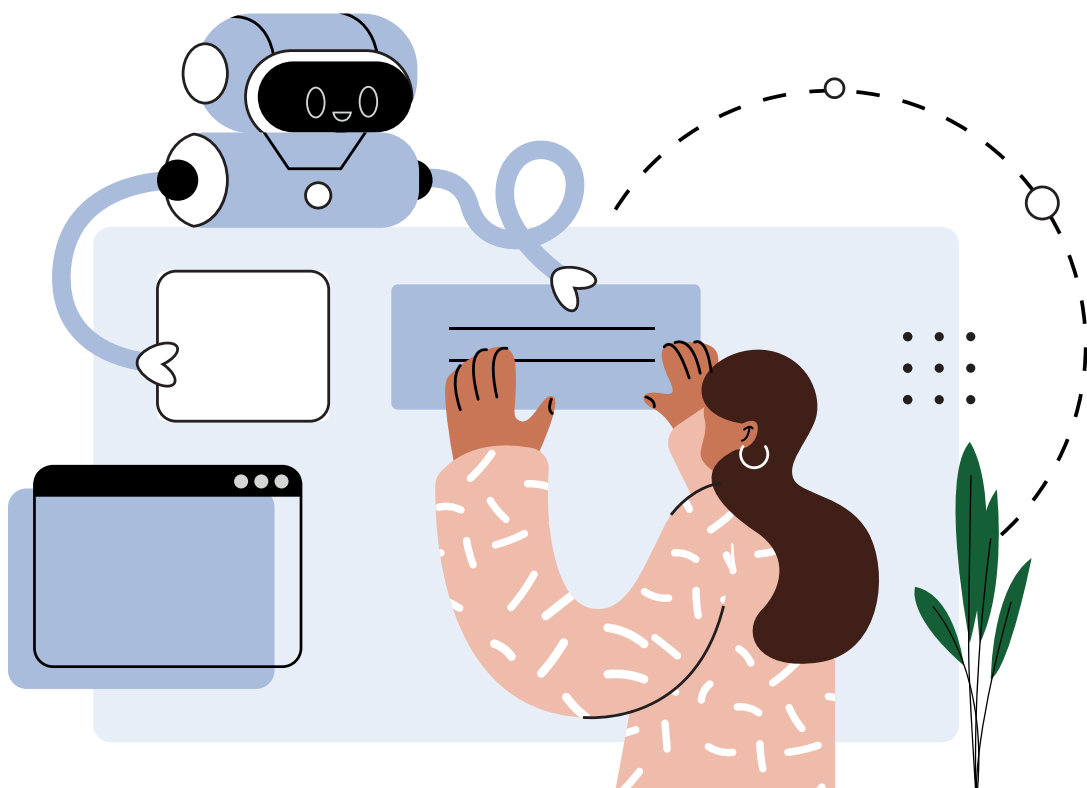


# 生成式人工智能 教育与研究应用指南



## 教科文组织——全球教育领导机构

教育是联合国教科文组织工作的重中之重，它既是一项基本人权，也是建设和平推动可持续发展的基础。教科文组织是主管教育的联合国专门机构，在全球和地区的教育领域发挥领导作用，以推动各国教育系统的发展，增强其韧性和能力，从而服务所有学习者。教科文组织通过变革性学习引领应对当今全球挑战的行动，并在所有业务领域重视性别平等和非洲。



联合国教育、  
科学及文化组织

## 2030年全球教育议程

教科文组织作为主管教育的联合国专门机构，负责领导并协调2030年教育议程——旨在通过17项可持续发展目标在2030年前消除贫穷的全球运动的一部分。教育既是实现各项可持续发展目标的关键，同时自身也是单独一项目标（可持续发展目标4），即“**确保包容和公平的优质教育，让全民终身享有学习机会**”。《2030年教育行动框架》为落实这一宏伟目标及各项承诺提供了指导方针。



由联合国教育、科学及文化组织于2025年出版  
法国巴黎丰特努瓦广场7号，邮编75352

© 联合国教科文组织2025

ISBN 978-92-3-500058-0

DOI <https://doi.org/10.54675/PYTW8189>



本出版物为开放获取出版物，授权协议为Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<https://creativecommons.org/licenses/by-sa/3.0/igo/>)。用户使用本出版物内容，即表明同意接受教科文组织开放获取资源库使用条件的约束 (<https://www.unesco.org/zh/open-access/cc-sa>)。

