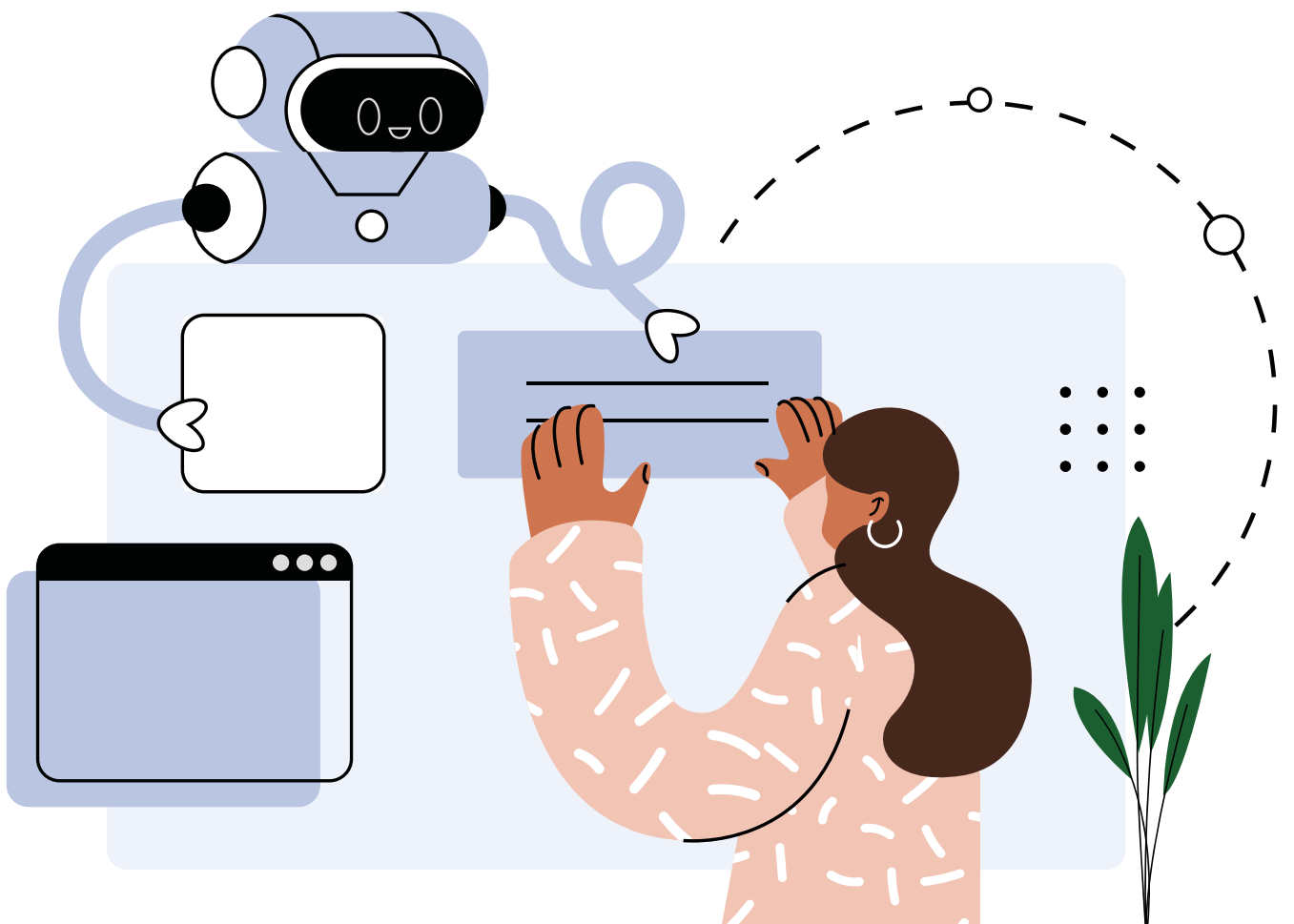


# Guidance for generative AI in education and research



---

## UNESCO – a global leader in education

Education is UNESCO's top priority because it is a basic human right and the foundation for peace and sustainable development. UNESCO is the United Nations' specialized agency for education, providing global and regional leadership to drive progress, strengthening the resilience and capacity of national systems to serve all learners. UNESCO also leads efforts to respond to contemporary global challenges through transformative learning, with special focus on gender equality and Africa across all actions.



## The Global Education 2030 Agenda

UNESCO, as the United Nations' specialized agency for education, is entrusted to lead and coordinate the Education 2030 Agenda, which is part of a global movement to eradicate poverty through 17 Sustainable Development Goals by 2030. Education, essential to achieve all of these goals, has its own dedicated Goal 4, which aims to *"ensure inclusive and equitable quality education and promote lifelong learning opportunities for all."* The Education 2030 Framework for Action provides guidance for the implementation of this ambitious goal and commitments.



---

Published in 2023 by the United Nations Educational, Scientific and Cultural Organization,  
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2023

ISBN 978-92-3-100612-8



This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (<https://www.unesco.org/en/open-access/cc-sa>).

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Images marked with an asterisk (\*) do not fall under the CC-BY-SA license and may not be used or reproduced without the prior permission of the copyright holders.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

Cover credit: Olexandra Simkina/Shutterstock.com\*

Designed and printed by UNESCO

Printed in France

# SHORT SUMMARY

## Towards a human-centred approach to the use of generative AI

Publicly available generative AI (GenAI) tools are rapidly emerging, and the release of iterative versions is outpacing the adaptation of national regulatory frameworks. The absence of national regulations on GenAI in most countries leaves the data privacy of users unprotected and educational institutions largely unprepared to validate the tools.

UNESCO's first global guidance on GenAI in education aims to support countries to implement immediate actions, plan long-term policies and develop human capacity to ensure a human-centred vision of these new technologies.

The Guidance presents an assessment of potential risks GenAI could pose to core humanistic values that promote human agency, inclusion, equity, gender equality, and linguistic and cultural diversities, as well as plural opinions and expressions.

It proposes key steps for governmental agencies to regulate the use of GenAI tools including mandating the protection of data privacy and considering an age limit for their use. It outlines requirements for GenAI providers to enable their ethical and effective use in education.

The Guidance stresses the need for educational institutions to validate GenAI systems on their ethical and pedagogical appropriateness for education. It calls on the international community to reflect on their long-term implications for knowledge, teaching, learning and assessment.

The publication offers concrete recommendations for policy-makers and educational institutions on how the uses of GenAI tools can be designed to protect human agency and genuinely benefit learners, teachers and researchers.

While **ChatGPT** reached **100 million** monthly active users in January 2023, only **one country** had released regulations on generative AI as of July 2023



**unesco**

*"Since wars begin in the minds of men and women it is in the minds of men and women that the defences of peace must be constructed"*



unesco

# Guidance for generative AI in education and research

## Foreword



© UNESCO

Generative artificial intelligence (GenAI) burst into the public awareness in late 2022 with the launch of ChatGPT, which became the fastest growing app in history. With the power to imitate human capabilities to produce outputs such as text, images, videos, music and software codes, these GenAI applications have caused a stir. Millions of people are now using GenAI in their daily lives and the potential of adapting the models to domain-specific AI applications seems unlimited.

Such wide-ranging capacities for information processing and knowledge production have potentially huge implications for education, as they replicate the higher-order thinking that constitutes the foundation of human learning. As GenAI tools are increasingly able to automate some basic levels of writing and artwork creation, they are forcing education policy-makers and institutions to revisit why, what and how we learn. These are now critical considerations for education in this new phase of the digital era.

This publication aims to support the planning of appropriate regulations, policies and human capacity development, to ensure that GenAI becomes a tool that genuinely benefits and empowers teachers, learners and researchers.

It proposes key steps for governmental agencies to regulate the use of generative AI. It also presents frameworks and concrete examples for policy formulation and instructional design that enable ethical and effective uses of this technology in education. Finally, it calls on the international community to consider the profound longer-term implications of generative AI for how we understand knowledge and define learning content, methods and outcomes, as well as the way in which we assess and validate learning.

Building on UNESCO's 2021 *Recommendation on the Ethics of Artificial Intelligence*, the guidance is anchored in a humanistic approach to education that promotes human agency, inclusion, equity, gender equality, and cultural and linguistic diversity, as well as plural opinions and expressions. Furthermore, it responds to the call of the 2021 report of the International Commission on the Futures of Education, *Reimagining our futures together: A new social contract for education* to redefine our relationship with technology, as an integral part of our efforts to renew the social contract for education.

AI must not usurp human intelligence. Rather, it invites us to reconsider our established understandings of knowledge and human learning. It is my hope that this guidance will help us redefine new horizons for education and inform our collective thinking and collaborative actions that can lead to human-centred digital learning futures for all.

A handwritten signature in black ink, which appears to read 'Stefania Giannini'.

**Stefania Giannini,**  
UNESCO Assistant Director-General for Education

# Acknowledgements

Under the leadership of Stefania Giannini, Assistant-Director for Education, and the guidance of Sobhi Tawil, Director of the Future of Learning and Innovation Division at UNESCO, the drafting of the publication was led by Fengchun Miao, Chief of Unit for Technology and AI in Education.

Particular thanks go to Wayne Holmes, Associate Professor at University College London, who co-drafted the publication.

This publication is the fruit of a collective effort of education leaders and experts in the field of AI and education.

It benefited from the insights and inputs of many experts including: Mutlu Cukurova, Professor at University College London; Colin de la Higuera, UNESCO Chair in Technologies for the Training of Teachers with Open Educational Resources at Nantes University; Shafika Isaacs, Research Associate at the University of Johannesburg; Natalie Lao, Executive Director of the App Inventor Foundation; Qin Ni, Associate Professor at Shanghai Normal University; Catalina Nicolin, ICT in Education Expert at the European Digital Education Hub in Romania; John Shaw-Taylor, UNESCO Chair in AI and Professor of Computational Statistics and Machine Learning at University College London; Kelly Shirohira, Executive Manager at Jet Education Services; Ki-Sang Song, Professor at Korea National University of Education; and Ilkka Tuomi, Chief Scientist at Meaning Processing Ltd in Finland.

Many colleagues across UNESCO also contributed in various ways including: Dafna Feinholz, Chief of Section for Bioethics and the Ethics of Science and Technology; Francesc Pedró, Director of the International Institute for Higher Education in Latin America and the Caribbean; Prateek Sibal, Programme Specialist, Section for Digital Policies and Digital Transformation; Saurabh Roy, Senior Project Officer at the Section for Teacher Development, Division for Policies and Lifelong Learning Systems; Benjamin Vergel De Dios, Programme Specialist in ICT in Education, Section for Educational Innovation and Skills Development in the Bangkok Office; the colleagues in the Diversity of Cultural Expressions Entity in the Culture Sector; and Mark West, Programme Specialist, Future of Learning and Innovation Division.

Appreciation is also due to Glen Hertelendy, Luisa Ferrara and Xianglei Zheng, Unit for Technology and AI in Education, Future of Learning and Innovation, for coordinating the production of the publication.

Gratitude is also extended to Jenny Webster for copy-editing and proofreading the text, and to Ngoc-Thuy Tran for designing the layout.

# Table of contents

<b>Foreword</b> .....	<b>2</b>
<b>Acknowledgements</b> .....	<b>3</b>
<b>List of acronyms and abbreviations</b> .....	<b>6</b>
<b>Introduction</b> .....	<b>7</b>
<b>1. What is generative AI and how does it work?</b> .....	<b>8</b>
1.1 What is generative AI? .....	8
1.2 How does generative AI work? .....	8
1.2.1 How text GenAI models work .....	9
1.2.2 How image GenAI models work .....	11
1.3 Prompt-engineering to generate desired outputs .....	11
1.4 Emerging EdGPT and its implications .....	13
<b>2. Controversies around generative AI and their implications for education</b> .....	<b>14</b>
2.1 Worsening digital poverty .....	14
2.2 Outpacing national regulatory adaptation .....	14
2.3 Use of content without consent .....	15
2.4 Unexplainable models used to generate outputs .....	15
2.5 AI-generated content polluting the internet .....	16
2.6 Lack of understanding of the real world .....	16
2.7 Reducing the diversity of opinions and further marginalizing already marginalized voices .....	17
2.8 Generating deeper deepfakes .....	17
<b>3. Regulating the use of generative AI in education</b> .....	<b>18</b>
3.1 A human-centred approach to AI .....	18
3.2 Steps to regulate GenAI in education .....	18
3.3 Regulations on GenAI: Key elements .....	20
3.3.1 Governmental regulatory agencies .....	20
3.3.2 Providers of GenAI tools .....	21
3.3.3 Institutional users .....	23
3.3.4 Individual users .....	23
<b>4. Towards a policy framework for the use of generative AI in education and research</b> .....	<b>24</b>
4.1 Promote inclusion, equity, and linguistic and cultural diversity .....	24
4.2 Protect human agency .....	24

4.3	Monitor and validate GenAI systems for education	25
4.4	Develop AI competencies including GenAI-related skills for learners	25
4.5	Build capacity for teachers and researchers to make proper use of GenAI	26
4.6	Promote plural opinions and plural expressions of ideas	26
4.7	Test locally relevant application models and build a cumulative evidence base	27
4.8	Review long-term implications in an intersectoral and interdisciplinary manner	27
<b>5.</b>	<b>Facilitating creative use of GenAI in education and research</b>	<b>28</b>
5.1	Institutional strategies to facilitate responsible and creative use of GenAI	28
5.2	A ‘human-centred and pedagogically appropriate interaction’ approach	29
5.3	Co-designing the use of GenAI in education and research	29
5.3.1	Generative AI for research	29
5.3.2	Generative AI to facilitate teaching	30
5.3.3	Generative AI as a 1:1 coach for the self-paced acquisition of foundational skills	31
5.3.4	Generative AI to facilitate inquiry or project-based learning	33
5.3.5	Generative AI to support learners with special needs	34
<b>6.</b>	<b>GenAI and the future of education and research</b>	<b>36</b>
6.1	Uncharted ethical issues	36
6.2	Copyright and intellectual property	36
6.3	Sources of content and learning	36
6.4	Homogenized responses versus diverse and creative outputs	37
6.5	Rethinking assessment and learning outcomes	37
6.6	Thinking processes	37
	Concluding remarks	38
<b>References</b>		<b>39</b>

