**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Có thể đặt ra giới hạn tuổi cho các cuộc trò chuyện độc lập do kết quả đầu ra của các hệ thống GenAI có thể không nhạy cảm về mặt văn hóa hoặc không phù hợp với độ tuổi. Người học phải có động lực nội tại ban đầu để tham gia vào cuộc trò chuyện với hệ thống AI. Người học nên có năng lực tiếp cận mang tính phản biện đối với các đề xuất của GenAI và kiểm tra xem chúng có chính xác hay không.	Khi sử dụng các nền tảng GenAI chung, nhà giáo con người có thể hướng dẫn người học tương tác với các công cụ GenAI để yêu cầu câu trả lời cải thiện, sửa lỗi phát âm hoặc ví dụ về viết. Ví dụ: *Tham gia với tôi trong một cuộc trò chuyện bằng ngôn ngữ [x], giúp tôi liên tục cải thiện.* *Đề xuất một số ý tưởng để giúp tôi viết về [chủ đề x].*	Cần cảnh giác với ngôn ngữ không nhạy cảm về văn hóa hoặc không chính xác về ngữ cảnh, và việc vô tình duy trì các định kiến hoặc thiên kiến văn hóa Nếu không có chiến lược sư phạm thích hợp để kích thích động lực nội tại của người học, nó có thể hạn chế óc sáng tạo và tính độc đáo của trẻ em, dẫn đến viết lách theo công thức. Nó cũng có thể hạn chế cơ hội tương tác trong cuộc sống thực, ý kiến đa dạng, biểu đạt đa dạng và tư duy phản biện
Huấn luyện viên nghệ thuật 1:1	Kỹ năng trong các lĩnh vực nghệ thuật như âm nhạc và vẽ.	Cung cấp gợi ý cho các kỹ thuật nhạc họa (ví dụ: mẹo về phối cảnh và màu sắc), hoặc sáng tác âm nhạc (ví dụ: giai điệu và tiến trình hợp âm). **Tiềm năng chuyển đổi: Giáo viên nghệ thuật 1:1 ở mức cơ bản**	Bắt đầu với danh sách trong Phần 1.2, đánh giá xem các công cụ GenAI có thể truy cập được tại địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người	Người học phải có một số mục tiêu ban đầu để tạo ra nghệ thuật hoặc âm nhạc, hiểu biết cơ bản về các yếu tố chính của lĩnh vực nghệ thuật hoặc âm nhạc, và năng lực cơ bản để phân tích các tác phẩm nghệ thuật hoặc sáng tác âm nhạc.	Giáo viên con người nên yêu cầu người học so sánh kỹ thuật nghệ thuật của công cụ AI với tác phẩm nghệ thuật của chính họ. Giáo viên hoặc huấn luyện viên con người phải khuyến khích người học phát triển và áp dụng trí tưởng tượng và sáng tạo của họ, điều mà GenAI không thể thay thế. Ví dụ về lời nhắc: *Đề xuất một số ý tưởng để truyền cảm hứng cho tôi tạo ra một hình ảnh về [chủ đề/ý tưởng].*	Có thể khiến trẻ em tiếp xúc với nội dung không phù hợp hoặc phản cảm, điều này có thể vi phạm quyền được bảo vệ và an toàn của các em. Các công cụ GenAI cản trở việc người học phát triển trí tưởng tượng và sáng tạo của họ.
Huấn luyện viên 1:1 cho lập trình hoặc toán	Kiến thức lập trình khái niệm và kỹ năng ở cấp độ giới thiệu. Nó cũng có thể áp dụng cho việc học toán cơ bản.	Hỗ trợ tự học kiến thức và kỹ năng lập trình cơ bản, tìm lỗi trong mã của người học và cung cấp câu trả lời ngay lập tức, và điều chỉnh câu trả lời cho các câu hỏi. **Tiềm năng chuyển đổi:** **Giáo viên lập trình 1:1 ở mức cơ bản**	Bắt đầu với danh sách trong Phần 1.3, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Tìm và xác định vấn đề, và thiết kế thuật toán để giải quyết vấn đề, vẫn là các khía cạnh cốt lõi của việc học lập trình và lập trình. Người học phải có động lực nội tại để sử dụng mã, cùng với một số kiến thức và kỹ năng cơ bản trong việc sử dụng ngôn ngữ lập trình.	Giáo viên và huấn luyện viên con người nên dạy kiến thức và kỹ năng cơ bản, và truyền cảm hứng cho người học sử dụng tư duy tính toán và lập trình để giải quyết vấn đề bao gồm cả thông qua lập trình cộng tác. Ví dụ về lời nhắc: *Đề xuất một số ý tưởng mới lạ thường cho việc lập trình.*	Độ chính xác của câu trả lời và đề xuất vẫn là một vấn đề có vấn đề vì GenAI không phải lúc nào cũng đúng. Có nguy cơ cao là các công cụ GenAI sẽ ngăn cản người học phát triển kỹ năng tư duy tính toán và khả năng tìm và xác định các vấn đề có ý nghĩa cho việc lập trình.

5.3.4 AI tạo sinh để thúc đẩy học tập dựa trên tra cứu hoặc dự án

Nếu không được sử dụng có mục đích để thúc đẩy tư duy bậc cao hoặc sáng tạo, các công cụ GenAI có xu hướng khuyến khích đạo văn hoặc kết quả đầu sơ sài kiểu 'vẹt ngẫu nhiên'. Tuy nhiên, do các mô hình GenAI đã được huấn luyện dựa trên dữ liệu quy mô lớn, chúng có tiềm năng đóng vai trò như một đối thủ trong các cuộc đối thoại Socrates hoặc như một trợ lý nghiên cứu trong học tập dựa trên dự án. Tuy nhiên, những tiềm năng này chỉ có thể được tận dụng thông qua các quy trình thiết kế hướng dẫn/học tập nhằm kích hoạt tư duy bậc cao như được minh họa trong **Bảng 6**.