原版出版物名称: *Guidance for generative AI in education and research*  
联合国教育、科学及文化组织2023年出版

本出版物中使用的名称和材料的呈现方式，不代表联合国教科文组织对任何国家、地区、城市或地区的法律地位或其当局，或对其边界或界限的划定表达任何意见。

本出版物中标有星号(\*)的图像不属于CC-BY-SA许可，未经复制权所有者的明确许可，不得以任何方式复制。

本出版物中表达的观点和意见属于作者本人；它们不一定代表联合国教科文组织的观点，也不对组织构成承诺。

同时，对浙江大学教育学院阚阅、李艳、翟雪松、阮玉芳、袁婧、朱雨萌、许家奇、楚肖燕、许洁、吴庭辉、王之圣，以及浙江省教育技术中心王会军在本出版物中文版的翻译及文本校对方面的贡献表示感谢。

封面图片来源: Olexandra Simkina/Shutterstock.com\*

由联合国教科文组织设计和印刷

法国印刷

CLD\_PAO (416\_25)

## 摘要

# 通向以人为本的生成式人工智能应用方法

公开可用的生成式人工智能工具正迅速涌现，其迭代版本的发布速度超过了国家监管框架的适应速度。大多数国家对生成式人工智能缺乏国家层面的法规，导致用户数据隐私缺乏保护，教育机构在验证这些工具方面也普遍准备不足。

联合国教科文组织发布的首份教育领域生成式人工智能全球指南旨在支持各国实施即时行动、规划长期政策，并培养人才，以确保对这些新技术采取以人为本的视角。

本指南评估了生成式人工智能可能对促进人的能动性、包容、公平、性别平等、语言和文化多样性以及多元意见和表达的核心人文价值构成的潜在风险。

它提出了政府机构规范生成式人工智能工具使用的关键步骤，包括强制保护数据隐私并考虑使用年龄限制。它概述了生成式人工智能提供商的要求，以促进其在教育中的伦理和有效使用。

本指南强调教育机构需要验证生成式人工智能系统在伦理和教学适应性方面是否适合教育。它呼吁国际社会反思这些问题对知识、教学、学习和评价的长期影响。

本出版物为政策制定者和教育机构提供了具体建议，说明如何设计生成式人工智能工具的使用，以保护人的能动性并真正惠及学习者、教师和研究人员。

虽然ChatGPT  
在2023年1月的  
月活跃用户就已突破

**1亿**

但截至2023年7月，  
只有一个国家发布了  
关于生成式人工智能的  
监管规定



unesco

“战争起源于人之思想，  
故务需于人之思想中  
筑起保卫和平之屏障”

**生成式人工智能**

**教育与研究应用指南**

# 序 言



© UNESCO

2022年末，随着ChatGPT的推出，生成式人工智能突然进入公众视野，成为历史上发展最快的应用程序。生成式人工智能应用程序可以模仿人类生成文本、图像、视频、音乐和软件代码等，在全球范围内引起了轰动。如今，数百万人在日常生活中使用生成式人工智能，将这些模型适配到特定领域的人工智能应用的潜力似乎是无限的。

这种广泛的信息处理和知识生产能力对教育可能有着巨大的影响，因为它们复制了构成人类学习基础的高阶思维。随着生成式人工智能工具越来越能够自动化一些基础的写作和艺术创作，它们迫使教育政策制定者和机构重新审视我们学习的目的、内容和方式。在数字时代的这一新阶段，这些现在是教育的关键考量。

本出版物旨在支持规划适当的法规、政策和人才发展，以确保生成式人工智能成为真正有益于并赋权教师、学习者和研究人员的工具。

它提出了政府机构规范生成式人工智能使用的关键步骤。它还介绍了政策制定和教学设计框架及具体示例，以促进在教育中伦理和有效地使用这项技术。最后，它呼吁国际社会考虑生成式人工智能对我们理解知识和定义学习内容、方法和成果以及我们评估和验证学习方式的深远长期影响。