### List of tables

Table 1.	Techniques used in generative AI	8
Table 2.	OpenAI GPTs	9
Table 3.	Co-designing uses of GenAI for research	30
Table 4.	Co-designing uses of GenAI to support teachers and teaching	31
Table 5.	Co-designing uses of GenAI as a 1:1 coach for the self-paced acquisition of foundational skills in languages and the arts	32
Table 6.	Co-designing uses of GenAI to facilitate inquiry or project-based learning	33
Table 7.	Co-designing uses of GenAI to support learners with special needs	34



# List of acronyms and abbreviations

## Concepts and technologies

<b>AGI</b>	Artificial general intelligence
<b>AI</b>	Artificial intelligence
<b>API</b>	Application programming interface
<b>ANN</b>	Artificial neural network
<b>DAI</b>	Distributed artificial intelligence
<b>GAN</b>	Generative adversarial networks
<b>GB</b>	Gigabytes
<b>GDPR</b>	General Data Protection Regulation
<b>GenAI</b>	Generative artificial intelligence
<b>GPT</b>	Generative pre-trained transformer
<b>ICT</b>	Information and communication technology
<b>LaMDA</b>	Language model for dialogue applications
<b>LLM</b>	Large language model
<b>ML</b>	Machine learning
<b>VAE</b>	Variational autoencoders

## Organizations

<b>AGCC</b>	AI Government Cloud Cluster (Singapore)
<b>CAC</b>	Cyberspace Administration of China
<b>EU</b>	European Union
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>UNCTAD</b>	United Nations Conference on Trade and Development
<b>UNESCO</b>	United Nations Educational, Scientific and Cultural Organization

# Introduction

The release of ChatGPT in late 2022, the first easy-to-use generative artificial intelligence (GenAI) tool made widely available to the public,<sup>1</sup> followed by iteratively more sophisticated versions, sent shock waves worldwide, and is fuelling the race among large technology companies to position themselves in the field of GenAI model development.<sup>2</sup>

Across the world, the initial concern in education was that ChatGPT and similar GenAI tools would be used by students to cheat on their assignments, thus undermining the value of learning assessment, certification and qualifications (Anders, 2023). While some educational institutions banned the use of ChatGPT, others cautiously welcomed the arrival of GenAI (Tlili, 2023). Many schools and universities, for instance, adopted a progressive approach believing that ‘rather than seek to prohibit their use, students and staff need to be supported in using GenAI tools effectively, ethically and transparently’ (Russell Group, 2023). This approach acknowledges that GenAI is widely available, is likely only to become more sophisticated, and has both specific negative and unique positive potential for education.

Indeed, GenAI has a myriad of possible uses. It can automate information processing and the presentation of outputs across all key symbolic representations of human thinking. It enables the delivery of final outputs by furnishing semi-finished knowledge products. By freeing humans from some categories of lower-order thinking skills, this new generation of AI tools might have profound implications for how we understand human intelligence and learning.

But GenAI also raises multiple immediate concerns related to issues such as safety, data privacy, copyright, and manipulation. Some of these are broader risks related to artificial intelligence that have been further exacerbated by GenAI, while others have newly emerged with this latest generation of tools. It is now urgent that each of these issues and concerns be fully understood and addressed.

This Guidance is designed to respond to this urgent need. However, a thematic set of guidance on GenAI for education should not be understood as a claim that GenAI is the solution to education’s fundamental challenges. Despite the media hyperbole, it is unlikely that GenAI alone will solve any of the problems facing education systems around the world. In responding to long-standing educational issues, it is key to uphold the idea that human capacity and collective action, and not technology, is the determining factor in effective solutions to fundamental challenges faced by societies.

This Guidance therefore aims to support the planning of appropriate regulations, policies and human capacity development programmes, to ensure that GenAI becomes a tool that genuinely benefits and empowers teachers, learners and researchers. Building on UNESCO’s *Recommendation on the Ethics of Artificial Intelligence*, the Guidance is anchored in a human-centred approach that promotes human agency, inclusion, equity, gender equality, and cultural and linguistic diversity, as well as plural opinions and expressions.

The Guidance first looks into what GenAI is and how it works, presenting the diverse technologies and models available (Section 1), before identifying a range of controversial ethical and policy issues around both AI in general, and GenAI specifically (Section 2). This is followed by a discussion of the steps and key elements to be examined when seeking to regulate GenAI based on a human-centred approach – one that ensures ethical, safe, equitable and meaningful use (Section 3). Section 4 then proposes measures that can be taken to develop coherent, comprehensive policy frameworks to regulate the use of GenAI in education and research, while Section 5 looks into the possibilities for creatively using GenAI in curriculum design, teaching, learning and research activities. Section 6 concludes the Guidance with considerations around the long-term implications of GenAI for education and research.

# 1. What is generative AI and how does it work?

## 1.1 What is generative AI?

Generative AI (GenAI) is an artificial intelligence (AI) technology that automatically generates content in response to prompts written in natural-language conversational interfaces. Rather than simply curating existing webpages, by drawing on existing content, GenAI actually produces new content. The content can appear in formats that comprise all symbolic representations of human thinking: texts written in natural language, images (including photographs, digital paintings and cartoons), videos, music and software code. GenAI is trained using data collected from webpages, social media conversations and other online media. It generates its content by statistically analysing the distributions of words, pixels or other elements in the data that it has ingested and identifying and repeating common patterns (for example, which words typically follow which other words).

While GenAI can produce new content, it cannot generate new ideas or solutions to real-world challenges, as it does not understand real-world objects or social relations that underpin language. Moreover, despite its fluent and impressive output, GenAI cannot be trusted to be accurate. Indeed, even

the provider of ChatGPT acknowledges, ‘While tools like ChatGPT can often generate answers that sound reasonable, they cannot be relied upon to be accurate.’ (OpenAI, 2023). Most often, the errors will go unnoticed unless the user has a solid knowledge of the topic in question.

## 1.2 How does generative AI work?

The specific technologies behind GenAI are part of the family of AI technologies called machine learning (ML) which uses algorithms to enable it to continuously and automatically improve its performance from data. The type of ML which has led to many of the advances in AI that we have seen in recent years, such as the use of AI for facial recognition, is known as artificial neural networks (ANNs), which are inspired by how the human brain works and its synaptic connections between neurons. There are many types of ANNs.

Both text and image generative AI technologies are based on a set of AI technologies that have been available to researchers for several years.<sup>1</sup> ChatGPT, for instance, uses a generative pre-trained transformer (GPT), while image GenAI typically uses what are known as generative adversarial networks (GANs) (see **Table 1**).<sup>3</sup>

**Table 1. Techniques used in generative AI**

<b>Machine learning (ML)</b>		A type of AI that uses data to automatically improve its performance.
<b>Artificial neural network (ANN)</b>		A type of ML that is inspired by the structure and functioning of the human brain (e.g. the synaptic connections between neurons).
<b>Text generative AI</b>	<b>General-purpose transformers</b>	A type of ANN that is capable of focusing on different parts of data to determine how they relate to each other
	<b>Large language models (LLM)</b>	A type of general-purpose transformer that is trained on vast amounts of text data.
	<b>Generative pre-trained transformer (GPT)<sup>4</sup></b>	A type of LLM that is pre-trained on even larger amounts of data, which allows the model to capture the nuances of language and generate coherent context-aware text.
<b>Image generative AI</b>	<b>Generative adversarial networks (GANs)</b>	Types of neural network used for image generation.
	<b>Variational autoencoders (VAEs)</b>	

### 1.2.1. How text GenAI models work

Text generative AI uses a type of ANN known as a general-purpose transformer, and a type of general-purpose transformer called a large language model. This is why AI Text GenAI systems are often referred to as large language models, or LLMs. The type of LLM used by text GenAI is known as a generative pre-trained transformer, or GPT (hence the 'GPT' in 'ChatGPT').

ChatGPT is built on GPT-3 which was developed by OpenAI. This was the third iteration of their GPT, the first being launched in 2018 and the most recent, GPT-4, in March 2023 (see **Table 2**). Each OpenAI GPT iteratively improved upon the previous through advances in AI architectures, training methods and optimization techniques. One well-known facet of its

continuous progress is the use of growing amounts of data to train its exponentially increasing number of 'parameters'. Parameters might be thought of as metaphorical knobs that can be adjusted to fine-tune the GPT's performance. They include the model's 'weights', numerical parameters that determine how the model processes input and produces output.

In addition to the advancements in optimizing AI architectures and training methods, this rapid iteration has been made possible also due to the massive amounts of data<sup>5</sup> and improvements in computing capabilities available to the big companies. Since 2012, computing capabilities used for training GenAI models have been doubling every 3-4 months. By comparison, Moore's Law had a two-year doubling period (OpenAI, 2018; Stanford University, 2019).

**Table 2. OpenAI GPTs**

Model	Launched	Amount of training data	Number of parameters	Characteristics
GPT-1	2018	40 GB	117 million	Capable of natural-language-processing tasks such as completing texts and answering questions.
GPT-2	2019	40 GB	1,500 million	Capable of more complex natural-language-processing tasks such as machine translation and summarizing.
GPT-3	2020	17,000 GB	175,000 million	Capable of advanced natural-language-processing tasks such as writing coherent paragraphs and generating entire articles. Also capable of adapting to new tasks with just a few examples.
GPT-4 <sup>6</sup>	2023	1,000,000 GB (reported but not confirmed)	170,000,000 million (reported but not confirmed)	Enhanced reliability and is capable of processing more complex instructions.

Once the GPT has been trained, generating a text response to a prompt involves the following steps:

1. The prompt is broken down into smaller units (called tokens) that are inputted into the GPT.
2. The GPT uses statistical patterns to predict likely words or phrases that might form a coherent response to the prompt.
  - The GPT identifies patterns of words and phrases that commonly co-occur in its prebuilt large data model (which comprises text scraped from the internet and elsewhere).
  - Using these patterns, the GPT estimates the probability of specific words or phrases appearing in a given context.
3. Beginning with a random prediction, the GPT uses these estimated probabilities to predict the next likely word or phrase in its response.
3. The predicted words or phrases are converted into readable text.
4. The readable text is filtered through what are known as 'guardrails' to remove any offensive content.
5. Steps 2 to 4 are repeated until a response is finished. The response is considered finished when it reaches a maximum token limit or meets predefined stopping criteria.

6. The response is post-processed to improve readability by applying formatting, punctuation and other enhancements (such as beginning the response with words that a human might use, such as 'Sure', 'Certainly' or 'I'm sorry').

While GPTs and their ability to automatically generate text have been available to researchers since 2018, what made the launch of ChatGPT so novel was its free access via an easy-to-use interface, meaning that anyone with internet access could explore the tool. The launch of ChatGPT set off shock waves around the world, and quickly led to other global tech companies playing catch-up, alongside numerous start-up companies, either by launching their own similar systems or by building new tools on top.

By July 2023, some of the alternatives to ChatGPT included the following:

- **Alpaca:**<sup>7</sup> A fine-tuned version of Meta's Llama, from Stanford University, which aims to address LLMs' false information, social stereotypes and toxic language.
- **Bard:**<sup>8</sup> An LLM from Google, based on its LaMDA and PaLM 2 systems, that has access to the internet in real time, which means it can provide up-to-date information.
- **Chatsonic:**<sup>9</sup> Made by Writesonic, it builds on ChatGPT while also crawling data directly from Google. Accordingly, it has less chance of producing factually incorrect answers.
- **Ernie** (also known as **Wenxin Yiyan** 文心一言):<sup>10</sup> A bilingual LLM from Baidu, still in development, which integrates extensive knowledge with massive datasets to generate text and images.
- **Hugging Chat:**<sup>11</sup> Made by Hugging Face, who emphasized ethics and transparency throughout its development, training and deployment. In addition, all data used to train their models are open source.
- **Jasper:**<sup>12</sup> A suite of tools and APIs that, for example, can be trained to write in a user's particular preferred style. It can also generate images.
- **Llama:**<sup>13</sup> An open-source LLM from Meta that requires less computing power and fewer resources to test new approaches, validate others' work and explore new use cases.
- **Open Assistant:**<sup>14</sup> An open-source approach designed to enable anyone with sufficient expertise to develop their own LLM. It was built on training data curated by volunteers.
- **Tongyi Qianwen** (通义千问):<sup>15</sup> An LLM from Alibaba that can respond to prompts in English or Chinese. It is being integrated into Alibaba's suite of business tools.
- **YouChat:**<sup>16</sup> An LLM that incorporates real-time search capabilities to provide additional context and insights in order to generate more accurate and reliable results.

Most of these are free to use (within certain limits), while some are open-source. Many other products are being launched that are based one of these LLMs. Examples include the following:

- **ChatPDF:**<sup>17</sup> Summarizes and answers questions about submitted PDF documents.
- **Elicit: The AI Research Assistant:**<sup>18</sup> Aims to automate parts of researchers' workflows, identifying relevant papers and summarizing key information.
- **Perplexity:**<sup>19</sup> Provides a 'knowledge hub' for people seeking quick, accurate answers tailored to their needs.

Similarly, LLM-based tools are being embedded into other products, such as web browsers. For example, extensions for the Chrome browser that are built on ChatGPT include the following:

- **WebChatGPT:**<sup>20</sup> Gives ChatGPT internet access to enable more accurate and up-to-date conversations.
- **Compose AI:**<sup>21</sup> Autocompletes sentences in emails and elsewhere.
- **TeamSmart AI:**<sup>22</sup> Provides a 'team of virtual assistants'.
- **Wiseone:**<sup>23</sup> Simplifies online information.

In addition, ChatGPT has been incorporated into some search engines,<sup>24</sup> and is being implemented across large portfolios of productivity tools (e.g. Microsoft Word and Excel), making it even more available in offices and educational institutions worldwide (Murphy Kelly, 2023).

Finally, as an interesting transition to image GenAI, the most recent GPT from OpenAI, GPT-4, is able to accept images as well as text in its prompts. In this sense, it is multimodal. Accordingly, some argue that the name ‘large language model’ (LLM) is becoming less appropriate, which is one reason why researchers at Stanford University have proposed the term ‘foundation model’ (Bommasani et al., 2021). This alternative is yet to be widely adopted.

### 1.2.2. How image GenAI models work

Image GenAI and music GenAI typically use a different type of ANN known as generative adversarial networks (GANs) which can also be combined with variational autoencoders. Some image GenAI models like Dall-E and Stable Diffusion use Diffusion Models, a different generative ANN. Taking GANs as example to explain how image GenAI models work: GANs have two parts (two ‘adversaries’), the ‘generator’ and the ‘discriminator’. In the case of image GANs, the generator creates a random image in response to a prompt, and the discriminator tries to distinguish between this generated image and real images. The generator then uses the result of the discriminator to adjust its parameters, in order to create another image. The process is repeated, possibly thousands of times, with the generator making more and more realistic images that the discriminator is less and less able to distinguish from real images. For example, a successful GAN trained on a dataset of thousands of landscape photographs might generate new but unreal images of landscapes that are almost indistinguishable from real photographs. Meanwhile, a GAN trained on a dataset of popular music (or even music by a single artist) might generate new pieces of music that follow the structure and complexity of the original music.