Bảng 6. Các trường hợp sử dụng GenAI đồng thiết kế nhằm thúc đẩy học tập dựa trên tra cứu hoặc dự án

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Thách thức Socrates	III-Các vấn đề có cấu trúc lỏng lẻo.	Tham gia với người học trong cuộc đối thoại gợi nhớ về việc đặt câu hỏi Socrates về kiến thức trước đó, dẫn đến khám phá kiến thức mới hoặc hiểu biết sâu sắc hơn. **Tiềm năng chuyển đổi: Đối thủ** **Socrates 1:1**	Bắt đầu với danh sách trong Phần 1.3, đánh giá xem các công cụ GenAI cụ thể có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt và được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Người học phải đạt đến độ tuổi cho phép họ tiến hành các cuộc trò chuyện độc lập với các công cụ GenAI. Người học phải có kiến thức trước và khả năng kiểm tra xem các lập luận và thông tin được trình bày có chính xác hay không.	Giáo viên con người có thể giúp chuẩn bị một danh sách các câu hỏi ngày càng sâu sắc hơn làm ví dụ để người học điều chỉnh thành các câu lệnh. Người học cũng có thể bắt đầu với một câu lệnh rộng như '*Cùng tôi tham gia vào một cuộc đối thoại Socrates để giúp tôi có góc nhìn phản biện về [chủ đề x]*' và sau đó dần dần làm sâu sắc cuộc đối thoại thông qua các câu lệnh ngày càng tinh hơn.	Các công cụ GenAI hiện tại có thể tạo ra các câu trả lời tương tự hoặc tiêu chuẩn hạn chế việc người học tiếp xúc với các quan điểm đa dạng và góc nhìn thay thế, dẫn đến hiệu ứng phòng vang, và cản trở sự phát triển tư duy độc lập.
Cố vấn cho việc học theo dự án	III-Các vấn đề nghiên cứu có cấu trúc lỏng lẻo trong khoa học tự nhiên hoặc xã hội.	Hỗ trợ tạo ra kiến thức thông qua việc giúp người học tiến hành học tập dựa trên dự án. Điều này bao gồm GenAI đóng vai trò tương tự như cố vấn nghiên cứu được mô tả trong Bảng 3. **Tiềm năng chuyển đổi:** **Huấn luyện viên học theo dự án 1:1**	Bắt đầu với danh sách trong Phần 1.3, đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được kiểm định nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Người học có thể đóng vai trò như các nhà nghiên cứu trẻ trong việc lập kế hoạch và triển khai học tập dựa trên dự án. Người học phải đủ tuổi để sử dụng độc lập các nền tảng GenAI. Người học phải có động lực và khả năng tham gia vào các hoạt động học tập dựa trên dự án tự định hướng, để họ không bị cám dỗ thụ động sao chép và dán các câu trả lời do GenAI cung cấp.	Giáo viên con người hướng dẫn người học yêu cầu GenAI cung cấp ý tưởng cơ bản để xác định vấn đề nghiên cứu như đã đề xuất trong 5.3.1. Cá nhân và nhóm người học sử dụng các công cụ GenAI để tiến hành đánh giá tài liệu, thu thập và xử lý dữ liệu, và tạo báo cáo.	Người học không có nền tảng kiến thức vững chắc và khả năng cần thiết để xác minh độ chính xác của câu trả lời có thể bị đánh lừa bởi thông tin mà các công cụ GenAI cung cấp. Nó cũng có thể hạn chế các cuộc thảo luận và tương tác của người học với bạn bè và giảm cơ hội học tập hợp tác, có khả năng gây cản trở sự phát triển xã hội của họ.

5.3.5 AI tạo sinh để hỗ trợ người học có nhu cầu đặc biệt

Về mặt lý thuyết, các mô hình GenAI có tiềm năng giúp người học khiếm thính hoặc khiếm thị. Các thực hành mới nổi bao gồm phụ đề hoặc chú thích được hỗ trợ bởi GenAI cho người học bị khiếm thính hoặc khó nghe, và mô tả âm thanh do GenAI tạo ra cho người học khiếm thị. Các mô hình GenAI cũng có thể chuyển đổi văn bản thành giọng nói và giọng nói thành văn bản để cho phép người khiếm thị, khiếm thính hoặc khiếm khuyết ngôn ngữ tiếp cận nội dung, đặt câu hỏi và giao tiếp với bạn bè của họ. Tuy nhiên, chức năng này vẫn chưa được tận dụng ở quy mô lớn. Theo cuộc khảo sát được đề cập trước đây, được UNESCO triển khai vào năm 2023 về việc sử dụng AI trong giáo dục của chính phủ, chỉ có 4 quốc gia (Trung Quốc, Gioóc-đa-ni, Ma-lai-xi-a và Ca-ta) báo cáo rằng các cơ quan chính phủ của họ đã kiểm nghiệm và khuyến nghị các công cụ hỗ trợ AI để hỗ trợ tiếp cận hòa nhập cho người học khuyết tật (UNESCO, 2023*c*).

Cũng có xu hướng hướng tới các lần lặp của các mô hình GenAI được huấn luyện để hỗ trợ người học sử dụng ngôn ngữ của riêng họ, bao gồm các ngôn ngữ thiểu số và bản địa, để học tập và giao tiếp. Chẳng hạn, PaLM 2, mô hình ngôn ngữ lớn thế hệ tiếp theo của Google, được huấn luyện trên dữ liệu song song bao gồm hàng trăm ngôn ngữ dưới dạng các cặp văn bản nguồn và đích. Việc đưa vào dữ liệu đa ngôn ngữ song song được thiết kế để cải thiện hơn nữa khả năng của mô hình trong việc hiểu và tạo ra văn bản đa ngôn ngữ (Google, 2023*b*).