基于联合国教科文组织2021年发布的《人工智能伦理建议书》，本指南以促进人的能动性、包容、公平、性别平等以及文化和语言多样性和多元意见及表达的人文主义教育方法为基础。此外，它回应了2021年“教育的未来”国际委员会的报告《一起重新构想我们的未来：为教育打造新的社会契约》，作为我们努力更新教育社会契约、重新定义我们与技术关系的不可分割的一部分。

人工智能绝不能取代人类智能。相反，它激励我们重新思考对知识和人类学习的固有理解。我希望这份指南帮助我们重新定义教育的视野，并为我们的集体思维和协同行动提供参考，引领我们迈向以人为本的全民数字学习未来。

斯特凡妮娅·贾尼尼 (Stefania Giannini)

联合国教科文组织教育助理总干事

# 致 谢

在联合国教科文组织教育助理总干事斯特凡尼娅·贾尼尼（Stefania Giannini）的领导下，以及未来学习和创新司司长索比·塔维勒（Sobhi Tawil）的指导下，本出版物的起草工作由教育技术与人工智能处处长苗逢春（Fengchun Miao）领导。

特别感谢伦敦大学学院韦恩·霍尔姆斯（Wayne Holmes）副教授共同起草了本出版物。

本出版物是教育领导者和人工智能与教育领域专家的集体努力的成果。

它得益于包括伦敦大学学院穆特鲁·丘库罗瓦（Mutlu Cukurova）教授；联合国教科文组织南特大学开放教育资源教师培训技术教席主持人科林·德拉·希格拉（Colin dela Higuera）；约翰内斯堡大学研究助理沙菲卡·艾萨克斯（Shafika Isaacs）；应用发明者基金会（App Inventor Foundation）执行董事娜塔莉·劳（Natalie Lao）；上海师范大学倪琴副教授；罗马尼亚欧洲数字教育中心教育信息通信技术专家卡塔琳娜·尼古林（Catalina Nicolin）；联合国教科文组织伦敦大学学院人工智能教席主持人、伦敦大学学院计算统计学和机器学习教授约翰·肖·泰勒（John Shaw-Taylor）；Jet教育服务执行经理凯利·希罗希拉（Kelly Shiohira）；韩国国立教育大学宋基相（Ki-Sang Song）教授；以及芬兰意义处理有限公司首席科学家伊尔卡·图米（Ilkka Tuomi）等许多专家的意见和投入。

联合国教科文组织的许多同事也以各种方式做出了贡献，包括生物伦理和科学技术伦理处处长达夫纳·费恩霍尔兹（Dafna Feinholz）；拉丁美洲和加勒比地区高等教育国际研究所所长弗朗切斯克·佩德罗（Francesc Pedró）；数字政策和数字化转型处项目专家普拉提克·西巴尔（Prateek Sibal）；政策和终身学习系统司教师发展处高级项目官员索拉布·罗伊（Saurabh Roy）；曼谷办事处教育创新和技能发展处信息通信技术教育项目专员本杰明·维格尔·德·迪奥斯（Benjamin Vergel De Dios）；文化部门文化表达多样性实体的同事们；以及未来学习和创新司项目专家马克·韦斯特（Mark West）。

本出版物的制作得到了未来学习和创新司教育技术与人工智能处格伦·赫特伦迪（Glen Hertelendy）、路易莎·费拉拉（Luisa Ferrara）和郑祥蕾（Xianglei Zheng）的协调支持。