As of July 2023, the **image GenAI** models that are available include the following, all of which generate images from text prompts. Most are free to use, within certain limits:

- **Craiyon:**<sup>25</sup> Previously known as DALL•E mini.
- **DALL•E 2:**<sup>26</sup> OpenAI’s image GenAI tool.
- **DreamStudio:**<sup>27</sup> Stable Diffusion’s image GenAI tool.
- **Fotor:**<sup>28</sup> Incorporates GenAI in a range of image-editing tools.

- **Midjourney:**<sup>29</sup> An independent image GenAI tool.
- **NightCafe:**<sup>30</sup> Interface to Stable Diffusion and DALL•E 2.
- **Photosonic:**<sup>31</sup> WriteSonic’s AI art generator.

Examples of easy-to-access **video GenAI** include the following:

- **Elai:**<sup>32</sup> Can convert presentations, websites and text into videos.
- **GliaCloud:**<sup>33</sup> Can generate videos from news content, social media posts, live sporting events and statistical data.
- **Pictory:**<sup>34</sup> Can automatically create short videos from long-form content.
- **Runway:**<sup>35</sup> Offers a range of video (and imaging) generation and editing tools.

Finally, these are some examples of easy-to-access **music GenAI**:

- **Aiva:**<sup>36</sup> Can automatically create personalized soundtracks.
- **Boomy,**<sup>37</sup> **Soundraw,**<sup>38</sup> and **Voicemod:**<sup>39</sup> Can generate songs from any text, and require no musical composition knowledge.

## 1.3 Prompt-engineering to generate desired outputs

While using GenAI can be as simple as typing in a question or other prompt, the reality is that it is still not straightforward for the user to get exactly the output that they want. For example, the breakthrough AI image *Théâtre D’opéra Spatial* which won a prize at the Colorado State Fair in the United States of America, took weeks of writing prompts and fine-tuning hundreds of images in order to generate the final submission (Roose, 2022). The similar challenge of writing effective prompts for text GenAI has led to an increasing number of prompt-engineering jobs appearing on recruitment websites (Popli, 2023). ‘Prompt-engineering’ refers to the processes and techniques for composing input to produce GenAI output that more closely resembles the user’s desired intent.

Prompt-engineering is most successful when the prompt articulates a coherent chain of reasoning centred on a particular problem or a chain of thought in a logical order. Specific recommendations include:

- Use **simple**, clear and straightforward language that can be easily understood, avoiding complex or ambiguous wording.
- Include **examples** to illustrate the desired response or format of generated completions.
- Include **context**, which is crucial for generating relevant and meaningful completions.
- **Refine** and iterate as necessary, experimenting with different variations.
- Be **ethical**, avoiding prompts that may generate inappropriate, biased or harmful content.

It is also important to recognize immediately that GenAI outputs cannot be relied upon without critical evaluation. As OpenAI write about their most sophisticated GPT:<sup>40</sup>

“Despite its capabilities, GPT-4 has similar limitations as earlier GPT models. Most importantly, it still is not fully reliable (it ‘hallucinates’ facts and makes reasoning errors). Great care should be taken when using language model outputs, particularly in high-stakes contexts, with the exact protocol (such as human review, grounding with additional context, or avoiding high-stakes uses altogether) matching the needs of a specific use-case.”



### Implications for education and research

While GenAI might help teachers and researchers generate useful text and other outputs to support their work, it is not necessarily a straightforward process. It can take multiple iterations of a prompt before the desired output is achieved. A worry is that young learners, because they are by definition less expert than teachers, might unknowingly and without critical engagement accept GenAI output that is superficial, inaccurate or even harmful.

In light of the quality of GenAI’s outputs, rigorous user tests and performance evaluations should be conducted before validating the tools for large-scale or high-stakes adoption. Such exercises should be designed with a performance metric that is most relevant to the type of task for which users ask GenAI to provide outputs. For example, for solving math problems, ‘accuracy’ could be used as the main metric to quantify how often a GenAI tool produces the correct answer; for responding to sensitive questions, the main metric to measure performance might be ‘answer rate’ (the frequency with which the GenAI directly answers a question); for code generation, the metric may be ‘the fraction of the generated codes that are directly executable’ (whether the generated code could be directly executed in a programming environment and pass the unit tests); and for visual reasoning, the metric could be ‘exact match’ (whether the generated visual objects exactly match the ground truth) (Chen et al., 2023).

In summary, at a superficial level, GenAI is easy to use; however, more sophisticated outputs need skilled human input and must be critically evaluated before they are used.

## 1.4 Emerging EdGPT and its implications

Given that GenAI models can serve as the basis or starting point for developing more specialized or domain-specific models, some researchers have suggested that GPTs should be renamed 'foundation models' (Bommasani et al., 2021). In education, developers and researchers have started to fine-tune a foundation model to develop 'EdGPT'.<sup>41</sup> EdGPT models are trained with specific data to serve educational purposes. In other words, EdGPT aims to refine the model that has been derived from massive amounts of general training data with smaller amounts of high-quality, domain-specific education data.

This potentially gives EdGPT more scope to support the achievement of the transformations listed in Section 4.3. For example, EdGPT models targeting curriculum co-design may allow educators and learners to generate appropriate educational materials such as lesson plans, quizzes and interactive activities that closely align with an effective pedagogical approach and specific curricular objectives and levels of challenge for particular learners. Similarly, in the context of a 1:1 language skills coach, a foundation model refined with texts appropriate for a particular language might be used to generate exemplar sentences, paragraphs or conversations for practice. When learners interact with the model, it can respond with relevant and grammatically accurate text at the right level for them. Theoretically, the outputs of EdGPT models could also contain fewer general biases or otherwise objectionable content than standard GPT,

but still might generate errors. It is critical to note that, unless the underlying GenAI models and approach change significantly, EdGPT may still generate errors and demonstrate other limitations. Accordingly, it is still important that the main users of EdGPT, especially teachers and learners, need to take a critical perspective to any outputs.

Currently, the refining of foundation models for more targeted use of GPT in education is at an early stage. Existing examples include EduChat, a foundation model developed by East China Normal University to provide services for teaching and learning, and whose codes, data and parameters are shared as open source.<sup>42</sup> Another example is MathGPT being developed by the TAL Education Group – a LLM that focuses on mathematics-related problem-solving and lecturing for users worldwide.<sup>43</sup>

However, before significant progress is possible, it is essential that efforts are put into refining foundation models not only through adding subject knowledge and de-biasing, but also through adding knowledge about relevant learning methods, and how this can be reflected in the design of algorithms and models. The challenge is to determine the extent to which EdGPT models can go beyond subject knowledge to also target student-centred pedagogy and positive teacher-student interactions. The further challenge is to determine the extent to which learner and teacher data may ethically be collected and used in order to inform an EdGPT. Finally, there is also a need for robust research to ensure that EdGPT does not undermine students' human rights nor disempower teachers.



## 2. Controversies around generative AI and their implications for education

Having previously discussed what GenAI is and how it works, this section examines controversies and ethical risks raised by all GenAI systems and considers some of the implications for education.

### 2.1 Worsening digital poverty

As noted earlier, GenAI relies upon huge amounts of data and massive computing power in addition to its iterative innovations in AI architectures and training methods, which are mostly only available to the largest international technology companies and a few economies (mostly the United States, People's Republic of China, and to a lesser extent Europe). This means that the possibility to create and control GenAI is out of reach of most companies and most countries, especially those in the Global South.

As access to data becomes increasingly essential for the economic development of countries and for the digital opportunities of individuals, those countries and people who do not have access to or cannot afford enough data are left in a situation of 'data poverty' (Marwala, 2023). The situation is similar for access to computing power. The rapid pervasion of GenAI in technologically advanced countries and regions has accelerated exponentially the generation and processing of data, and has simultaneously intensified the concentration of AI wealth in the Global North. As an immediate consequence, the data-poor regions have been further excluded and put at long-term risk of being colonized by the standards embedded in the GPT models. The current ChatGPT models are trained on data from online users which reflect the values and norms of the Global North, making them inappropriate for locally relevant AI algorithms in data-poor communities in many parts of the Global South or in more disadvantaged communities in the Global North.



#### Implications for education and research

Researchers, teachers and learners should take a critical view of the value orientations, cultural standards and social customs embedded in GenAI training models. Policy-makers should be aware of and take action to address the worsening of inequities caused by the widening divide in training and controlling GenAI models.

### 2.2 Outpacing national regulatory adaptation

Dominant GenAI providers have also been criticized for not allowing their systems to be subject to rigorous independent academic review (Dwivedi et al., 2023).<sup>44</sup> The foundational technologies of a company's GenAI tend to be protected as corporate intellectual property. Meanwhile many of the companies that are starting to use GenAI are finding it increasingly challenging to maintain the security of their systems (Lin, 2023). Moreover, despite calls for regulation from the AI industry itself,<sup>45</sup> the drafting of legislation on the creation and use of all AI, including GenAI, often lags behind the rapid pace of development. This partly explains the challenges experienced by national or local agencies in understanding and governing the legal and ethical issues.<sup>46</sup>

While GenAI may augment human capacities in completing certain tasks, there is limited democratic control of the companies that are promoting GenAI. This raises the question of regulations, in particular in respect of access to, and use of, domestic data including data on local institutions and individuals as well as data generated on the countries' territory. Appropriate legislation is needed so that local governmental agencies may gain some control over the surging waves of GenAI to ensure its governance as a public good.



### Implications for education and research

Researchers, teachers and learners should be aware of the lack of appropriate regulations to protect the ownership of domestic institutions and individuals and the rights of domestic users of GenAI, and to respond to legislative issues triggered by GenAI.

## 2.3 Use of content without consent

As noted earlier, GenAI models are built from large amounts of data (e.g. text, sounds, code and images) often scraped from the internet and usually without any owner's permission. Many image GenAI systems and some code GenAI systems have consequently been accused of violating intellectual property rights. At the time of writing, there are several ongoing international legal cases that relate to this issue.

Furthermore, some have pointed out that GPTs may contravene laws such as the European Union's (2016) General Data Protection Regulation or GDPR, especially people's right to be forgotten, as it is currently impossible to remove someone's data (or the results of that data) from a GPT model once it has been trained.



### Implications for education and research

- Researchers, teachers and learners need to know the rights of data owners and should check whether the GenAI tools they are using contravene any existing regulations.
- Researchers, teachers and learners should also be aware that the images or codes created with GenAI might violate someone else's intellectual property rights, and that images, sounds or code that they create and share on the internet might be exploited by other GenAI.

## 2.4 Unexplainable models used to generate outputs

It has long been recognized that artificial neural networks (ANNs) are usually 'black boxes'; that is, that their inner workings are not open to inspection. As a result, ANNs are not 'transparent' or 'explainable', and it is not possible to ascertain how their outputs were determined.

While the overall approach, including the algorithms used, is generally explainable, the particular models and their parameters, including the model's weights, are not inspectable, which is why a specific output that is generated cannot be explained. There are billions of parameters/weights in a model like GPT-4 (see **Table 2**) and it is the weights collectively that hold the learned patterns that the model uses to generate its outputs. As parameters/weights are not transparent in ANNs (**Table 1**), one cannot explain the precise way a specific output is created by these models.

GenAI's lack of transparency and explainability is increasingly problematic as GenAI becomes ever more complex (see **Table 2**), often producing unexpected or undesired results. In addition, GenAI models inherit and perpetuate biases present in their training data which, given the non-transparent nature of the models, are hard to detect and address. Finally, this opacity is also a key cause of trust issues around GenAI (Nazaretsky et al., 2022a). If users don't understand how a GenAI system arrived at a specific output, they are less likely to be willing to adopt it or use it (Nazaretsky et al., 2022b).



### Implications for education and research

Researchers, teachers and learners should be aware that GenAI systems operate as black boxes and that it is consequently difficult, if not impossible, to know why particular content has been created. A lack of explanation of how the outputs are generated tends to lock users in the logic defined by parameters designed in the GenAI systems. These parameters may reflect specific cultural or commercial values and norms that implicitly bias the content produced.

## 2.5 AI-generated content polluting the internet

Because GPT training data is typically drawn from the internet, which all too frequently includes discriminatory and other unacceptable language, developers have had to implement what they call ‘guardrails’ to prevent GPT output from being offensive and/or unethical. However, due to the absence of strict regulations and effective monitoring mechanisms, biased materials generated by GenAI are increasingly spreading throughout the internet, polluting one of the main sources of content or knowledge for most learners across the world. This is especially important because the material generated by GenAI can appear to be quite accurate and convincing, when often it contains errors and biased ideas. This poses a high risk for young learners who do not have solid prior knowledge of the topic in question. It also poses a recursive risk for future GPT models that will be trained on text scraped from the Internet that GPT models have themselves created which also include their biases and errors.



### Implications for education and research

- Researchers, teachers and learners need to be aware that GenAI systems are capable of outputting offensive and unethical materials.
- They also need to know about the long-term issues that will potentially arise for the reliability of knowledge when future GPT models are based on text that previous GPT models have generated.

## 2.6 Lack of understanding of the real world

Text GPTs are sometimes pejoratively referred to as ‘stochastic parrots’ because, as has been noted earlier, while they can produce text that appears convincing, that text often contains errors and can include harmful statements (Bender et al., 2021). This all occurs because GPTs only repeat language patterns found in their training data (usually text drawn from the internet),

starting with random (or ‘stochastic’) patterns, and without understanding their meaning – just as a parrot can mimic sounds without actually comprehending what it is saying.

The disconnect between GenAI models ‘appearing’ to understand the text that they use and generate, and the ‘reality’ that they do not understand the language and the real world can lead teachers and students to place a level of trust in the output that it does not warrant. This poses serious risks for future education. Indeed, GenAI is not informed by observations of the real world or other key aspects of the scientific method, nor is it aligned with human or social values. For these reasons, it cannot generate genuinely novel content about the real world, objects and their relations, people and social relations, human-object relations, or human-tech relations. Whether the apparently novel content generated by GenAI models can be recognized as scientific knowledge is contested.