Bằng cách cung cấp bản dịch, diễn giải và sửa lỗi tự động theo thời gian thực, các công cụ GenAI có tiềm năng giúp người học sử dụng ngôn ngữ thiểu số truyền đạt ý tưởng và tăng cường sự hợp tác với bạn bè từ các nền tảng ngôn ngữ khác nhau. Tuy nhiên, điều này sẽ không tự nhiên xảy ra ở quy mô lớn. Chỉ với thiết kế có mục đích, tiềm năng này mới có thể được tận dụng để khuếch đại tiếng nói của các nhóm yếu thế.

Cuối cùng, cũng có ý kiến đề xuất rằng các hệ thống GenAI có tiềm năng triển khai chẩn đoán dựa trên cuộc trò chuyện, xác định các vấn đề tâm lý hoặc cảm xúc-xã hội cũng như khó khăn trong học tập. Tuy nhiên, vẫn còn ít bằng chứng cho thấy cách tiếp cận này hoặc là hiệu quả hoặc an toàn, và bất kỳ chẩn đoán nào cũng sẽ đòi hỏi sự diễn giải của các chuyên gia có chuyên môn.

Bảng 7. Các trường hợp sử dụng GenAI đồng thiết kế nhằm hỗ trợ người học có nhu cầu đặc biệt

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Chẩn đoán qua hội thoại về các khó khăn trong học tập	Điều này có thể hữu ích cho người học đang phải đối mặt với khó khăn học tập do các vấn đề tâm lý, xã hội hoặc cảm xúc gây ra.	Sử dụng tương tác ngôn ngữ tự nhiên để xác định nhu cầu của người học có vấn đề tâm lý, xã hội hoặc cảm xúc hoặc khó khăn trong học tập, để cung cấp cho họ hỗ trợ hoặc hướng dẫn liên quan. **Tiềm năng chuyển đổi:** **Cố vấn chính 1:1 cho người học có vấn đề xã hội hoặc cảm xúc hoặc khó khăn trong học tập**	Ngoài các công cụ GenAI chung, tìm kiếm các chatbot được hỗ trợ bởi GenAI. Đánh giá xem chúng có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Các nhà giáo hoặc chuyên gia làm việc với nhóm người học này sẽ cần đảm bảo rằng lời khuyên chính được đề xuất bởi hệ thống GenAI là chính xác.	Các nhà giáo hoặc người hỗ trợ cần cung cấp môi trường thoải mái để thu hút người học vào cuộc trò chuyện nhằm chẩn đoán các vấn đề tâm lý, xã hội hoặc cảm xúc, hoặc khó khăn trong học tập.	Có thể vô tình chẩn đoán sai thách thức cụ thể của người học, dẫn đến việc cung cấp hỗ trợ sai.

Các ứng dụng tiềm năng nhưng chưa được chứng minh	Lĩnh vực kiến thức hoặc vấn đề phù hợp	Các kết quả mong đợi	Các công cụ GenAI phù hợp và lợi thế so sánh	Yêu cầu đối với người dùng	Yêu cầu về phương pháp sư phạm của con người và ví dụ về câu lệnh	Các nguy cơ tiềm tàng
Công cụ tiếp cận dựa trên AI	Những công cụ này cho phép người học khiếm thính hoặc khiếm thị tiếp cận nhiều nội dung hơn, từ đó cải thiện chất lượng học tập của họ.	Đáp ứng nhu cầu tiếp cận của người học và hỗ trợ việc tiếp thu kiến thức môn học cụ thể bằng cách cung cấp phụ đề và/hoặc thông dịch ngôn ngữ ký hiệu được hỗ trợ bởi GenAI cho nội dung âm thanh hoặc video, và mô tả âm thanh cho văn bản hoặc tài liệu trực quan khác. **Tiềm năng chuyển đổi:** **Công cụ hỗ trợ ngôn ngữ được hỗ trợ bởi AI được cá nhân hóa 1:1**	Ngoài các công cụ GenAI chung, tìm kiếm các trình tạo phụ đề và mô tả âm thanh được hỗ trợ bởi AI liên quan và đáng tin cậy. Đánh giá xem chúng có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Các nhà giáo dục hoặc người hỗ trợ phải giúp người học truy cập và học cách vận hành các công cụ GenAI. Họ cũng cần đảm bảo rằng kết quả đầu ra của các công cụ thực sự hỗ trợ những người học này và không củng cố những thách thức và định kiến mà họ phải đối mặt.	Cần kiểm tra khả năng tiếp cận của các nền tảng hoặc công cụ để xác định và khắc phục các vấn đề về khả năng tiếp cận trước khi chúng được sử dụng. Các công cụ GenAI chỉ có thể cung cấp quyền truy cập vào nội dung, vì vậy các nhà giáo dục và người hỗ trợ nên tập trung vào việc nâng cao chất lượng học tập và phúc lợi xã hội của họ. Các nhà giáo dục và người hỗ trợ cần dạy người học tạo câu lệnh bằng giọng nói hoặc văn bản dựa trên khả năng của họ.	Phụ đề hoặc mô tả âm thanh được tạo ra bởi các nền tảng GenAI không được thiết kế đặc biệt để hỗ trợ thị giác hoặc thính giác thường không chính xác và có thể gây hiểu lầm cho người học có nhu cầu đặc biệt. Các công cụ này có thể vô tình củng cố các thiên kiến hiện có.
Bộ khuếch đại tạo sinh cho người học yếu thế	Nó có thể hữu ích cho người học từ nền tảng ngôn ngữ hoặc văn hóa thiểu số để thể hiện và khuếch đại tiếng nói của họ, tham gia trực tuyến, và tiến hành các nghiên cứu xã hội hợp tác.	Cung cấp bản dịch thời gian thực, diễn giải, và sửa lỗi tự động trong viết để hỗ trợ người học từ các nhóm yếu thế sử dụng ngôn ngữ của riêng họ để giao tiếp với bạn bè từ các nền tảng ngôn ngữ khác nhau. **Tiềm năng chuyển đổi:** **LLM hòa nhập cho người học thuộc nhóm yếu thế**	Một ví dụ cụ thể để xem xét là PaLM 2. Đánh giá xem các công cụ GenAI có thể truy cập được ở địa phương, mã nguồn mở, được thử nghiệm nghiêm ngặt hoặc được kiểm chứng bởi các cơ quan có thẩm quyền hay không. Tiếp tục xem xét các ưu điểm và thách thức của bất kỳ công cụ GenAI cụ thể nào, và đảm bảo rằng nó giải quyết đúng nhu cầu cụ thể của con người.	Người học nên có kiến thức hoặc ý kiến có ý nghĩa về chủ đề của cuộc trò chuyện hoặc nghiên cứu hợp tác. Họ cần có khả năng đóng góp có trách nhiệm và không phân biệt đối xử và tránh phát ngôn thù địch.	Giáo viên hoặc nhà giáo dục nên thiết kế các nghiên cứu và nhiệm vụ viết cho người học về các chủ đề xã hội hoặc văn hóa, hoặc tổ chức các hội thảo trực tuyến hoặc hợp tác liên văn hóa để kích thích người học tạo ra ý tưởng và chia sẻ ý kiến.	Cần xác định và sửa chữa các lỗi trong bản dịch và diễn giải của AI có thể gây ra hiểu lầm giữa các nền văn hóa. Việc sử dụng này có thể cung cấp cơ hội cho người học thuộc nhóm yếu thế để khuếch đại tiếng nói của họ, nhưng sẽ không giải quyết nguyên nhân gốc rễ của sự nghèo nàn về dữ liệu và do đó không thể phi thực dân hóa các công cụ AI.