同时，对詹妮·韦伯斯特（Jenny Webster）在文本校对和审查方面的工作，以及陈水玉（Ngoc-Thuy Tran）在版面设计上的贡献表示感谢。

# 目 录

- 序 言 .....6
- 致 谢 .....7
- 缩略词和缩写词列表 .....10
- 引 言 .....11
- 1. 什么是生成式人工智能及如何运行 .....12
  - 1.1 什么是生成式人工智能？ .....12
  - 1.2 生成式人工智能如何运行？ .....12
    - 1.2.1 文本生成式人工智能模型如何运行？ .....13
    - 1.2.2 图像生成式人工智能模型如何运行？ .....15
  - 1.3 通过提示工程生成期望的输出结果 .....15
  - 1.4 新兴的教育生成式预训练转换器（EdGPT）及其影响 .....17
- 2. 关于生成式人工智能及其对教育影响的争议 .....18
  - 2.1 数字贫困的加剧 .....18
  - 2.2 国家监管调控的滞后 .....18
  - 2.3 未经授权内容的使用 .....19
  - 2.4 生成模型的不可解释性 .....19
  - 2.5 人工智能生成内容对网络的污染 .....20
  - 2.6 真实世界理解的缺失 .....20
  - 2.7 观点多样性的减少及少数者声音的进一步边缘化 .....21
  - 2.8 更加深度伪造内容的生成 .....21
- 3. 生成式人工智能在教育领域的规范应用 .....22
  - 3.1 以人为本的人工智能方法 .....22
  - 3.2 规范生成式人工智能在教育应用中的步骤 .....22
  - 3.3 生成式人工智能监管：关键要素 .....24
    - 3.3.1 政府监管机构 .....24
    - 3.3.2 生成式人工智能工具提供者 .....25
    - 3.3.3 机构用户 .....26
    - 3.3.4 个人用户 .....27
- 4. 为在教育和研究中使用生成式人工智能制定政策框架 .....28
  - 4.1 促进包容、公平以及语言和文化多样性 .....28
  - 4.2 保护人的能动性 .....28

4.3	监控并验证用于教育的生成式人工智能系统.....	29
4.4	培养学习者包括生成式人工智能相关技能在内的人工智能素养.....	29
4.5	培养教师和研究人员正确使用生成式人工智能的能力.....	30
4.6	促进多元意见和多元观点表达.....	30
4.7	测试本地相关的应用模型并建立不断迭代的证据库.....	31
4.8	以跨部门和跨学科的方式审查长期影响.....	31
<b>5.</b>	<b>促进创造性地将生成式人工智能应用在教育和研究中.....</b>	<b>32</b>
5.1	促进负责任和创造性使用生成式人工智能的制度战略.....	32
5.2	融合“以人为本和适切教学互动”的方法.....	33
5.3	共同设计生成式人工智能在教育和研究中的应用.....	33
5.3.1	用于研究的生成式人工智能.....	33
5.3.2	促进教学的生成式人工智能.....	34
5.3.3	作为一对一教练促进个性化基本技能学习的生成式人工智能.....	35
5.3.4	促进探究性学习或项目式学习的生成式人工智能.....	37
5.3.5	支持有特殊需求学习者的生成式人工智能.....	38
<b>6.</b>	<b>生成式人工智能与未来教育和研究的关系.....</b>	<b>40</b>
6.1	未知的伦理问题.....	40
6.2	版权与知识产权.....	40
6.3	学习内容来源.....	40
6.4	同质化反应与多样化和创造性的产出.....	41
6.5	对评价方式和学习结果的再思考.....	41
6.6	思维过程.....	41
	结语.....	41
	<b>参考文献.....</b>	<b>43</b>

## 表目录

表1. 生成式人工智能使用的技术.....	12
表2. OpenAI GPTs.....	13
表3. 生成式人工智能在研究中应用的共同设计.....	34
表4. 生成式人工智能支持教师和教学应用的共同设计.....	35
表5. 生成式人工智能作为一对一教练实现自主习得语言和艺术基本技能的共同设计.....	36
表6. 生成式人工智能促进探究式学习或项目式学习应用的共同设计.....	37
表7. 生成式人工智能支持有特殊需求学习者应用的共同设计.....	38

# 缩略词和缩写词列表

## 概念与技术

<b>AGI</b>	通用人工智能
<b>AI</b>	人工智能
<b>API</b>	应用程序编程接口
<b>ANN</b>	人工神经网络
<b>DAI</b>	分布式人工智能
<b>GAN</b>	生成对抗网络
<b>GB</b>	千兆字节
<b>GDPR</b>	通用数据保护条例
<b>GenAI</b>	生成式人工智能
<b>GPT</b>	生成式预训练转换器
<b>ICT</b>	信息和通信技术
<b>LaMDA</b>	对话应用语言模型
<b>LLM</b>	大型语言模型
<b>ML</b>	机器学习
<b>VAE</b>	变分自编码器