As already noted, GPTs can frequently produce inaccurate or unreliable text. In fact, it is well known that GPTs make up some things that do not exist in real life. Some call this ‘hallucination’, although others criticize the use of such an anthropomorphic and therefore misleading term. This is acknowledged by the companies producing GenAI. The bottom of the ChatGPT public interface, for instance, states: ‘ChatGPT may produce inaccurate information about people, places, or facts’<sup>2</sup>

It has also been suggested by a few advocates that GenAI represents a significant step in the journey towards artificial general intelligence (AGI), a term suggesting a class of AI that is more intelligent than humans. However, this has long been critiqued, with the argument that AI will never progress towards AGI at least until it in some way brings together, in symbiosis, both knowledge-based AI (also known as symbolic or rule-based AI) and data-based AI (also known as machine learning) (Marcus, 2022). The AGI or sentience claims also distract us from more careful consideration of current harms being perpetrated with AI, such as hidden discrimination against already discriminated-against groups (Metz, 2021).



### Implications for education and research

- The output of a text GenAI can look impressively human-like, as if it understood the text that it generated. However, GenAI does not understand anything. Instead, these tools string words together in ways that are common on the internet. The text that is generated can also be incorrect.
- Researchers, teachers and learners need to be aware that a GPT does not understand the text that it generates; that it can, and often does, generate incorrect statements; and that they therefore need to take a critical approach to everything that it does generate.

## 2.7 Reducing the diversity of opinions and further marginalizing already marginalized voices

ChatGPT and similar such tools tend to output only standard answers that assume the values of the owners/creators of the data used to train the models. Indeed, if a sequence of words appears frequently in the training data – as is the case with common and uncontroversial topics and mainstream or dominant beliefs – it is likely to be repeated by the GPT in its output.

This risks constraining and undermining the development of plural opinions and plural expressions of ideas. Data-poor populations, including marginalized communities in the Global North, have minimal or limited digital presence online. Their voices are consequently not being heard and their concerns are not represented in the data being used to train GPTs, and so rarely appear in the outputs. For these reasons, given the pre-training methodology based on data from internet web pages and social media conversations, GPT models can further marginalize already disadvantaged people.



### Implications for education and research

- While the developers and providers of GenAI models have the primary responsibility for continuously addressing biases in the datasets and outputs of these models, the user-side researchers, teachers and learners need to know that the output of text GenAI represents only the most common or dominant view of the world at the time when its training data was produced and that some of it is problematic or biased (e.g. stereotypical gender roles).
- Learners, teachers and researchers should never accept the information provided by the GenAI at face value and should always critically assess it.
- Researchers, teachers and learners also must be aware of how minority voices can be left out, because minority voices are by definition less common in the training data.

## 2.8 Generating deeper deepfakes

In addition to the controversies common to all GenAI, GAN GenAI can be used to alter or manipulate existing images or videos to generate fake ones that are difficult to distinguish from real ones. GenAI is making it increasingly easy to create these 'deepfakes' and so-called 'fake news'. In other words, GenAI is making it easier for certain actors to commit unethical, immoral and criminal acts, such as spreading disinformation, promoting hate speech and incorporating the faces of people, without their knowledge or consent, into entirely fake and sometimes compromising films.



### Implications for education and research

While it is the obligation of GenAI providers to protect the copyright and portrait rights of users, researchers, teachers and learners also need to be aware that any images they share on the internet may be incorporated into GenAI training data and might be manipulated and used in unethical ways.

## 3. Regulating the use of generative AI in education

In order to address the controversies around generative AI and to harness the potential benefits of GenAI in education, it first needs to be regulated. Regulation of GenAI for educational purposes requires a number of steps and policy measures based on a human-centred approach to ensure its ethical, safe, equitable and meaningful use.

### 3.1 A human-centred approach to AI

UNESCO's 2021 *Recommendation on the Ethics of Artificial Intelligence* provides the requisite normative framework to start addressing the multiple controversies around generative AI, including those that pertain to education and research. It is based on a human-centred approach to AI which advocates that the use of AI should be at the service of the development of human capabilities for inclusive, just and sustainable futures. Such an approach must be guided by human rights principles, and the need to protect human dignity and the cultural diversity that defines the knowledge commons. In terms of governance, a human-centred approach requires proper regulation that can ensure human agency, transparency and public accountability.

The 2019 *Beijing Consensus on Artificial Intelligence (AI) and Education* further elaborates what a human-centred approach implies for the use of AI in the context of education. The Consensus affirms that the use of AI technologies in education should enhance human capacities for sustainable development and effective human-machine collaboration in life, learning and work. It also calls for further actions to ensure equitable access to AI to support marginalized people and address inequalities, while promoting linguistic and cultural diversities. The Consensus suggests adopting whole-of-government, intersectoral and multistakeholder approaches to the planning of policies on AI in education.

*AI and education: Guidance for policy-makers* (UNESCO, 2022b) further refines what a human-centred approach means when examining the benefits and risks of AI in education and the role of education as a means of developing AI competencies. It proposes concrete

recommendations for the formulation of policies to steer the use of AI to (i) enable inclusive access to learning programmes, especially for vulnerable groups such as learners with disabilities; (ii) support personalized and open learning options; (iii) improve data-based provisions and management to expand access and improve quality in learning; (iv) monitor learning processes and alert teachers to failure risks; and (v) develop understanding and skills for the ethical and meaningful use of AI.

### 3.2 Steps to regulate GenAI in education

Prior to the release of ChatGPT, governments had been developing or adapting frameworks for regulating the collection and use of data and the adoption of AI systems across sectors including in education, which provided a legislative and policy context for the regulation of newly emergent AI applications. In the aftermath of the release of multiple competitive GenAI models starting in November 2022, governments have been adopting different policy responses – from banning GenAI to assessing needs for adapting existing frameworks, to urgently formulating new regulations.

Governmental strategies for regulating and facilitating the creative use of GenAI were mapped and reviewed in April 2023 (UNESCO, 2023b).<sup>47</sup> The review suggests a series of seven steps that governmental agencies can take to regulate generative AI and reassert public control in order to leverage its potentials across sectors, including in education.

#### **Step 1: Endorse international or regional general data protection regulations or develop national ones**

The training of GenAI models has involved collecting and processing online data from citizens across many countries. The use, by GenAI models, of data and content without consent is further challenging the issue of data protection.

General data protection regulations, with the EU's GDPR enacted in 2018 as one of the forerunner

examples, provide the necessary legal framework to regulate the collection and processing of personal data by the suppliers of GenAI. According to the Data Protection and Privacy Legislation Worldline portal of the United Nations Conference on Trade and Development (UNCTAD), 137 out of 194 countries have established legislation to safeguard data protection and privacy.<sup>48</sup>

The extent to which these frameworks are being implemented in those countries, however, remains unclear. It is therefore ever more critical to ensure that these are properly implemented, including regular monitoring of the operations of GenAI systems. It is also urgent for countries that do not yet have general data protection laws to develop them.

### **Step 2: Adopt/revise and fund whole-of-government strategies on AI**

Regulating generative AI must be part and parcel of broader national AI strategies that can ensure safe and equitable use of AI across development sectors, including in education. The formulation, endorsement, funding and implementation of national AI strategies requires a whole-of-government approach. Only such an approach can ensure the coordination of intersectoral actions required for integrated responses to emerging challenges.

By early 2023, some 67 countries<sup>49</sup> had developed or planned national strategies on AI, with 61 of them taking the form of a standalone AI strategy, and 7 being chapters on AI integrated within broader national ICT or digitalization strategies. Understandably, given its novelty, none of these national strategies had yet covered generative AI as a specific issue at the time of writing.

It is critical that countries revise existing national AI strategies, or develop them, ensuring provisions to regulate the ethical use of AI across sectors including in education.

### **Step 3: Solidify and implement specific regulations on the ethics of AI**

In order to address the ethical dimensions posed by the use of AI, specific regulations are required.

The UNESCO 2023 review of existing national AI strategies indicates that the identification of such ethical issues and the formulation of guiding principles is only common to some 40 national AI strategies.<sup>50</sup> And even here, the ethical principles will need to be translated into enforceable laws or regulations. This is seldom the case. Indeed, only around 20 countries had defined any clear regulations on the ethics of AI including as they relate to education, either as part of national AI strategies or otherwise. Interestingly, while education is highlighted as a policy domain across some 45 national AI strategies,<sup>51</sup> references to education are articulated more in terms of AI skills and talent development required to support national competitiveness, and less in terms of ethical issues.

Countries that do not yet have regulations on ethics of AI must urgently articulate and implement them.

### **Step 4: Adjust or enforce existing copyright laws to regulate AI-generated content**

The increasingly pervasive use of GenAI has introduced new challenges for copyright, both concerning the copyrighted content or work that models are trained on, as well as the status of the 'non-human' knowledge outputs they produce.

At present, only China, EU countries and the United States have adjusted copyright laws to account for the implications of generative AI. The US Copyright Office, for instance, has ruled that the output of GenAI systems such as ChatGPT are not protectable under US copyright law, arguing that 'copyright can protect only material that is the product of human creativity' (US Copyright Office, 2023). Meanwhile in the EU, the proposed EU AI Act requires AI tool developers to disclose the copyrighted materials they used in building their systems (European Commission, 2021). China, through its regulation on GenAI released in July 2023, requires the labelling of outputs of GenAI as AI-generated content, and only recognizes them as outputs of digital synthesis.

Regulating the use of copyrighted materials in the training of GenAI models and defining the copyright status of GenAI outputs are emerging as new accountabilities of copyright laws. It is urgent that existing laws be adjusted to account for this.

### Step 5: Elaborate regulatory frameworks on generative AI

The rapid pace of development of AI technologies is forcing national/local governance agencies to speed up their renewal of regulations. As of July 2023, only one country, China, had released specific official regulations on GenAI. The Provisional Regulations on Governing the Service of Generative AI released on 13 July 2023 (Cyberspace Administration of China, 2023a) requires providers of GenAI systems to label AI-generated content, images and videos properly and lawfully in accordance with its existing Regulation on Deep Synthesis in the Framework of Online Information Services. More of such national GenAI-specific frameworks need to be developed based upon an assessment of the gaps in existing local regulations and laws.

### Step 6: Build capacity for proper use of GenAI in education and research

Schools and other educational institutions need to develop capacities to understand the potential benefits and risks of AI, including GenAI, for education. It is only based on such understanding that they can validate the adoption of AI tools. Moreover, teachers and researchers need to be supported to strengthen their capacities for the proper use of GenAI, including through training and continuous coaching. A number of countries have launched such capacity-building programmes, including Singapore, which has been offering a dedicated platform for the AI capacity development of educational institutions through its AI Government Cloud Cluster which includes a dedicated repository of GPT models (Ocampo, 2023).

### Step 7: Reflect on the long-term implications of GenAI for education and research

The impact of current versions of GenAI is just beginning to unfold, and their effects on education are yet to be fully explored and understood. Meanwhile, stronger versions of GenAI and other classes of AI continue to be developed and deployed. Crucial questions remain, however, around the implications of GenAI for knowledge creation, transmission and validation – for teaching and learning, for curriculum design and assessment, and for research and copyright. Most countries are at the early stage of the adoption of GenAI in education, even as the longer-term impacts

have yet to be understood. To ensure a human-centred use of AI, open public debate and policy dialogues on the long-term implications should urgently be conducted. Inclusive debate involving government, the private sector and other partners, should serve to provide insights and inputs for the iterative renewal of regulations and policies.

## 3.3 Regulations on GenAI: Key elements

All countries need to properly regulate GenAI in order to ensure it benefits development in education and other contexts. This section proposes actions around key elements that can be taken by: (1) governmental regulatory agencies, (2) providers of AI-enabled tools, (3) institutional users, and (4) individual users. While many of the elements in the framework are of a transnational nature, all should also be considered in light of the local context, that is, the specific country's educational systems and general regulatory frameworks already in place.

### 3.3.1. Governmental regulatory agencies

A whole-of-government approach is required for the coordination of the design, alignment and implementation of regulations on GenAI. The following seven key elements and actions are recommended:

- **Intersectoral coordination:** Establish a national body to lead on the whole-of-government approach to GenAI and coordinate cooperation across sectors.
- **Alignment of legislation:** Align the framework with the relevant legislative and regulatory contexts of each country – with, for example, general data protection laws, regulations on internet security, laws on the security of data produced from or used to serve citizens, and other relevant legislation and usual practices. Assess the appropriateness of existing regulations and any necessary adaptations in response to new issues raised by GenAI.
- **Balance between the regulation of GenAI and the promotion of AI innovation:** Promote intersectoral cooperation among companies, organizations, and education

and research institutions, as well as relevant public agencies to jointly develop trustworthy models; encourage the building of open-source eco-systems to promote the sharing of super-computing resources and high-quality pre-training datasets; and foster the practical application of GenAI across sectors and the creation of high-quality content for the public good.

- **Assessment and classification of the potential risks of AI:** Establish principles and a process for the assessment and categorization of the efficacy, safety and security of GenAI services, before they are deployed and throughout the system's life cycle. Consider categorization mechanisms based on the levels of risk that GenAI may imply for citizens. Classify them into strict regulations (i.e. banning AI-enabled applications or systems with unacceptable risks), special regulations for high-risk applications, and general regulations on applications that are not listed as high risk. See the EU's draft AI Act for an example of this approach.
- **Protection of data privacy:** Account for the fact that the use of GenAI almost always involves users sharing their data with the GenAI provider. Mandate the drafting and implementation of laws for the protection of users' personal information and identify and combat unlawful data storage, profiling and sharing.
- **Definition and enforcement of age limit for the use of GenAI:** Most GenAI applications are primarily designed for adult users. These applications often entail substantial risks for children, including exposure to inappropriate content as well as the potential for manipulation. In light of these risks and given the considerable uncertainty that continues to surround iterative GenAI applications, age restrictions are strongly recommended for general-purpose AI technologies in order to protect children's rights and wellbeing.

Currently, the terms of use for ChatGPT require that users must be at least 13 years old, and users under 18 must have their parent or legal guardian's permission to use the services.<sup>52</sup> These age restrictions or thresholds are derived from the Children's Online Privacy

Protection Act of the United States of America (Federal Trade Commission, 1998). Passed in 1998 before widespread social media use and well before the creation of easy-to-use and powerful GenAI applications such as ChatGPT, the US law specifies that organizations or individual social media providers are not allowed to provide services for children under the age of 13 without parental permission. Many commentators understand this threshold to be too young and have advocated for legislation to raise the age to 16. The GDPR of the European Union (2016) specifies that users must be at least 16 years old to use the services of social media without parental permissions.