6. GenAI và tương lai của giáo dục và nghiên cứu

Các công nghệ GenAI vẫn đang phát triển nhanh chóng và có khả năng có tác động sâu sắc đến giáo dục và nghiên cứu, và vẫn chưa được hiểu đầy đủ. Do đó, những hàm ý tiềm tàng, lâu dài của chúng đối với giáo dục và nghiên cứu cần được quan tâm ngay tức thì và đánh giá sâu hơn.

6.1 Các vấn đề đạo đức chưa được khai phá

Các công cụ GenAI ngày càng tinh vi sẽ làm dấy lên thêm những lo ngại về đạo đức cần được xem xét chi tiết. Tiếp theo Phần 2 và 3, cần có những phân tích sâu hơn và hướng tới tương lai hơn để tiết lộ và giải quyết các vấn đề đạo đức chưa được khám phá từ ít nhất 5 góc độ sau:

- **Tiếp cận và công bằng**: Các hệ thống GenAI trong giáo dục có thể làm trầm trọng thêm sự chênh lệch hiện có trong việc tiếp cận công nghệ và tài nguyên giáo dục, làm sâu sắc thêm bất bình đẳng.
- **Kết nối con người**: Các hệ thống GenAI trong giáo dục có thể giảm tương tác giữa người với người và các khía cạnh cảm xúc-xã hội quan trọng của việc học tập.
- **Phát triển trí tuệ con người**: Các hệ thống GenAI trong giáo dục có thể hạn chế quyền tự chủ và quyền tự quyết của người học bằng cách cung cấp các giải pháp định sẵn hoặc thu hẹp phạm vi trải nghiệm học tập. Tác động lâu dài của chúng đối với sự phát triển trí tuệ của người học trẻ cần được điều tra thêm.
- **Tác động tâm lý**: Các hệ thống GenAI mô phỏng tương tác của con người có thể có những tác động tâm lý chưa được biết đến đến đối với người học, gây ra lo ngại về sự phát triển nhận thức và sức khỏe cảm xúc của họ, và về khả năng bị chỉnh sửa.
- **Thiên kiến và phân biệt đối xử ngầm**: Khi các hệ thống GenAI tinh vi hơn đang được phát triển và áp dụng trong giáo dục, chúng có khả năng tạo ra những thiên kiến và hình thức phân biệt đối xử mới dựa trên dữ liệu huấn luyện và phương pháp được sử dụng bởi các mô hình, điều này có thể dẫn đến những kết quả không biết và có khả năng gây hại.

6.2 Bản quyền và sở hữu trí tuệ

Sự xuất hiện của GenAI đang nhanh chóng thay đổi cách thức các tác phẩm khoa học, nghệ thuật và văn học được tạo ra, phân phối và tiêu thụ. Việc sao chép, phân phối hoặc sử dụng trái phép các tác phẩm có bản quyền mà không được sự cho phép của chủ sở hữu bản quyền vi phạm quyền độc quyền của họ và có thể dẫn đến hậu quả pháp lý. Chẳng hạn, việc huấn luyện các mô hình GenAI đã bị cáo buộc vi phạm bản quyền. Trong một trong những trường hợp gần đây, bài hát do AI tạo ra có đặc điểm của 'Drake' và 'The Weeknd' (Abel Tesfaye) đã tiếp cận được hàng triệu người nghe trước khi bị gỡ bỏ do tranh chấp bản quyền (Coscarelli, 2023). Mặc dù các khuôn khổ quy định mới nổi có ý định yêu cầu các nhà cung cấp GenAI công nhận và bảo vệ quyền sở hữu trí tuệ của chủ sở hữu nội dung được sử dụng bởi mô hình, việc xác định quyền sở hữu và tính nguyên bản của lượng tác phẩm được tạo ra quá lớn ngày càng trở nên khó khăn. Việc thiếu khả năng truy lại nguồn gốc này không chỉ gây ra lo ngại về việc bảo vệ quyền của người sáng tạo và đảm bảo bồi thường công bằng cho đóng góp trí tuệ của họ, mà còn đưa ra những thách thức trong bối cảnh giáo dục về cách sử dụng kết quả đầu ra của các công cụ GenAI một cách có trách nhiệm. Điều này có thể có những hệ quả sâu sắc đối với hệ thống nghiên cứu.