## 组织机构

<b>AGCC</b>	人工智能政府云集群（新加坡）
<b>CAC</b>	中国国家互联网信息办公室
<b>EU</b>	欧洲联盟
<b>OECD</b>	经济合作与发展组织
<b>UNCTAD</b>	联合国贸易和发展会议
<b>UNESCO</b>	联合国教育、科学及文化组织

# 引言

2022年末生成式人工智能正式问世。这是首个向公众广泛提供的生成式人工智能工具，<sup>1</sup>随后推出的更复杂的版本，在全球范围内引发巨震，并推动各大科技公司在生成式人工智能模型开发领域的竞争。<sup>2</sup>

在世界范围内，教育界最初的担忧是，ChatGPT和类似的生成式人工智能工具会被学生用来在作业中作弊，从而损害学习评价、认证和资格的价值。（Anders, 2023）虽然有些教育机构禁止使用ChatGPT，但也有一些机构谨慎地欢迎生成式人工智能的到来。（Tili, 2023）例如，很多学校和大学采取了一种渐进的方法，认为“需要向学生和教师提供支持，使其有效地、规范地、透明地使用生成式人工智能工具，而非一味持禁用态度。（Russell Group, 2023）。这一应用策略既回应了生成式人工智能的发展是大势所趋，又提示我们生成式人工智能在教育中的应用也必将伴随着复杂与多变，既会对推动教育发展带来独特价值，也会引发诸多负面影响。

事实上，生成式人工智能具有大量可能的用途。它可以对人类思维特征进行仿真，并对仿真的过程和结论实现自动化分析和表征。它可以通过对半成品知识的再加工，输出更为完整的知识产品。它有助于将人类从某些低阶思维技能中解放出来，从而更深一步地认识和理解人类智能和学习的本质。

然而，生成式人工智能也引发了诸多亟待破解的难题，如安全性、数据隐私、版权保护和算法操纵等。其中一些风险是传统人工智能技术的附赘悬疣，这些风险因生成式人工智能的到来而愈加严重；而另一些风险则是随着这一最新一代工具的出现而滋生滋长。因此，深度探究这些问题的本质，并寻求有效破解方案成为当前紧迫的研究议题。

本指南的发布旨在回应这一紧迫需求。然而，应当明确指出的是，生成式人工智能与教育的专题指南并不意味着将生成式人工智能视为解决教育领域基本挑战的独立方案。尽管媒体可能夸大了它的作用，但单靠生成式人工智能不太可能解决全球各地教育系统面临的所有问题。在面对长期存在的教育问题时，关键在于坚持一种理念，即人类的能力和集体行动而非技术，才是解决社会面临的根本性挑战的决定性因素。

因此，本指南的目标是为适当的法规、政策以及人力发展项目的规划提供支持，以确保生成式人工智能成为真正助益和赋能教师、学生、以及研究人员的工具。基于联合国教科文组织《人工智能伦理建议书》，本指南侧重于以人为本的方法，以推动人类的能动性、包容、公平、性别平等、文化和语言多样性，以及多元的观点和表达。

本指南首先介绍了生成式人工智能及其工作原理，涵盖了各种现有技术和模式（第1章），之后确定了涉及通用人工智能和特定生成式人工智能的一系列有争议的伦理和政策问题（第2章）。接下来，它探讨了以人为本方法的基础——一种确保道德、安全、公平和有意义使用生成式人工智能的方法，并阐述了监管生成式人工智能时需要研究的步骤和关键要素（第3章）。第4章主要涉及可采取的措施，以制定连贯、全面的政策框架，规范生成式人工智能在教育和研究中的使用。第5章探讨了在课程设计、教学、学习和研究活动中创造性地使用生成式人工智能的可能性。最后，第6章围绕生成式人工智能对教育和研究的长期影响，并对整个文件进行了总结。