The emergence of various GenAI chatbots demand that countries carefully consider – and publicly deliberate – the appropriate age threshold for independent conversations with GenAI platforms. The minimum threshold should be 13 years of age. Countries will also need to decide if self-reporting age remains an appropriate means of age verification. Countries will need to mandate the accountabilities of GenAI providers for age verification and accountabilities of parents or guardians for monitoring the independent conversations of underage children.

- **National data ownership and the risk of data poverty:** Take legislative measures to protect national data ownership and regulate providers of GenAI that operate within its borders. For datasets generated by citizens that are being used for commercial purposes, establish regulations to promote mutual beneficial cooperation so that this category of data shall not be drained from the country to be exploited exclusively by the big tech companies.

### 3.3.2. Providers of GenAI tools

Providers of GenAI include organizations and individuals who are responsible for developing and making available GenAI tools, and/or are using GenAI technologies to provide services including through programmable application programming interfaces (APIs). Most of the influential providers of GenAI tools are extremely well-funded companies. It should be made clear to GenAI providers that they are accountable for ethics by design, including



for implementing the ethical principles stipulated in the regulations. The following ten categories of accountabilities should be covered:

- **Human accountabilities:** GenAI providers should be held responsible for ensuring adherence to core values and lawful purposes, respecting intellectual property, and upholding ethical practices, while also preventing the spread of disinformation and hate speech.
- **Trustworthy data and models:** GenAI providers should be required to evidence the trustworthiness and ethics of the data sources and methods used by their models and outputs. They must be mandated to adopt data and foundation models with proven legal sources, and abide by the relevant intellectual property laws (e.g. if the data are protected by intellectual property rights). In addition, when the models need to use personal information, the collection of said information should take place only with the informed and explicit consent of the owners.
- **Non-discriminatory content generation:** Providers of GenAI must prohibit the design and deployment of GenAI systems that generate biased or discriminatory content based on race, nationality, gender or other protected characteristics. They should ensure that robust 'guardrails' are in place to prevent GenAI producing offensive, biased or false content, while ensuring that the humans involved in informing the guardrails are protected and not exploited.
- **Explainability and transparency of GenAI models:** Providers should submit to public governance agencies their explanations of the sources, scale and types of data used by the models; their rules for labelling data in pre-training; the methods or algorithms that their models use to generate content or responses; and the services that their GenAI tools are providing. When necessary, they should offer support to help governance agencies understand the technology and data. GenAI's propensity to generate content with errors and contestable responses should be made transparent for users.
- **Labelling of GenAI content:** In accordance with relevant laws or regulations on the AI-assisted synthesis of online information, providers need to label GenAI-generated papers, reports, images and videos properly and lawfully. For example, GenAI output should be clearly labelled as having been produced by a machine.
- **Security and safety principles:** Providers of GenAI should ensure secure, robust and sustainable service throughout the life cycle of a GenAI system.
- **Specifications on appropriateness for access and use:** Providers of GenAI should provide clear specifications on the appropriate audience for, and use scenarios and purposes of, their services and help users of GenAI tools to make rational and responsible decisions.
- **Acknowledging the limitations and preventing predictable risks:** Providers of GenAI should clearly advertise the limitations of the methods used by the systems and their outputs. They need to develop technologies to ensure that the input data, methods, and outputs do no predictable harm to users, together with protocols to mitigate unpredictable harms when they occur. They must also provide guidance to help users understand GenAI-generated content based on ethical principles, and to prevent their over-reliance on and addiction to the generated content.
- **Mechanisms for complaints and remedies:** Providers of GenAI need to establish mechanisms and channels for the collection of complaints from users and the wider public, and take timely actions to accept and process these complaints.
- **Monitoring and reporting of unlawful use:** Providers shall cooperate with public governance agencies to facilitate the monitoring and reporting of unlawful use. This includes when people use GenAI products in ways that are illegal or violate ethical or social values such as promoting disinformation or hate speech, generating spam or composing malware.

### 3.3.3. Institutional users

Institutional users include educational authorities and institutions such as universities and schools that hold responsibilities for determining whether GenAI should be adopted and which types of GenAI tools should be procured and deployed within the institution.

- **Institutional auditing of GenAI algorithms, data and outputs:** Implement mechanisms to monitor as best as possible the algorithms and data used by GenAI tools and the outputs they generate. This should include regular audits and assessments, the protection of user data, and automatically filtering out inappropriate content.
- **Validating proportionality and protecting users' well-being:** Implement national classification mechanisms or build an institutional policy for categorizing and validating GenAI systems and applications. Ensure that the GenAI systems adopted by the institution are in line with locally validated ethical frameworks and do no predictable harm to the institutions' target users, especially children and vulnerable groups.
- **Review and address the long-term impacts:** Over time, relying on GenAI tools or content in education may have profound effects on the development of human capacities such as critical thinking skills and creativity. These potential effects should be evaluated and addressed.
- **Age appropriateness:** Consider implementing minimum age restrictions for the independent use of GenAI in the institution.

### 3.3.4. Individual users

Individual users potentially include all people globally who have access to the Internet and at least one type of GenAI tool. The term 'individual users', as employed here, mainly refer to individual teachers, researchers and learners in formal educational institutions or those participating in non-formal programmes of study.

- **Awareness of terms of reference on the use of GenAI:** Upon signing or expressing consent to service agreements, users should be aware of the obligations of abiding by the ToR stipulated in the agreement and the laws or regulations behind the agreement.
- **Ethical use of GenAI applications:** Users should deploy GenAI responsibly and avoid exploiting it in ways that might damage other people's reputations and lawful rights.
- **Monitoring and reporting unlawful GenAI applications:** When discovering GenAI applications that violate one or more regulations, users should notify the governmental regulatory agencies.

## 4. Towards a policy framework for the use of generative AI in education and research

Regulating GenAI to harness the potential benefits for education and research requires the development of appropriate policies. The 2023 survey data cited above indicate that only a handful of countries have adopted specific policies or plans for the use of AI in education. The preceding section outlined a vision, the steps required and the key elements and actions that can be taken by various stakeholders. This section provides measures that can be taken to develop coherent, comprehensive policy frameworks to regulate the use of GenAI in education and research.

A starting point for this is the 2022 *AI and education: guidance for policy-makers* (UNESCO, 2022b). It proposes a comprehensive set of recommendations to guide governments in the development and implementation of sector-wide policies on AI and education with a focus on promoting quality education, social equity and inclusion. Most of the recommendations remain applicable and can be further adapted to guide the formulation of specific policies on GenAI in education. The following eight specific measures for the planning of policies on GenAI in education and research are proposed here to complement this existing guidance.

### 4.1 Promote inclusion, equity, and linguistic and cultural diversity

The critical importance of inclusion must be recognized and addressed throughout the life cycle of GenAI. More specifically, GenAI tools will not help address the fundamental challenges in education or the achievement of SDG 4 commitments unless such tools are made inclusively accessible (irrespective of gender, ethnicity, special educational needs, socio-economic status, geographic location, displacement status and so on), and if they do not by design advance equity, linguistic diversities and cultural pluralism. To achieve this, the following three policy measures are recommended:

- Identify those who do not have or cannot afford internet connectivity or data, and take

action to promote universal connectivity and digital competencies in order to reduce the barriers to equitable and inclusive access to AI applications. Establish sustainable funding mechanisms for the development and provision of AI-enabled tools for learners who have disabilities or special needs. Promote the use of GenAI to support lifelong learners of all ages, locations, and backgrounds.

- Develop criteria for the validation of GenAI systems to ensure that there is no gender bias, discrimination against marginalized groups, or hate speech embedded in data or algorithms.
- Develop and implement inclusive specifications for GenAI systems and implement institutional measures to protect linguistic and cultural diversities when deploying GenAI in education and research at scale. Relevant specifications should require providers of GenAI to include data in multiple languages, especially local or indigenous languages, in the training of GPT models to improve GenAI's ability to respond to and generate multilingual text. Specifications and institutional measures should strictly prevent AI providers from any intentional or unintentional removal of minority languages or discrimination against speakers of indigenous languages, and require providers to stop systems promoting dominant languages or cultural norms.

### 4.2 Protect human agency

As GenAI becomes increasingly sophisticated, a key danger is its potential to undermine human agency. As more individuals use GenAI to support their writing or other creative activities, they might unintentionally come to rely upon it. This can compromise the development of intellectual skills. While GenAI may be used to challenge and extend human thinking, it should not be allowed to usurp human thinking. The

protection and enhancement of human agency should always be core considerations when designing and adopting GenAI from the following seven perspectives:

- Inform learners about the types of data that GenAI may collect from them, how these data are used, and the impact it may have on their education and wider lives.
- Protect learners' intrinsic motivation to grow and learn as individuals. Reinforce human autonomy over their own approaches to research, teaching, and learning in the context of using increasingly sophisticated GenAI systems.
- Prevent the use of GenAI where it would deprive learners of opportunities to develop cognitive abilities and social skills through observations of the real world, empirical practices such as experiments, discussions with other humans, and independent logical reasoning.
- Ensure sufficient social interaction and appropriate exposure to creative output produced by humans and prevent learners becoming addicted to or dependent on GenAI.
- Use GenAI tools to minimize the pressure of homework and exams, rather than to exacerbate it.
- Consult researchers, teachers and learners about their views on GenAI and use the feedback to decide whether and how specific GenAI tools should be deployed at an institutional scale. Encourage learners, teachers and researchers to critique and question the methodologies behind the AI systems, the accuracy of the output content, and the norms or pedagogies that they may impose.
- Prevent ceding human accountability to GenAI systems when making high-stakes decisions.

### 4.3 Monitor and validate GenAI systems for education

As noted, the development and deployment of GenAI should be ethical by design. Subsequently, once the GenAI is in use, and throughout its life cycle, it needs to be carefully monitored and validated – for its ethical

risks, its pedagogical appropriateness and rigour, and its impact on students, teachers and classroom/school relationships. In this respect, the following five actions are recommended:

- Build validation mechanisms to test whether GenAI systems used in education and research are free of biases, especially gender biases, and whether they are trained on data representative of diversity (in terms of gender, disability, social and economic status, ethnic and cultural background, and geographic location).
- Address the complex issue of informed consent, particularly in contexts where children or other vulnerable learners are not capable of giving genuinely informed consent.
- Audit whether outputs of GenAI include deepfake images, fake (inaccurate or false) news, or hate speech. If the GenAI is found to be generating inappropriate content, institutions and educators should be willing and able to take swift and robust action to mitigate or eliminate the problem.
- Exercise strict ethical validation of GenAI applications before they are officially adopted in educational or research institutions (i.e. adopt an ethics-by-design approach).
- Before making decisions on institutional adoption, ensure that the GenAI applications in question do no predictable harm to students, are educationally effective and valid for the ages and abilities of the target learners, and are aligned with sound pedagogical principles (i.e. based on the relevant domains of knowledge and the expected learning outcomes and development of values).

### 4.4 Develop AI competencies including GenAI-related skills for learners

The development of AI competencies among learners is key to the safe, ethical and meaningful use of AI in education and beyond. However, according to UNESCO data, only some 15 countries had developed and implemented, or were in the process of developing, government-endorsed AI curricula in schools in early 2022 (UNESCO, 2022c). The latest developments of GenAI have further reinforced the urgent need for

everyone to achieve an appropriate level of literacy in both the human and technological dimensions of AI, understanding how it works in broad terms, as well as the specific impact of GenAI. In order to do so, the following five actions are now urgently needed:

- Commit to the provision of government-sanctioned AI curricula for school education, in technical and vocational education and training, as well as for lifelong learning. AI curricula should cover the impact of AI on our lives, including the ethical issues it raises, as well as age-appropriate understanding of algorithms and data, and skills for the proper and creative use of AI tools including GenAI applications;
- Support higher education and research institutions to enhance programmes to develop local AI talent;
- Promote gender equality in developing advanced AI competencies and create a gender-balanced pool of professionals;
- Develop intersectoral forecasts of the national and global job shifts caused by the latest GenAI automation, and enhance future-proof skills at all levels of education and lifelong learning systems based on prospective shifts in demand; and
- Provide special programmes for older workers and citizens who may need to learn new skills and adapt to new environments.

#### 4.5 Build capacity for teachers and researchers to make proper use of GenAI

According to 2023 survey data on the governmental use of AI for education (UNESCO, 2023c), only some seven countries (China, Finland, Georgia, Qatar, Spain, Thailand and Türkiye) reported that they had developed or were developing frameworks or training programmes on AI for teachers. Only the Ministry of Education of Singapore reported building an online repository centred on the use of ChatGPT in teaching and learning. This clearly shows that teachers in most countries do not have access to well-structured training on the use of AI in education, not least on the use of GenAI.

To prepare teachers for the responsible and effective use of GenAI, countries need to take the following four actions:

- Formulate or adjust guidance based on local tests to help researchers and teachers to navigate widely available GenAI tools, and steer the design of new domain-specific AI applications.
- Protect the rights of teachers and researchers and the value of their practices when using GenAI. More specifically, analyse teachers' unique roles in facilitating higher-order thinking, organizing human interaction, and fostering human values.
- Define the value orientation, knowledge and skills that teachers need in order to understand and use GenAI systems effectively and ethically. Enable teachers to create specific GenAI-based tools to facilitate learning in the classroom and in their own professional development.
- Dynamically review the competencies needed by teachers to understand and use AI for teaching, learning and for their professional development, and integrate emerging sets of values, understanding and skills on AI into the competency frameworks and programmes for training in-service and pre-service teachers.

#### 4.6 Promote plural opinions and plural expressions of ideas

As noted earlier, GenAI understands neither the prompt nor the response. Instead, its responses are based on probabilities of language patterns found in the data (from the internet) that it ingested when its model was trained. To address some of the fundamental problems of its outputs, new methods are currently being researched such as connecting GenAI with knowledge databases and reasoning engines. Nonetheless, because of how it works, its source materials and the tacit perspectives of its developers, GenAI, by definition, reproduces dominant worldviews in its outputs and undermines minority and plural opinions. Accordingly, if human civilizations are to flourish, it is essential that we recognize that GenAI can never be an authoritative source of knowledge on whatever topic it engages with.