6.3 Nguồn nội dung và học tập

Các công cụ GenAI đang thay đổi cách mà nội dung giảng dạy và học tập có thể được sản xuất và truyền tải. Trong tương lai, nội dung được tạo ra thông qua các cuộc trò chuyện giữa người và AI có thể trở thành một trong những nguồn sản xuất kiến thức chính. Điều này có khả năng làm yếu đi sự tham gia trực tiếp của người học với nội dung giáo dục dựa trên các tài nguyên, sách giáo khoa và chương trình giảng dạy được tạo ra và kiểm chứng bởi con người. Vẻ bề ngoài uy tín của văn bản GenAI có thể đánh lừa những người học trẻ tuổi không có đủ kiến thức trước đó để có thể nhận ra những điểm không chính xác hoặc để đặt câu hỏi về nó một cách hiệu quả. Liệu việc tham gia của người học với nội dung chưa được kiểm chứng có nên được công nhận là 'học tập' hay không cũng là điều còn phải tranh luận.

Hệ quả là sự tập trung vào thông tin tổng hợp từ nguồn thứ cấp, vốn có thể làm mất đi cơ hội tự xây dựng kiến thức thông qua các phương pháp đã được chứng thực như trực tiếp nhận thức và trải nghiệm thế giới thực, học hỏi từ thử nghiệm và sai lầm, triển khai các thí nghiệm thực nghiệm, và phát triển ý thức thông thường của người học. Nó cũng có thể đe dọa việc xây dựng kiến thức xã hội và nuôi dưỡng các giá trị xã hội thông qua các hoạt động hợp tác trong lớp học

6.4 Câu trả lời đồng nhất so với kết quả đầu ra đa dạng và sáng tạo

GenAI thu hẹp các tường thuật đa dạng vì các kết quả đầu ra được tạo ra có xu hướng đại diện và củng cố các quan điểm chiếm ưu thế. Sự đồng nhất hóa kiến thức kết quả hạn chế tư duy đa dạng và sáng tạo. Sự phụ thuộc ngày càng tăng của nhà giáo và người học vào các công cụ GenAI để tìm kiếm gợi ý có thể dẫn đến sự tiêu chuẩn hóa và đồng nhất của các câu trả lời, làm suy yếu giá trị của tư duy độc lập và tự định hướng. Khả năng đồng nhất hóa biểu đạt trong các tác phẩm viết và nghệ thuật có thể hạn chế trí tưởng tượng, sáng tạo và góc nhìn thay thế của người học.

Các nhà cung cấp GenAI và nhà giáo dục cần xem xét mức độ mà EdGPT có thể được phát triển và sử dụng để thúc đẩy sáng tạo, hợp tác, tư duy phản biện và các kỹ năng tư duy bậc cao khác.

6.5 Tư duy lại về đánh giá và chuẩn đầu ra

Những tác động của GenAI đối với đánh giá vượt ra ngoài những lo ngại tức thời về việc người học gian lận trong các bài tiểu luận. Chúng ta phải đối mặt với thực tế là GenAI có thể tạo ra các bài báo và bài luận được tổ chức tương đối tốt và các tác phẩm nghệ thuật ấn tượng, và có thể vượt qua một số bài kiểm tra dựa trên kiến thức trong một số lĩnh vực môn học nhất định. Do đó, chúng ta cần phải suy nghĩ lại về việc chính xác nên học cái gì và với mục đích gì, và làm thế nào để đánh giá và xác nhận việc học tập.

Thảo luận mang tính phản biện bởi các nhà giáo dục, nhà hoạch định chính sách, người học và các bên liên quan khác cần xem xét 4 loại chuẩn đầu ra sau:

Giá trị: Các giá trị cần thiết để đảm bảo thiết kế và sử dụng công nghệ lấy con người làm trung tâm là trọng tâm của việc suy nghĩ lại về chuẩn đầu ra và đánh giá chúng trong kỷ nguyên số. Trong việc xem xét lại mục đích của giáo dục, các giá trị thông tin cho cách công nghệ liên quan đến giáo dục nên được nêu rõ. Chính thông qua lăng kính quy chuẩn này mà chuẩn đầu ra và việc đánh giá và xác nhận chúng cần được cập nhật liên tục để đáp ứng với việc sử dụng ngày càng phổ biến của công nghệ, bao gồm cả AI, trong xã hội.

Kiến thức và kỹ năng nền tảng: Ngay cả trong các lĩnh vực mà các công cụ GenAI có thể làm tốt hơn con người, người học vẫn sẽ cần kiến thức và kỹ năng nền tảng vững chắc. Các kỹ năng đọc viết, tính toán và hiểu biết khoa học cơ bản vẫn sẽ là chìa khóa cho giáo dục trong tương lai. Phạm vi và bản chất của các kỹ năng nền tảng này sẽ cần được xem xét lại thường xuyên để phản ánh môi trường ngày càng giàu AI mà chúng ta đang sống.

Kỹ năng tư duy bậc cao: Chuẩn đầu ra sẽ cần bao gồm các kỹ năng cần thiết để hỗ trợ tư duy bậc cao và giải quyết vấn đề dựa trên sự hợp tác giữa người và AI và việc sử dụng kết quả đầu ra do GenAI tạo ra. Những kỹ năng này có thể bao gồm hiểu biết về vai trò của kiến thức thực tế và khái niệm trong việc đặt nền tảng cho tư duy bậc cao, và đánh giá mang tính phản biện nội dung do AI tạo ra.