# 1. 什么是生成式人工智能及如何运行

## 1.1 什么是生成式人工智能？

生成式人工智能是一种通过自然语言对话界面响应提示，并自动生成内容的人工智能技术。生成式人工智能不仅仅局限于对现有网页内容的简单整合，而是能够在利用现有内容的基础上，创造出全新的内容。生成式人工智能所生成的内容形式多样，涵盖了能够代表人类思维的各种符号形态，包括自然语言文本、图像（包括照片、数字绘画和卡通等多种类型）、视频、音乐和软件代码等。生成式人工智能的训练数据来源广泛，包括网页、社交媒体对话以及其他形式的在线媒体。它通过对所吸收的数据中词语、像素或其他元素的分布进行统计分析，识别并复现常见模式(例如，某些词语通常跟在哪些词语之后)来生成内容。

虽然生成式人工智能能够生成新颖的内容，但它无法产生解决现实世界问题的新想法或方案。这主要是因为生成式人工智能对构成语言基础的现实世界事物和社会关系缺乏真正的理解。此外，尽管其生成的内容在表面上看起来流畅自然、令人印象深刻，但内容的准确性和可靠性却无法得到完全保证。事实上，即

便是ChatGPT的供应方也承认：“尽管ChatGPT等工具通常可以生成听起来合理的答案，但我们不能完全依赖其准确性。”在实际应用中，除非用户本身对相关话题有深厚的知识积累，否则很难察觉生成式人工智能所输出内容中的错误和偏差。

## 1.2 生成式人工智能如何运行？

生成式人工智能所依托的核心技术属于机器学习的范畴。机器学习是一种利用算法和数据实现持续自动性能提升的技术族。近年来，人工智能在人脸识别等领域取得的显著进展，主要得益于一种被称为人工神经网络（ANNs）的机器学习技术。人工神经网络的设计灵感源于对人脑工作方式及其神经元之间的突触连接。人工神经网络可以细分为多种类型。

文本和图像生成式人工智能技术都基于一套研究人员已经掌握多年的人工智能技术。<sup>1</sup>以ChatGPT为例，其采用了生成式预训练转换器（GPT）模型；而在图像生成领域，生成对抗网络（GANs）则是一种被广泛应用的技术路线（详见表1）。<sup>3</sup>

表1. 生成式人工智能使用的技术		
机器学习		一种利用数据自动提高性能的人工智能类型
神经网络		一种受人脑结构和功能(如神经元之间的突触连接)启发的机器学习类型
文本生成式人工智能	通用转换器	一种能够专注于数据的不同部分，以确定它们之间如何相互关联的人工神经网络类型
	大语言模型	一种在大量文本数据上训练的通用转换器类型
	生成式预训练转换器 <sup>1</sup>	一种在更大量数据上预训练的大型语言模型类型，使模型能够捕捉语言的细微差别，并生成连贯的上下文感知文本
图像生成式人工智能	生成对抗网络	用于图像生成的神经网络类型
	变分自编码器(VAEs)	

### 1.2.1 文本生成式人工智能模型如何运行？

文本生成式人工智能使用一种称为通用变换器（General-Purpose Transformer）的人工神经网络，以及一种称为大语言模型的通用变换器。这就是文本生成式人工智能系统常被称为大语言模型或LLMs的原因。文本生成式人工智能所使用的LLM类型被称为生成式预训练变换器（Generative Pre-trained Transformer），即GPT（因此有了“ChatGPT”中的“GPT”）。

ChatGPT是基于OpenAI开发的GPT-3模型构建而成的。GPT-3是OpenAI公司GPT系列的第三次迭代升级，该系列的首个版本GPT-1于2018年推出，而最新的GPT-4则于2023年3月发布（详见表2）。OpenAI在GPT系列的每一次迭代中，都通过在人工智能架构、训练方法和优化技术等

方面的创新和突破，对前一代模型进行了全面的改进和升级。GPT持续进步的一个显著特点，就是使用了规模不断增长的数据来训练其指数级增长的“参数”数量。这里的“参数”可以比喻为一组可调节的旋钮，用于对GPT的性能进行微调。这些参数包括了模型的“权重”参数，即决定模型在处理输入和生成输出时所采用的具体数值参数。

除了在优化人工智能架构和训练方法方面取得进展外，大公司可用的海量数据和计算能力的提升，也使这种快速迭代成为可能。自2012年以来，用于训练生成式人工智能模型的计算能力呈现出惊人的增长态势，平均每3-4个月就会实现一次翻倍。作为对比，摩尔定律大约每两年翻一番（OpenAI, 2018；Stanford University, 2019）。