As a result, users need to view GenAI's outputs critically. In particular:

- Understand the role of GenAI as a fast but frequently unreliable source of information. While some plugins and LLM-based tools mentioned earlier are designed to support the need to access validated and up-to-date information, there is little robust evidence as yet that these are effective.
- Encourage learners and researchers to critique the responses provided by GenAI. Recognize that GenAI typically only repeats established or standard opinions, thus undermining plural and minority opinions and plural expressions of ideas.
- Provide learners with sufficient opportunities to learn from trial and error, empirical experiments, and observations of the real world.

#### 4.7 Test locally relevant application models and build a cumulative evidence base

GenAI models are thus far dominated by information from the Global North and under-representing voices from the Global South and indigenous communities. Only by means of determined efforts, for example harnessing synthetic data (Marwala, 2023), will GenAI tools be made sensitive to the context and needs of local communities, particularly those from the Global South. To explore approaches relevant to local needs, while collaborating more widely, the following eight actions are recommended:

- Ensure the design and adoption of GenAI are strategically planned rather than facilitating a passive and non-critical procurement process.
- Incentivize the designers of GenAI to target open-ended, exploratory and diverse learning options.
- Test and scale up evidence-based use cases of applying AI in education and research in accordance with educational priorities, rather than novelty, myth or hype.
- Guide the use of GenAI to trigger innovation in research, including through leveraging computing capabilities, large-scale data, and GenAI outputs to inform and inspire the improvement of research methodologies.
- Review the social and ethical implications of incorporating GenAI into research processes.
- Establish specific criteria based on evidenced pedagogical research and methodologies and build an evidence base for the effectiveness of GenAI in terms of supporting the provision of inclusive learning opportunities, meeting learning and research objectives, and promoting linguistic and cultural diversities.
- Take iterative steps to strengthen evidence on the social and ethical impact of GenAI.
- Analyse the environmental costs of leveraging AI technologies at scale (e.g. the energy and resources required for training GPT models), and develop sustainable targets to be met by AI providers in a bid to avoid adding to climate change.

#### 4.8 Review long-term implications in an intersectoral and interdisciplinary manner

Intersectoral and interdisciplinary approaches are essential for the effective and ethical use of GenAI in education and research. Only by drawing on a range of expertise, while bringing together multiple stakeholders, will key challenges be identified promptly and addressed effectively to minimize long-term negative implications while leveraging ongoing and cumulative benefits. Therefore, these three actions are recommended:

- Collaborate with AI providers, educators, researchers, and representatives of parents and students to plan system-wide adjustments in curriculum frameworks and assessment methodologies, to fully leverage the potential and mitigate the risks of GenAI for education and research.
- Bring together intersectoral and interdisciplinary expertise including educators, researchers, learning scientists, AI engineers, and representatives of other stakeholders to examine the long-term implications of GenAI for learning and knowledge production, research and copyright, curriculum and assessment, and human collaboration and social dynamics.
- Provide timely advice to inform the iterative updates of regulations and policies.

## 5. Facilitating creative use of GenAI in education and research

When ChatGPT was first launched, educators across the world expressed their concerns about its potential to generate essays and how it might help students to cheat. More recently, many people and organizations including some of the world's leading universities have argued that 'the genie is out of the bottle' and tools like ChatGPT are here to stay and may be used productively in educational settings. Meanwhile, the internet is now awash with suggestions for the use of GenAI in education and research. These include using it to inspire new ideas, generate multi-perspective examples, develop lesson plans and presentations, summarize existing materials, and stimulate image creation. Although new ideas appear on the internet almost every day, researchers and educators are still working out exactly what GenAI means for teaching, learning and research. In particular, the people behind many of the proposed uses may not have properly considered ethical principles, while others are driven by the technological potentials of GenAI rather than the needs of researchers, teachers or learners. This section outlines ways in which the creative use of GenAI in education can be facilitated.

### 5.1 Institutional strategies to facilitate responsible and creative use of GenAI

As stated earlier, educational and research institutions should develop, implement and validate appropriate strategies and ethical frameworks to guide the responsible and ethical use of GenAI systems and applications to meet the needs of teaching, learning and research. This can be achieved through the following four strategies:

- Institutional implementation of ethical principles:** Ensure that researchers, teachers and learners use GenAI tools responsibly and ethically, and critically approach the accuracy and validity of the outputs.
- Guidance and training:** Provide guidance and training to researchers, teachers and learners about GenAI tools to ensure that they understand the ethical issues such as biases in data labelling and algorithms, and that they comply with the appropriate regulations on data privacy and intellectual property.
- Building GenAI prompt-engineering capacities:** In addition to subject-specific knowledge, researchers and teachers will also need expertise in engineering and critically evaluating the prompts generated by GenAI. Given that the challenges raised by GenAI are complex, researchers and teachers must receive high-quality training and support to do this.
- Detecting GenAI-based plagiarism in written assignments:** GenAI might allow students to pass off text that they did not write as their own work, a new type of 'plagiarism'. GenAI providers are required to label their outputs with 'generated by AI' watermarks, while tools are being developed to identify material that has been produced by AI. However, there is little evidence that these measures or tools are effective. The immediate institutional strategy is to uphold academic integrity and reinforce accountability through rigorous detection by humans. The long-term strategy is for institutions and educators to rethink the design of written assignments so that they are not used to assess tasks that GenAI tools can do better than human learners. Instead, they should address what humans can do that GenAI and other AI tools cannot do, including applying human values such as compassion and creativity to complex real-world challenges.

## 5.2 A ‘human-centred and pedagogically appropriate interaction’ approach

Researchers and educators should prioritize human agency and responsible, pedagogically appropriate interaction between humans and AI tools when deciding on whether and how to use GenAI. This includes the following five considerations:

- the use of the tool(s) should contribute to humans’ needs and make learning or research more effective than a no-tech or other alternative approach;
- educators’ and learners’ use of the tool(s) should be based on their intrinsic motivation;
- the process of using the tool(s) should be controlled by the human educators, learners or researchers;
- the choice and organization of the tool(s) and the content they generate should be proportionate, based on the learners’ age range, the expected results, and the type of target knowledge (e.g. factual, conceptual, procedural, or metacognitive) or target problem (e.g. well-structured or ill-structured); and
- the usage processes should ensure humans’ interactive engagement with GenAI and higher-order thinking, as well as human accountability for decisions related to the accuracy of AI-generated content, teaching or research strategies, and their impact on human behaviours.

## 5.3 Co-designing the use of GenAI in education and research

The use of GenAI in education and research should be neither imposed in a top-down approach nor driven by commercial hyperbole. Instead, its safe and effective use should be co-designed by teachers, learners, and researchers. It also needs a robust process of piloting and evaluation to examine the effectiveness and the long-term impact of different uses.

To facilitate the recommended co-design, this Guidance proposes a framework composed of the following six perspectives to consolidate pedagogically appropriate interactions and the prioritization of human agency:

- appropriate domains of knowledge or problems;
- expected outcomes;
- appropriate GenAI tools and comparative advantages;
- requirements for users;
- required human pedagogical methods and example prompts; and
- ethical risks.

This section provides examples of how a process of co-design in the use of GenAI can inform research practices, assist in teaching, provide coaching for the self-paced acquisition of foundational skills, facilitate higher-order thinking, and support learners with special needs. These examples represent only the tip of the iceberg of the increasing number of domains in which GenAI may have potential.

### 5.3.1 Generative AI for research

GenAI models have demonstrated their potential to expand views on research outlines and to enrich data exploration as well as literature reviews (see **Table 3**). While a wider range of use cases may emerge, novel research is needed to define the potential domain of research problems and expected outcomes, to demonstrate the efficacy and accuracy, and to ensure that human agency in understanding the real world through research will not be undermined by the use of AI tools.



**Table 3. Co-designing uses of GenAI for research**

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
AI advisor for research outlines	Might be useful in well-structured domains of research problems.	Developing and answering research questions, suggesting appropriate methodologies.  <b>Potential transformation: 1:1 coach for research planning</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open-source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The researcher must have a basic understanding of the topic(s).  The researcher should develop the ability to verify the information, and be especially capable of detecting citations of non-existent research papers.	Basic ideas for the definition of research problems (e.g. target audience, issues, context), as well as methodologies, expected outcomes and formats.  <b>Example prompt:</b> <i>Write 10 potential research questions for [topic x] and rank them in importance for [the field of research y].</i>	Need to be alert to the high risk of GenAI making up information (such as non-existent research publications), and of users being tempted to copy and paste AI-generated research outlines, which may reduce junior researchers' opportunities to learn from trial and error.
Generative data explorer and literature reviewer	Might be useful in ill-structured domains of research problems.	Automatic gathering of information, exploration of a wide range of data, proposing drafts of literature reviews, and automating parts of data interpretation.  <b>Potential transformation: AI trainers for data exploration and literature reviews</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The researchers must have a robust knowledge of methodologies and techniques for analysing data.	Progressive definitions of the problems, the scope of data and sources of literature, the methodologies used for data exploration and literature reviews, and the expected outcomes and their formats.	Need to beware of GenAI-fabricated information, the improper handling of data, possible breaches of privacy, unauthorized profiling, and gender bias.  Need to be alert to the propagation of dominant norms and their threat to alternative norms and plural opinions.

### 5.3.2 Generative AI to facilitate teaching

The use of both general GenAI platforms and specific educational GenAI tools should be designed to enhance teachers' understanding of their subject as well as their knowledge on teaching methodologies, including through teacher-AI co-designing of lesson plans, course packages, or entire curricula. The GenAI-assisted conversational teachers' assistants or 'generative twins of teaching assistants'<sup>53</sup> that are

pre-trained based on data from experienced teachers and libraries, have been tested in some educational institutions and may hold unknown potential as well as uncharted ethical risk. The practical application processes and further iterations of these models still need to be carefully audited through the framework recommended in this Guidance and safeguarded by human supervision as exemplified in **Table 4**.

**Table 4. Co-designing uses of GenAI to support teachers and teaching**

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
Curriculum or course co-designer	Conceptual knowledge on certain teaching topics and procedural knowledge on teaching methodologies.	Assisting with the curriculum and lesson design process, including outlining or extending views on key areas of the target topic and defining the curriculum structure. It may also help teachers prepare tests and exams by offering examples of questions and rubrics for evaluation.  <b>Potential transformation: AI-generated curriculum</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The teachers must understand and carefully specify what they want the curriculum, courses, lessons, or tests to cover and achieve, whether they want to address procedural or conceptual knowledge, and what teaching theory they wish to apply.	Questions to GenAI on suggesting the structure and examples of factual knowledge on topic(s), suggesting teaching methods and processes for topics or problems, or creating course packages or lesson plans based on topic(s) and formatting.  Human curriculum designers need to verify the factual knowledge and check the appropriateness of the suggested course packages.	The risk of GenAI imposing dominant norms and pedagogical methods is high.  It may inadvertently perpetuate exclusionary practices in favour of the already data-rich groups and reinforce inequalities in access to relevant and high-quality educational opportunities, disadvantaging data-poor groups.
Generative chatbot as teaching assistant	Conceptual knowledge across multiple domains in well-structured problems.	Providing individualized support, answering questions and identifying resources.  <b>Potential transformation: Generative twins of teachers' assistants</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	It supports teachers but targets learners directly, so this requires learners to have sufficient prior knowledge, abilities and metacognitive skills to the verify the outputs of GenAI and notice the misinformation. Thus it might be more appropriate for learners in higher education.	Requires the teachers to understand the problems clearly, to monitor the conversation and help learners to verify dubious answers provided by GenAI.	Based on the current capabilities of GenAI models, educational institutions need to guarantee human supervision of the responses provided by GenAI tools, being alert to the risk of misinformation.  It may also limit learners' access to human guidance and support, hindering the development of a strong teacher-student relationship, which is especially concerning for children.

### 5.3.3 Generative AI as a 1:1 coach for the self-paced acquisition of foundational skills

While higher-order thinking and creativity have been drawing increasing attention when defining learning outcomes, there is still no doubting the importance of foundational skills in children's psychological development and competency progression. Among a large spectrum of abilities, these foundational skills include listening, pronouncing, and writing a

mother tongue or foreign language, as well as basic numeracy, art, and coding. 'Drill and practice' should not be considered as an obsolete pedagogical method; instead, it should be reinvigorated and upgraded with GenAI technologies to foster learners' self-paced rehearsal of foundational skills. If guided by ethical and pedagogical principles, GenAI tools have the potential to become 1:1 coaches for such self-paced practice, as illustrated in **Table 5**.

**Table 5. Co-designing uses of GenAI as a 1:1 coach for the self-paced acquisition of foundational skills in languages and the arts**

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
1:1 language skills coach	Language learning, including conversational practice.	Engaging learners in conversational practice to help them improve listening, speaking and writing skills by offering feedback, corrections and modelling of the mother tongue or foreign language. Helping learners improve their writing skills.  <b>Potential transformation:</b> <b>1:1 language tutorials at beginner level</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	An age limit may be set for the independent conversations in view of the culturally insensitive or age-inappropriate output provided by GenAI systems.  The learner must have the initial intrinsic motivation to engage in a conversation with an AI system.  The learner should be able to take a critical approach to the GenAI's suggestions and check whether they are accurate.	When using general GenAI platforms, human teachers can guide learners to engage with GenAI tools to request feedback for improvement, correction of pronunciation or examples of writing.  For instance: <i>Engage me in a conversation in the [x] language, helping me to continuously improve.</i>  <i>Suggest some ideas to help me write about [topic x].</i>	Need to be alert to culturally insensitive or contextually inaccurate language, and the inadvertent perpetuation of stereotypes or cultural biases.  Without proper pedagogical strategies to simulate learners' intrinsic motivations, it may limit children's creativity and originality, leading to formulaic writing.  It may also limit opportunities for real-life interactions, plural opinions, plural expression, and critical thinking.
1:1 art coach	Technical skills in areas of art such as music and drawing.	Providing suggestions for art techniques (e.g. tips on perspective and colour), or musical composition (e.g. melody and chord progression).  <b>Potential transformation:</b> <b>1:1 art teacher at introductory levels</b>	Starting with the list in Section 1.2, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	Learners must have some initial aims for creating art or music, a foundational understanding of the key elements of the domain of art or music, and basic abilities to analyse the artworks or musical compositions.	Human teachers should ask learners to compare AI tools' art techniques with their own artwork. Human teachers or coaches must encourage learners to develop and apply their imagination and creativity, which GenAI cannot replace.  Example prompt: <i>Suggest some ideas to inspire me to create an image on [topics/ideas].</i>	May expose children to inappropriate or offensive content, which may violate their right to safeguarding and well-being.  GenAI tools raise the risk of stopping learners from developing their imagination and creativity.
1:1 coach for coding or arithmetic	Conceptual programming knowledge and skills at the introductory level. It might also apply to the learning of basic mathematics.	Supporting self-paced learning of basic coding knowledge and skills, finding bugs in learners' coding and providing immediate feedback, and tailoring answers to questions.  <b>Potential transformation:</b> <b>1:1 coding teacher at introductory level</b>	Starting with the list in Section 1.3, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	Finding and defining a problem, and designing algorithms to solve the problem, remain the core aspects of learning coding and programming. Learners must have intrinsic motivation to use the coding, along with some basic knowledge and skills in using the programming language.	Human teachers and coaches should teach basic knowledge and skills, and inspire learners to use computational thinking and programming to solve problems including through collaborative coding.  Example prompt: <i>Suggest some unusual ideas for coding.</i>	The accuracy of feedback and suggestions remains a problematic issue as GenAI will not always be right.  There is a high risk that GenAI tools will prevent learners from developing computational thinking skills and abilities to find and define meaningful problems for coding.