Kỹ năng nghề nghiệp cần thiết để làm việc với AI: Trong các lĩnh vực mà AI có thể làm tốt hơn con người và đang tự động hóa các đơn vị nhiệm vụ, người học cần nuôi dưỡng các kỹ năng mới cho phép họ phát triển, vận hành và làm việc với các công cụ GenAI. Việc thiết kế lại chuẩn đầu ra và đánh giá giáo dục sẽ cần phản ánh các kỹ năng nghề nghiệp cần thiết cho các công việc mới được tạo ra bởi AI.

6.6 Quy trình tư duy

Góc độ cơ bản nhất của những tác động lâu dài của GenAI đối với giáo dục và nghiên cứu vẫn là về mối quan hệ bổ sung giữa quyền tự chủ của con người và máy móc. Một trong những câu hỏi chính là liệu con người có thể nhường các cấp độ cơ bản của quá trình tư duy và tiếp thu kỹ năng cho AI và thay vào đó tập

trung vào các kỹ năng tư duy bậc cao dựa trên kết quả đầu ra do AI cung cấp hay không.

Viết lách thường được liên kết với việc cấu trúc tư duy. Với GenAI, thay vì bắt đầu từ đầu để lập kế hoạch cho mục tiêu, phạm vi và đề cương của một tập hợp ý tưởng, con người giờ đây có thể bắt đầu với một đề cương được cấu trúc tốt do GenAI cung cấp. Một số chuyên gia đã mô tả việc sử dụng GenAI để tạo văn bản theo cách này là 'viết mà không cần suy nghĩ' (Chayka, 2023). Khi những thực hành mới được hỗ trợ bởi GenAI này được áp dụng rộng rãi hơn, các phương pháp đã được thiết lập để tiếp thu và đánh giá kỹ năng viết sẽ cần phải thích ứng. Một lựa chọn trong tương lai là việc học viết có thể tập trung vào việc xây dựng kỹ năng lập kế hoạch và soạn thảo lời nhắc, đánh giá kĩ lưỡng kết quả đầu ra của GenAI, và tư duy bậc cao, cũng như về viết chung dựa trên đề cương của GenAI.

Kết luận

Từ góc độ của cách tiếp cận lấy con người làm trung tâm, các công cụ AI nên được thiết kế để mở rộng hoặc tăng cường năng lực trí tuệ và kỹ năng xã hội của con người, chứ không phải làm suy yếu, xung đột với chúng hoặc thay thế chúng. Từ lâu người ta đã kỳ vọng rằng các công cụ AI có thể được tích hợp thêm như một phần không thể thiếu trong các công cụ có sẵn cho con người để hỗ trợ phân tích và hành động hướng tới tương lai hòa nhập và bền vững hơn.

Để AI trở thành một phần đáng tin cậy của sự hợp tác giữa người và máy - ở cấp độ cá nhân, tổ chức và hệ thống - cách tiếp cận lấy con người làm trung tâm dựa trên *Khuyến nghị về Đạo đức AI năm 2021* của UNESCO cần được cụ thể hóa và triển khai thêm theo các đặc điểm cụ thể của các công nghệ mới nổi như GenAI. Chỉ thông qua cách này chúng ta mới có thể đảm bảo rằng GenAI trở thành một công cụ đáng tin cậy cho các nhà nghiên cứu, nhà giáo và người học.

Mặc dù GenAI nên được sử dụng để phục vụ giáo dục và nghiên cứu, tất cả chúng ta cần nhận thức rằng GenAI cũng có thể thay đổi các hệ thống đã được thiết lập và nền tảng của chúng trong các lĩnh vực này. Sự chuyển đổi của giáo dục và nghiên cứu được kích hoạt bởi GenAI, nếu có, nên được xem xét kỹ lưỡng và định hướng bởi cách tiếp cận lấy con người làm trung tâm. Chỉ làm như vậy chúng ta mới có thể đảm bảo rằng tiềm năng của AI nói riêng, và của tất cả các loại công nghệ khác được sử dụng trong giáo dục nói chung, giúp nâng cao năng lực của con người và xây dựng một tương lai số hòa nhập cho mọi người.

Tài liệu tham khảo

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 (Truy cập ngày 23 tháng 6 năm 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-altman-urges-congress-to-regulate-powerful-new-ai-technology (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 23 tháng 6 năm 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-ai-promises-chatgpt-earnings (Truy cập ngày 3 tháng 7 năm 2023.)

Chayka, K. 2023. *My A.I. Writing Report*. New York, The New Yorker. Available at: https://www.newyorker.com/culture/infinite-scroll/my-ai-writing-robot (Truy cập ngày 1 tháng 8 năm 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 (Truy cập ngày 31 tháng 7 năm 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-weeknd-fake.html (Truy cập ngày 30 tháng 8 năm 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 (Truy cập ngày 19 tháng 7 năm 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 (Truy cập ngày 19 tháng 7 năm 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 (Truy cập ngày 25 tháng 8 năm 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-about-openais-new-ai-model-a5977b42e495 (Truy cập ngày 1 tháng 8 năm 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 (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 23 tháng 6 năm 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/childrens-online-privacy-protection-rule-coppa (Truy cập ngày 4 tháng 9 năm 2023.)

Giannini, S. 2023. *Generative AI and the Future of Education*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000385877 (Truy cập ngày 29 tháng 8 năm 2023.)

Google. 2023*a*. *Recommendations for Regulating AI*. Mountain View, Google. Available at: https://ai.google/static/documents/recommendations-for-regulating-ai.pdf (Truy cập ngày 23 tháng 6 năm 2023.)

——. 2023*b*. *PaLM 2 Technical Report*. Mountain View, Google. Available at: https://doi.org/10.48550/arXiv.2305.10403 (Truy cập ngày 20 tháng 7 năm 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-than-companies-can-keep-up-a2bdedd4 (Truy cập ngày 25 tháng 8 năm 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 (Truy cập ngày 23 tháng 6 năm 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-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems (Truy cập ngày 31 tháng 7 năm 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-intelligence-google-bias.html (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 25 tháng 8 năm 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 AI-powered 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 (Truy cập ngày 1 tháng 8 năm 2023.)

Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Singapore, OpenGov Asia. Available at: https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster (Truy cập ngày 25 tháng 8 năm 2023.)

OpenAI. 2018. *AI and compute*. San Francisco, OpenAI. Available at: https://openai.com/research/ai-and-compute (Truy cập ngày 23 tháng 6 năm 2023.)

———. 2023. *Educator considerations for ChatGPT*. San Francisco, OpenAI. Available at: https://platform.openai.com/docs/chatgpt-education (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 23 tháng 6 năm 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-artificial-intelligence-artists.html (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 25 tháng 8 năm 2023.)

Stanford University. 2019. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: https://hai.stanford.edu/ai-index-2019 (Truy cập ngày 23 tháng 6 năm 2023.)

———. 2023. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: https://hai.stanford.edu/research/ai-index-2023 (Truy cập ngày 23 tháng 6 năm 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-sutskever-interview (Truy cập ngày 1 tháng 8 năm 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/openai-gpt-4-rumor-release-date-sam-altman-interview (Truy cập ngày 1 tháng 8 năm 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 (Truy cập ngày 23 tháng 6 năm 2023.)

UNESCO. 2019. *Beijing Consensus on Artificial Intelligence and Education*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000368303 (Truy cập ngày 3 tháng 7 năm 2023.)

———. 2022*a*. *Recommendation on the Ethics of Artificial Intelligence*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (Truy cập ngày 3 tháng 7 năm 2023.)

———. 2022*b*. *AI and education: guidance for policy-makers*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000376709 (Truy cập ngày 23 tháng 6 năm 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 (Truy cập ngày 20 tháng 7 năm 2023.)

———. 2022*d*. *Guidelines for ICT in education policies and masterplans*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380926 (Truy cập ngày 31 tháng 7 năm 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-governments-implement-global-ethical-framework-without (Truy cập ngày 3 tháng 7 năm 2023.)

——. 2023*b*. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. Không xuất bản.

——. 2023*c*. *Survey for the governmental use of AI as a public good for education*. Không xuất bản (Đệ trình lên 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 (Truy cập ngày 25 tháng 8 năm 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 (Truy cập ngày 25 tháng 8 năm 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 (Truy cập ngày 3 tháng 7 năm 2023.)

Chú giải

1 Xem https://chat.openai.com
2 Để đọc thêm giải thích về các kỹ thuật và công nghệ AI và mối quan hệ của chúng, xem UNESCO, 2022b, tr.8-10.
4 Lưu ý rằng, vì GenAI vẫn còn tương đối mới, các công ty khác nhau thường sử dụng các thuật ngữ này theo những cách khác nhau, và đôi khi sử dụng các từ khác nhau để chỉ cùng một thứ.
5 Lưu ý OpenAI, công ty đã phát triển các GPT trong bảng này, chưa công bố thông tin chi tiết về GPT-4 (The Verge, 2023a). Thực tế, số lượng tham số đã bị CEO của OpenAI bác bỏ (The Verge, 2023b). Tuy nhiên, các con số được đưa vào đây đã được báo cáo bởi một số nguồn tin (ví dụ, xem E2Analyst, 2023). Trong mọi trường hợp, điểm chính cần lưu ý là GPT-4 được xây dựng trên một bộ dữ liệu lớn hơn nhiều và sử dụng số lượng tham số lớn hơn nhiều so với GPT-3.
6 Lưu ý OpenAI, công ty đã phát triển các GPT trong bảng này, chưa công bố thông tin chi tiết về GPT-4 (The Verge, 2023a). Thực tế, số lượng tham số đã bị CEO của OpenAI bác bỏ (The Verge, 2023b). Tuy nhiên, các con số được đưa vào đây đã được báo cáo bởi một số nguồn tin (ví dụ, xem E2Analyst, 2023). Trong mọi trường hợp, điểm chính cần lưu ý là GPT-4 được xây dựng trên một bộ dữ liệu lớn hơn nhiều và sử dụng số lượng tham số lớn hơn nhiều so với GPT-3.
7 Xem https://crfm.stanford.edu/2023/03/13/alpaca.html
8 Xem https://bard.google.com
9 Xem https://writesonic.com/chat
10 Xem https://yiyan.baidu.com/welcome
11 Xem https://huggingface.co/chat
12 Xem https://www.jasper.ai
13 Xem https://ai.facebook.com/blog/large-language-model-llama-meta-ai
14 Xem https://open-assistant.io
15 Xem https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users
16 Xem https://you.com
17 Xem https://www.chatpdf.com
18 Xem https://elicit.org
19 Xem https://www.perplexity.ai
20 Xem https://tools.zmo.ai/webchatgpt
21 Xem https://www.compose.ai
22 Xem https://www.teamsmart.ai
23 Xem https://wiseone.io
24 Xem https://www.microsoft.com/en-us/bing
25 Xem https://www.craiyon.com
26 Xem https://openai.com/product/dall-e-2
27 Xem https://dream.ai/create
28 Xem https://www.fotor.com/features/ai-image-generator
29 Xem https://www.midjourney.com
30 Xem https://creator.nightcafe.studio
31 Xem https://writesonic.com/photosonic-ai-art-generator