### 5.3.4 Generative AI to facilitate inquiry or project-based learning

If not used purposefully to facilitate higher-order thinking or creativity, GenAI tools tend to encourage plagiarism or shallow ‘stochastic parroting’ outputs. However, given that GenAI models have been trained

based on large-scale data, they have potential for acting as an opponent in Socratic dialogues or as a research assistant in project-based learning. Yet these potentials can only be leveraged through instructional/ learning design processes that aim to trigger higher-order thinking as exemplified in **Table 6**.

**Table 6. Co-designing uses of GenAI to facilitate inquiry or project-based learning**

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
Socratic challenger	Ill-structured problems.	Engage learners in dialogue reminiscent of the Socratic questioning of prior knowledge, leading to the discovery of new knowledge or deeper understanding.  <b>Potential transformation:</b>  <b>1:1 Socratic opponent</b>	Starting with the list in Section 1.3, assess whether specific GenAI tools are locally accessible, open-source, rigorously tested and validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The learner must have reached the age that allows them to conduct independent conversations with GenAI tools. Learners must have prior knowledge and abilities to check whether the arguments and information presented are accurate.	Human teachers may help prepare a list of gradually deeper questions as examples for learners to adapt into prompts. Learners may also start with a broad prompt such as <i>‘Engage me in a Socratic dialogue in order to help me take a critical perspective towards [topic x]’</i> and then gradually deepen the dialogue through increasingly refined prompts.	The current GenAI tools may generate similar or standard answers that limit learners’ exposure to diverse viewpoints and alternative perspectives, leading to an echo-chamber effect, and hinder the development of independent thinking.
Advisor for project-based learning	Ill-structured research problems in science or social studies.	Support knowledge creation through helping learners to conduct project-based learning. This includes GenAI playing a role that is similar to the research advisor described in Table 3.  <b>Potential transformation:</b>  <b>1:1 project-based learning coach</b>	Starting with the list in Section 1.3, assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	Learners could act as junior researchers in planning and implementing project-based learning. The learners must be old enough for the independent use of GenAI platforms. Learners must have the motivation and ability to engage in self-directed project-based learning activities, so that they are not tempted to passively copy and paste the answers provided by GenAI tools.	Human teachers guide learners to ask GenAI to provide basic ideas for the definition of research problems as suggested in 5.3.1. Individual and group learners use GenAI tools to conduct literature reviews, collect and process data, and create reports.	Learners without the solid prior knowledge and the ability necessary to verify the accuracy of answers may be misled by the information that GenAI tools provide. It may also limit learners’ discussions and interactions with peers and reduce opportunities for collaborative learning, potentially harming their social development.

### 5.3.5 Generative AI to support learners with special needs

Theoretically, GenAI models have the potential to help learners with hearing or visual impairments. The emerging practices include GenAI-enabled subtitles or captions for deaf and hard-of-hearing learners, and GenAI-generated audio description for visually impaired learners. GenAI models can also convert text to speech and speech to text to enable people with visual, hearing, or speech impairments to access content, ask questions, and communicate with their peers. However, this function has not yet been leveraged at scale. According to the survey mentioned earlier, conducted by UNESCO in 2023 on governments’ use of AI in education, only four countries (China, Jordan, Malaysia and Qatar) reported that their governmental agencies had validated and recommended AI-assisted tools to support inclusive access for learners who have disabilities (UNESCO, 2023c).

There is also a trend toward iterations of GenAI models being trained to support learners to use their own languages, including minority and indigenous languages, to learn and communicate. For example,

PaLM 2, Google’s next-generation LLM, is trained on parallel data covering hundreds of languages in the form of source and target text pairs. The inclusion of parallel multilingual data is designed to further improve the model’s ability to understand and generate multilingual text (Google, 2023b).

By providing real-time translations, paraphrasing, and automatic correction, GenAI tools have the potential to help learners who use minority languages to communicate ideas and enhance their collaboration with peers from different linguistic backgrounds. However, this will not happen naturally at scale. Only with purposeful design can this potential be leveraged to amplify the voices of marginalized groups.

Finally, it has also been suggested that GenAI systems have the potential to carry out conversation-based diagnoses, identifying psychological or social-emotional problems as well as learning difficulties. However, there remains little evidence that this approach is either effective or safe, and any diagnoses would require interpretation by skilled professionals.

**Table 7. Co-designing uses of GenAI to support learners with special needs**

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
Conversational diagnosis of learning difficulties	This might be helpful for learners who are facing learning difficulties caused by psychological, social or emotional problems.	Using natural-language engagement to identify the needs of learners who have psychological, social or emotional problems or learning difficulties, in order to provide them with relevant support or instruction.  <b>Potential transformation:</b> <b>1:1 primary advisor for learners with social or emotional problems or learning difficulties</b>	In addition to general GenAI tools, search for chatbots powered by GenAI.  Assess whether they are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	Teachers or specialists who work with this group of learners will need to ensure that the primary advice suggested by the GenAI system is accurate.	Teachers or facilitators need to provide comfortable environments to engage the learner in a conversation in order to diagnose psychological, social, or emotional problems, or learning difficulties.	May inadvertently misdiagnose the learner’s specific challenges, leading to the wrong support being provided.

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methods and example prompts	Possible risks
AI-powered accessibility tools	These enable learners with hearing or visual impairment to access a wider range of content, thus improving the quality of their learning.	Meeting learners' access needs and supporting their acquisition of subject-specific knowledge by providing GenAI-enabled captioning and/or sign language interpretation for audio or video content, and audio descriptions for text or other visual material.  <b>Potential transformation:</b>  <b>1:1 personalized AI-powered language aids</b>	In addition to general GenAI tools, search for relevant and trusted AI-powered generators of captions and audio descriptions.  Assess whether they are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The educators or facilitators must help learners access and learn how to operate the GenAI tools. They also need to ensure that the tools' outputs genuinely support these learners and do not reinforce the challenges and biases that they face.	Need to test the accessibility of platforms or tools to identify and fix accessibility issues before they are used. GenAI tools can only provide access to content, so educators and facilitators should focus on enhancing their quality of learning and social well-being.  Educators and facilitators need to teach the learners to create voice or text prompts based on their abilities.	The captions or audio descriptions produced by GenAI platforms that are not designed specifically to support vision or hearing are often inaccurate and may mislead learners with special needs.  These tools may inadvertently reinforce existing biases.
Generative amplifier for marginalized learners	It might be helpful for learners from minority linguistic or cultural backgrounds to express and amplify their voices, to participate online, and to conduct collaborative social studies.	Providing real-time translations, paraphrasing, and automatic correction of writing to support learners from marginalized groups to use their own languages to communicate with peers from different linguistic backgrounds.  <b>Potential transformation:</b>  <b>Inclusive LLMs for marginalized learners</b>	A specific example for consideration is PaLM 2.  Assess whether the GenAI tools are locally accessible, open source, rigorously tested or validated by authorities.  Further consider the advantages and challenges of any particular GenAI tool, and ensure that it properly addresses specific human needs.	The learners should have knowledge or meaningful opinions on the topic of the conversation or collaborative study. They need to be capable of making responsible and non-discriminatory contributions and avoiding hate speech.	Teachers or educators should design studies and writing tasks for learners on social or cultural topics, or organize online seminars or intercultural collaborations to stimulate learners to generate ideas and share opinions.	Need to identify and correct the errors in AI translations and paraphrasing that may cause intercultural misunderstandings.  This use can provide opportunities for marginalized learners to amplify their voices, but will not touch the root cause of data poverty and therefore cannot decolonize AI tools.

## 6. GenAI and the future of education and research

GenAI technologies are still rapidly evolving and likely to have a profound impact on education and research, and are yet to be fully understood. Therefore, their potential long-term implications for education and research need immediate attention and further in-depth review.

### 6.1 Uncharted ethical issues

The increasingly sophisticated GenAI tools will raise additional ethical concerns that need to be examined in detail. Further to Sections 2 and 3, deeper and more forward-looking analyses are needed to reveal and address uncharted ethical issues from at least the following five perspectives:

- **Access and equity:** GenAI systems in education may exacerbate existing disparities in access to technology and educational resources, further deepening inequities.
- **Human connection:** GenAI systems in education may reduce human-to-human interaction and the critical social-emotional aspects of learning.
- **Human intellectual development:** GenAI systems in education may limit learners' autonomy and agency by providing predetermined solutions or narrowing the range of possible learning experiences. Their long-term impact on young learners' intellectual development needs to be investigated.
- **Psychological impact:** GenAI systems that mimic human interactions may have unknown psychological effects on learners, raising concerns about their cognitive development and emotional well-being, and about the potential for manipulation.
- **Hidden bias and discrimination:** As more sophisticated GenAI systems are being developed and applied in education, they are likely to generate new biases and forms of discrimination based on the training data and methods used by the models, which can result in unknown and potentially harmful outputs.

### 6.2 Copyright and intellectual property

The emergence of GenAI is rapidly changing the way in which scientific, artistic and literary works are created, distributed and consumed. Unauthorized copying, distribution or use of copyrighted works without permission from the copyright holder violates their exclusive rights and can lead to legal consequences. For example, the training of GenAI models has been accused of infringing copyright. In one of the recent cases, the AI-generated song featuring 'Drake' and 'The Weeknd' (Abel Tesfaye) reached millions of listeners before being taken offline due to a copyright dispute (Coscarelli, 2023). While the emerging regulatory frameworks intend to require GenAI providers to recognize and protect the intellectual property of the owners of the content used by the model, it is becoming increasingly challenging to determine the ownership and originality of the overwhelming amount of generated works. This lack of traceability not only raises concerns about protecting the rights of creators and ensuring fair compensation for their intellectual contributions, but also introduces challenges into educational contexts about how the output of GenAI tools may responsibly be used. This may have profound implications for the research system.

### 6.3 Sources of content and learning

GenAI tools are changing the way teaching and learning content can be generated and provided. In the future, content generated through human-AI conversations may become one of the main sources of knowledge production. This is likely to further undermine learners' direct engagement with educational content based on resources, textbooks and curricula created and validated by humans. The authoritative appearance of GenAI text may mislead young learners who do not have sufficient prior knowledge to be able to recognize inaccuracies or to question it effectively. Whether learners' engagement with unvalidated content should be recognized as 'learning' is also contestable.

The resultant concentration on aggregated second-hand information may also reduce learners' opportunities for constructing knowledge through proven methods such as directly perceiving and experiencing the real world, learning from trial and error, performing empirical experiments, and developing common sense. It may also threaten the social construction of knowledge and the fostering of social values through collaborative classroom practices.

## 6.4 Homogenized responses versus diverse and creative outputs

GenAI narrows plural narratives as the outputs generated tend to represent and reinforce dominant viewpoints. The resulting homogenization of knowledge limits pluralistic and creative thinking. The increased dependency of teachers and students on GenAI tools to seek suggestions may lead to the standardization and conformity of responses, weakening the value of independent thought and self-directed inquiry. The potential homogenization of expression in written pieces and artwork can limit learners' imagination, creativity and alternative perspectives of expressions.

GenAI providers and educators need to consider the extent to which EdGPT might be developed and used to foster creativity, collaboration, critical thinking and other higher-order thinking skills.

## 6.5 Rethinking assessment and learning outcomes

The implications of GenAI for assessment go far beyond the immediate concerns about learners cheating on written assignments. We must contend with the fact that GenAI can produce relatively well-organized papers and essays and impressive works of art, and can pass some knowledge-based exams in certain subject areas. We therefore need to rethink what exactly should be learned and to what ends, and how learning is to be assessed and validated.

Critical discussion by educators, policy-makers, learners and other stakeholders need to consider the following four categories of learning outcomes:

**Values:** The values required to ensure the human-centred design and use of technology are central to the rethinking of learning outcomes and their assessment in the digital era. In revisiting the purpose of education, the values that inform the way in which technology relates to education should be made explicit. It is through this normative lens that learning outcomes and their assessment and validation need to be iteratively updated to respond to the increasingly pervasive use of technology, including AI, in society.

**Foundational knowledge and skills:** Even in the domains of competencies where GenAI tools can do better than humans, learners will still need sound foundational knowledge and skills. Foundational literacy, numeracy and basic scientific literacy skills will remain key for education in the future. The scope and nature of these foundational skills will need to be regularly revisited to reflect the increasingly AI-rich environments we live in.

**Higher-order thinking skills:** Learning outcomes will need to include skills required to support higher-order thinking and problem solving based on human-AI collaboration and the use of GenAI-generated outputs. These may include understanding the roles of factual and conceptual knowledge in grounding higher-order thinking, and the critical evaluation of AI-generated content.

**Vocational skills needed to work with AI:** In the domains where AI can do better than humans and is automating task units, human learners need to nurture new skills that enable them to develop, operate and work with GenAI tools. The redesign of learning outcomes and educational assessment will need to reflect the vocational skills required for the new jobs created by AI.

## 6.6 Thinking processes

The most fundamental perspective of the long-term implications of GenAI for education and research is still about the complementary relationship between human agency and machines. One of the key questions is whether humans can possibly cede basic levels of thinking and skill-acquisition processes to AI and rather concentrate on higher-order thinking skills based on the outputs provided by AI.



Writing, for example, is often associated with the structuring of thinking. With GenAI, rather than starting from scratch to plan the aims, scope and outline of a set of ideas, humans can now start with a well-structured outline provided by GenAI. Some experts have characterized the use of GenAI to generate text in this way as ‘writing without thinking’ (Chayka, 2023). As these new GenAI-assisted practices become more widely adopted, established methods for the acquisition and assessment of writing skills will need to adapt. One option in the future is that the learning of writing may focus on building skills in planning and composing prompts, critical evaluation of the GenAI outputs, and higher-order thinking, as well as on co-writing based on GenAI’s outlines.