32 Xem https://elai.io
33 Xem https://www.gliacloud.com
34 Xem https://pictory.ai
35 Xem https://runwayml.com
36 Xem https://www.aiva.ai
37 Xem https://boomy.com
38 Xem https://soundraw.io
39 Xem https://www.voicemod.net/text-to-song
40 Xem https://openai.com/research/gpt-4
41 Xem https://www.educhat.top and https://www.mathgpt.com
42 Xem https://www.educhat.top
43 Xem https://www.mathgpt.com
44 Có một số trường hợp ngoại lệ, chẳng hạn như Hugging Face, một nhóm chuyên phát triển AI mã nguồn mở
45 Xem, đơn cử như lời kêu gọi từ Google (2023a) và OpenAI (Bass và Metz, 2023).
46 Đối với một dự án nhằm quy định AI, xem dự thảo Đạo luật AI của Ủy ban Châu Âu (2021).
47 Đánh giá này dựa trên dữ liệu thu thập từ một cuộc khảo sát của UNESCO được phân phối cho 193 Quốc gia Thành viên về việc sử dụng AI của chính phủ trong giáo dục (UNESCO, 2023c), Đài quan sát Chính sách AI của OECD, và Báo cáo Chỉ số AI của Đại học Stanford (Stanford University, 2023), và thông tin trực tiếp thu thập từ một nhóm chuyên gia quốc tế.
48 Xem https://unctad.org/page/data-protection-and-privacy-legislation-worldwide
49 Dựa trên kết quả sơ đồ hóa, tính đến tháng 4 năm 2023, các quốc gia sau đây đã công bố chiến lược quốc gia về AI: Ác-hen-ti-na, Ô-xtrây-li-a, Áo, Bỉ, Bê-nanh, Bra-xin, Ca-na-đa, Bun-ga-ri, Chi-lê, Trung Quốc, Cô-lôm-bi-a, Síp, Cộng hòa Séc, Đan Mạch, Ai Cập, E-xtô-ni-a, Phần Lan, Pháp, Đức, Hung-ga-ri, Ai-xơ-len, Ấn Độ, In-đô-nê-xi-a , Ai-len, Ý, Nhật Bản, Gioóc-đa-ni, Lát-vi-a, Lít-va, Lúc-xăm-bua, Ma-lai-xi-a, Man-ta, Mô-ri-xơ, Mê-hi-cô, Hà Lan (Vương quốc), Na Uy, Niu Di-lân, Ô-man, Peru, Ba Lan, Bồ Đào Nha, Phi-líp-pin, Ca-ta, Hàn Quốc, Ru-ma-ni, Liên bang Nga, A-rập Xê-út, Xéc-bi-a, Xinh-ga-po, Xlô-ven-nia, Tây Ban Nha, Thụy Điển, Thái Lan, Thổ Nhĩ Kỳ, Tuy-ni-di, Các Tiểu Vương quốc Ả Rập Thống nhất, Vương quốc Anh, Hoa Kỳ, U-ru-goay và Việt Nam. Ngoài ra, một số quốc gia đã tích hợp chiến lược AI trong các chiến lược CNTT hoặc số rộng hơn, bao gồm An-giê-ri, Bốt-xoa-na, Ca-dắc-xtan, Kê-ni-a, Xi-ê-ra Lê-ôn, Xlô-va-ki-a, Thụy Sĩ và U-gan-đa.
50 Theo đánh giá nhanh về tất cả các chiến lược AI quốc gia (UNESCO, 2023b), hơn 40 chiến lược có các phần riêng về vấn đề đạo đức.
51 Theo đánh giá nhanh về tất cả các chiến lược AI quốc gia (UNESCO, 2023b), khoảng 45 chiến lược có các phần riêng về vấn đề giáo dục.
52 Xem https://openai.com/policies/terms-of-use
53 Ở một số quốc gia, giáo viên sẽ có trợ giảng (TA) với nhiệm vụ trả lời các câu hỏi của từng học sinh về nội dung khóa học. GenAI có thể được sử dụng để phát triển một phiên bản kỹ thuật số của trợ giảng, nhằm hỗ trợ học sinh và các giáo viên khác, nhưng cũng có thể gây ra một số vấn đề tiêu cực (ví dụ như ảnh hưởng đến các mối quan hệ xã hội trong lớp học).

unesco
Tổ chức Giáo dục,
Khoa học và Văn hóa
của Liên Hợp Quốc

Hướng dẫn về AI tạo sinh trong giáo dục và nghiên cứu

Hướng dẫn này nhằm mục đích hỗ trợ việc lên kế hoạch xây dựng các quy định, chính sách và chương trình phát triển năng lực con người phù hợp để đảm bảo rằng AI tạo sinh (GenAI) trở thành một công cụ thực sự mang lại lợi ích và giúp trao quyền cho nhà giáo, người học và nhà nghiên cứu. Nó giải thích các kỹ thuật AI được sử dụng bởi GenAI và sơ đồ hóa các mô hình GPT được công bố công khai, đặc biệt là những mô hình hoạt động với giấy phép mã nguồn mở. Nó cũng mở ra cuộc thảo luận về sự xuất hiện của EdGPT - các mô hình GenAI được huấn luyện với dữ liệu đặc thù để phục vụ mục đích giáo dục. Hơn nữa, nó tóm tắt một số tranh cãi chính xung quanh GenAI, từ việc làm trầm trọng thêm tình trạng nghèo kỹ thuật số đến việc đồng nhất hóa ý kiến, và từ deepfake thật hơn đến các vấn đề bản quyền. Với một tầm nhìn nhân văn, Hướng dẫn đề xuất các bước chính để quản trị các công cụ GenAI, bao gồm việc đặt ra yêu cầu bắt buộc đối với bảo vệ quyền riêng tư dữ liệu và đặt ra giới hạn độ tuổi cho các cuộc hội thoại độc lập với các nền tảng GenAI. Để định hướng việc sử dụng đúng đắn các công cụ trong giáo dục và nghiên cứu, Hướng dẫn này đề xuất một cách tiếp cận lấy con người làm trung tâm cũng như phù hợp với lứa tuổi trong quy trình kiểm chứng đạo đức và thiết kế sư phạm.

9 789230 002497

4 GIÁO DỤC CHẤT LƯỢNG

Các Mục tiêu Phát triển Bền vững