## Concluding remarks

From the perspective of a human-centred approach, AI tools should be designed to extend or augment human intellectual abilities and social skills, and not undermine them, conflict with them or usurp them. It has long been expected that AI tools can be further integrated as part and parcel of the tools available to humans

to support analysis and action for more inclusive and sustainable futures.

For AI to be a trustable part and parcel of human-machine collaboration – at individual, institutional and system levels – the human-centred approach informed by the 2021 UNESCO *Recommendation on the Ethics of AI* is to be further specified and implemented according to the specific characteristics of emerging technologies such as GenAI. Only in this way can we ensure that GenAI becomes a trustworthy tool for researchers, teachers and learners.

While GenAI should be used to serve education and research, we all need to be cognizant that GenAI might also change the established systems and their foundations in these domains. The transformation of education and research to be triggered by GenAI, if any, should be rigorously reviewed and steered by a human-centred approach. Only by doing so can we ensure that the potentials of AI in particular, and all other categories of technologies used in education more broadly, enhance human capabilities to build inclusive digital futures for all.

## References

- Anders, B. A. 2023. *Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?* Cambridge, Cell Press. Available at: <https://doi.org/10.1016/j.patter.2023.100694> (Accessed 23 June 2023.)
- Bass, D. and Metz, R. 2023. *OpenAI's Sam Altman Urges Congress to Regulate Powerful New Technology*. New York, Bloomberg. Available at: <https://www.bloomberg.com/news/newsletters/2023-05-17/openai-s-sam-altman-urges-congress-to-regulate-powerful-new-ai-technology> (Accessed 23 June 2023.)
- Bender, E. M., Gebru, T., McMillan-Major, A. and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery. Available at: <https://doi.org/10.1145/3442188.3445922> (Accessed 23 June 2023.)
- Bommasani, R. et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, Stanford University. Available at: <https://crfm.stanford.edu/report.html> (Accessed 23 June 2023.)
- Bove, T. 2023. *Big tech is making big AI promises in earnings calls as ChatGPT disrupts the industry: 'You're going to see a lot from us in the coming few months'*. New York, Fortune. Available at: <https://fortune.com/2023/02/03/google-meta-apple-ai-promises-chatgpt-earnings> (Accessed 3 July 2023.)
- Chayka, K. 2023. *My A.I. Writing Report*. New York, The New Yorker. Available at: <https://www.newyorker.com/culture/infinite-scroll/my-ai-writing-robot> (Accessed 1 August 2023.)
- Chen, L., Zaharia, M., and Zou, J. 2023. *How Is ChatGPT's Behavior Changing over Time?* Ithaca, arXiv. Available at: <https://arxiv.org/pdf/2307.09009> (Accessed 31 July 2023.)
- Coscarelli, J. 2023. *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*. New York, New York Times. Available at: <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html> (Accessed 30 August 2023.)
- Cyberspace Administration of China. 2023a. 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](http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm) (Accessed 19 July 2023.)
- . 2023b. 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](http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm) (Accessed 19 July 2023.)
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Horn Nord, J., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst,

- W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J. and Wright, R. 2023. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, Vol. 71. Amsterdam, Elsevier, p. 102642. Available at: <https://doi.org/10.1016/j.ijinfomgt.2023.102642> (Accessed 25 August 2023.)
- E2Analyst. 2023. *GPT-4: Everything you want to know about OpenAI's new AI model*. San Francisco, Medium. Available at: <https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495> (Accessed 1 August 2023.)
- European Commission. 2021. *Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. Brussels, European Commission. Available at: <https://artificialintelligenceact.eu> (Accessed 23 June 2023.)
- European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Brussels, Official Journal of the European Union. Available at: <http://data.europa.eu/eli/reg/2016/679/oj> (Accessed 23 June 2023.)
- Federal Trade Commission. 1998. *Children's Online Privacy Protection Act of 1998*. Washington DC, Federal Trade Commission. Available at: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa> (Accessed 4 September 2023.)
- Giannini, S. 2023. *Generative AI and the Future of Education*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000385877> (Accessed 29 August 2023.)
- Google. 2023a. *Recommendations for Regulating AI*. Mountain View, Google. Available at: <https://ai.google/static/documents/recommendations-for-regulating-ai.pdf> (Accessed 23 June 2023.)
- . 2023b. *PaLM 2 Technical Report*. Mountain View, Google. Available at: <https://doi.org/10.48550/arXiv.2305.10403> (Accessed on 20 July 2023.)
- Lin, B. 2023. *AI Is Generating Security Risks Faster Than Companies Can Keep Up*. New York, The Wall Street Journal. Available at: <https://www.wsj.com/articles/ai-is-generating-security-risks-faster-than-companies-can-keep-up-a2bdedd4> (Accessed 25 August 2023.)
- Marcus, G. 2022. Hoping for the Best as AI Evolves. *Communications of the ACM*, Vol. 66, No. 4. New York, Association for Computing Machinery. Available at: <https://doi.org/10.1145/3583078> (Accessed 23 June 2023.)
- Marwala, T. 2023. *Algorithm Bias — Synthetic Data Should Be Option of Last Resort When Training AI Systems*. Tokyo, United Nation University. Available at: <https://unu.edu/article/algorithm-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems> (Accessed 31 July 2023.)
- Metz, C. 2021. *Who Is Making Sure the A.I. Machines Aren't Racist?* New York, The New York Times. Available at: <https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html> (Accessed 23 June 2023.)
- Murphy Kelly, S. 2023. *Microsoft is bringing ChatGPT technology to Word, Excel and Outlook*. Atlanta, CNN. Available at: <https://edition.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html> (Accessed 25 August 2023.)
- Nazaretsky, T., Cukurova, M. and Alexandron, G. 2022a. 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. 2022b. 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> (Accessed 1 August 2023.)
- Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Singapore, OpenGov Asia. Available at: <https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster> (Accessed 25 August 2023.)

- OpenAI. 2018. *AI and compute*. San Francisco, OpenAI. Available at: <https://openai.com/research/ai-and-compute> (Accessed 23 June 2023.)
- . 2023. *Educator considerations for ChatGPT*. San Francisco, OpenAI. Available at: <https://platform.openai.com/docs/chatgpt-education> (Accessed 23 June 2023.)
- Popli, N. 2023. *The AI Job That Pays Up to \$335K—and You Don't Need a Computer Engineering Background*. New York, TIME USA. Available at: <https://time.com/6272103/ai-prompt-engineer-job> (Accessed 23 June 2023.)
- Roose, K. 2022. *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*. New York, The New York Times. Available at: <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (Accessed 23 June 2023.)
- Russell Group, 2023. *Russell Group principles on the use of generative AI tools in education*. Cambridge, Russell Group. Available at: [https://russellgroup.ac.uk/media/6137/rg\\_ai\\_principles-final.pdf](https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf) (Accessed 25 August 2023.)
- Stanford University. 2019. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: <https://hai.stanford.edu/ai-index-2019> (Accessed 23 June 2023.)
- . 2023. *Artificial Intelligence Index Report*. Stanford, Stanford University. Available at: <https://hai.stanford.edu/research/ai-index-2023> (Accessed 23 June 2023.)
- The Verge. 2023a. *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> (Accessed 1 August 2023.)
- . 2023b. *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> (Accessed 1 August 2023.)
- Tlili, A., Shehata, B., Agyemang Adarkwah, M., Bozkurt, A., Hickey, D. T., Huang, R. and Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, Vol. 10, No. 15. Berlin, Springer. Available at: <https://doi.org/10.1186/s40561-023-00237-x> (Accessed 23 June 2023.)
- UNESCO. 2019. *Beijing Consensus on Artificial Intelligence and Education*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (Accessed 3 July 2023.)
- . 2022a. *Recommendation on the Ethics of Artificial Intelligence*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (Accessed 3 July 2023.)
- . 2022b. *AI and education: guidance for policy-makers*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000376709> (Accessed 23 June 2023.)
- . 2022c. *K-12 AI curricula: a mapping of government-endorsed AI curricula*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000380602> (Accessed 20 July 2023.)
- . 2022d. *Guidelines for ICT in education policies and masterplans*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000380926> (Accessed 31 July 2023.)
- . 2023a. *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> (Accessed 3 July 2023.)
- . 2023b. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. Unpublished.
- . 2023c. *Survey for the governmental use of AI as a public good for education*. Unpublished (Submitted to UNESCO).

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

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

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

# Endnotes

- 1 GenAI models became available to researchers and other interested parties far earlier than ChatGPT. For example, in 2015 Google released what they called 'DeepDream' (<https://en.wikipedia.org/wiki/DeepDream>).
- 2 See <https://chat.openai.com>
- 3 For an explanation of AI techniques and technologies and their relationship, see UNESCO, 2022b, pp. 8-10.
- 4 Note that, because GenAI is still relatively new, different companies often use these terms in different ways, and sometimes use different words to mean the same thing.
- 5 There is concern that the data used to train future iterations of OpenAI GPT will include substantial amounts of text generated by previous versions of GPT. This self-referential loop might contaminate the training data and thus compromise the capabilities of future GPT models.
- 6 NB OpenAI, the company that developed the GPTs in this table, has not publicly released detailed information about GPT-4 (The Verge, 2023a). In fact, the number of parameters has been debunked by OpenAI's CEO (The Verge, 2023b). However, the figures included here have been reported by a number of outlets (for example, see E2Analyst, 2023). In any case, the main takeaway is that GPT-4 is built on a massively larger dataset and uses a massively larger number of parameters than GPT-3.
- 7 See <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 8 See <https://bard.google.com>
- 9 See <https://writersonic.com/chat>
- 10 See <https://yiyian.baidu.com/welcome>
- 11 See <https://huggingface.co/chat>
- 12 See <https://www.jasper.ai>
- 13 See <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>
- 14 See <https://open-assistant.io>
- 15 See <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>
- 16 See <https://you.com>
- 17 See <https://www.chatpdf.com>
- 18 See <https://elicit.org>
- 19 See <https://www.perplexity.ai>
- 20 See <https://tools.zmo.ai/webchatgpt>
- 21 See <https://www.compose.ai>
- 22 See <https://www.teamsmart.ai>
- 23 See <https://wiseone.io>
- 24 See <https://www.microsoft.com/en-us/bing>
- 25 See <https://www.crayon.com>
- 26 See <https://openai.com/product/dall-e-2>
- 27 See <https://dream.ai/create>
- 28 See <https://www.fotor.com/features/ai-image-generator>
- 29 See <https://www.midjourney.com>
- 30 See <https://creator.nightcafe.studio>
- 31 See <https://writersonic.com/photosonic-ai-art-generator>

- 32 See <https://elai.io>
- 33 See <https://www.gliacloud.com>
- 34 See <https://pictory.ai>
- 35 See <https://runwayml.com>
- 36 See <https://www.aiva.ai>
- 37 See <https://boomy.com>
- 38 See <https://soundraw.io>
- 39 See <https://www.voicemod.net/text-to-song>
- 40 See <https://openai.com/research/gpt-4>
- 41 See <https://www.educhat.top> and <https://www.mathgpt.com>
- 42 See <https://www.educhat.top>
- 43 See <https://www.mathgpt.com>
- 44 There are a few exceptions, such as Hugging Face, a group that is dedicated to open-source AI development.
- 45 See, for example, calls from Google (2023a) and OpenAI (Bass and Metz, 2023).
- 46 For one project to regulate AI see the European Commission's draft AI Act (2021).
- 47 The review was based on data collected from a UNESCO survey distributed to its 193 Member States on the governmental use of AI in education (UNESCO, 2023c), the OECD AI Policy Observatory, and Stanford University's AI Index Report (Stanford University, 2023), and first-hand information elicited from a group of international experts.
- 48 See <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- 49 From the mapping, as of April 2023, the following countries had published national strategies on AI: Argentina, Australia, Austria, Belgium, Benin, Brazil, Canada, Bulgaria, Chile, China, Columbia, Cyprus, Czechia, Denmark, Egypt, Estonia, Finland, France, Germany, Hungary, Iceland, India, Indonesia, Ireland, Italy, Japan, Jordan, Latvia, Lithuania, Luxembourg, Malaysia, Malta, Mauritius, Mexico, Netherlands (Kingdom of the), Norway, New Zealand, Oman, Peru, Poland, Portugal, Philippines, Qatar, Republic of Korea, Romania, Russian Federation, Saudi Arabia, Serbia, Singapore, Slovenia, Spain, Sweden, Thailand, Türkiye, Tunisia, United Arab Emirates, United Kingdom, United States, Uruguay and Viet Nam. Additionally, some countries have incorporated AI strategies within broader ICT or digital strategies, including Algeria, Botswana, Kazakhstan, Kenya, Sierra Leone, Slovakia, Switzerland and Uganda.
- 50 According to a rapid review of all national AI strategies (UNESCO, 2023b), over 40 strategies have dedicated sections on the issue of ethics.
- 51 According to a rapid review of all national AI strategies (UNESCO, 2023b), around 45 strategies have dedicated sections on the issue of education.
- 52 See <https://openai.com/policies/terms-of-use>
- 53 In some countries, a teacher will have a teaching assistant (TA) whose role is to spend time answering the questions of individual students covering the course material. GenAI might be used to develop a generative twin of a TA, which can be supportive to the students and other teachers, but may also cause some negative issues (e.g. around social relationships in the classroom).



**unesco**

United Nations  
Educational, Scientific  
and Cultural Organization

# Guidance for generative AI in education and research

This Guidance aims to support the planning of appropriate regulations, policies and human capacity development programmes to ensure that generative artificial intelligence (GenAI) becomes a tool that genuinely benefits and empowers teachers, learners and researchers. It explains the AI techniques used by GenAI and maps out a list of GPT models that are made publicly available, especially those under open-source licences. It also opens a discussion on the emergence of EdGPT – GenAI models that are trained with specific data to serve educational purposes. Furthermore, it summarizes some of the key controversies around GenAI, from worsening digital poverty to the homogenization of opinions, and from deeper deepfakes to issues of copyright. Based on a humanistic vision, the Guidance proposes key steps for the regulation of GenAI tools, including mandating the protection of data privacy and setting an age limit for independent conversations with GenAI platforms. To guide the proper use of the tools in education and research, this Guidance proposes a human-agent and age-appropriate approach to the ethical validation and pedagogical design processes.