表2. OpenAI GPTs

模型	发布时间	训练数据量	参数数量	特点
GPT-1	2018年	40 GB	1.17亿	能够完成文本补全和问题回答等自然语言处理任务
GPT-2	2019年	40 GB	15亿	能够完成更复杂的自然语言处理任务，如机器翻译和摘要
GPT-3	2020年	17,000 GB	1,750亿	能够完成高级自然语言处理任务，如编写连贯的段落和生成完整的文章。还能够仅通过几个示例就适应新任务
GPT-4 <sup>6</sup>	2023年	1,000,000 GB (据报道但未经证实)	1,700,000亿 (据报道但未经证实)	增强了可靠性，能够处理更复杂的指令

一旦GPT完成训练，它便可以接受提示并生成相应的文本响应。这一过程通常包括以下几个步骤：

1. 将输入的提示分解为更小的语言单元（称为标记），并将其传入GPT模型。
2. GPT利用其学习到的统计模式，预测可能构成对提示的连贯响应的单词或短语。
  - GPT首先在其预先构建的大规模数据模型中（该模型包含了从互联网和其他来源抓取的文本数据）识别出通常共同出现的单词和短语模式。

- 基于这些模式，GPT估计特定单词或短语在给定上下文中出现的概率。
  - 从随机预测开始，GPT使用这些估计的概率来预测其响应中下一个可能出现的单词或短语。
3. 将预测得到的单词或短语转换为可读的文本形式。
  4. 将生成的可读文本通过一系列称为“防护措施”（Guardrails）的过滤机制，以去除其中可能存在的任何冒犯性内容。
  5. 重复执行步骤2到4，直到生成的响应达到预设的完成条件。这些条件可以是响应达到最大标记限制，或满足某些预定义的停止标准。

6. 对生成的响应进行后处理，以提高其可读性和自然性。这一过程可能包括应用格式、标点以及其他增强性的措施，例如在回应开头添加人类常用的词语，如“当然”、“确实”或“我很抱歉”等。

尽管自2018年以来，研究人员就已经可以使用GPT模型及其自动生成文本的能力，但ChatGPT的推出之所以如此新颖，是因为它通过一个易于使用的界面提供了免费访问。这意味着任何拥有互联网连接的人都可以探索这一工具。ChatGPT的推出在全球范围内引发了巨大震动，并迅速导致许多全球科技巨头竞相追赶。与此同时，众多初创公司也开始行动，或是推出与ChatGPT类似的系统，或是在其基础上构建新的工具和应用。

截至2023年7月，ChatGPT的一些替代方案包括：

- **Alpaca**:<sup>7</sup>由斯坦福大学开发，基于Meta的Llama模型进行微调。其主要目标是解决大型语言模型（LLMs）在虚假信息、社会刻板印象和有毒语言方面存在的问题。
- **Bard**:<sup>8</sup>谷歌基于其LaMDA和PaLM 2系统开发的大型语言模型。其特点是可以实时访问互联网，从而提供最新的信息和数据。
- **Chatsonic**:<sup>9</sup>由Writesonic公司推出，在ChatGPT的基础上增加了直接从谷歌抓取数据的能力，因此在产生事实错误答案方面的风险较小。
- **Ernie**（也称为文心一言）：<sup>10</sup>百度开发的双语大型语言模型，目前仍在开发中。它将广泛的知识与海量数据集相结合，能够生成文本和图像。
- **Hugging Chat**:<sup>11</sup>由“拥抱脸”（Hugging Face）公司推出，其开发、训练和部署过程都强调伦理和透明度。此外，用于训练其模型的所有数据都是开源的。
- **Jasper**:<sup>12</sup>一套工具和API，可以根据用户特定的首选风格进行写作训练。同时，它还具备图像生成能力。

- **Llama**:<sup>13</sup>Meta推出的开源大型语言模型，其优势在于需要更少的计算能力和资源，便于测试新方法、验证他人的工作并探索新的用例。
- **Open Assistant**:<sup>14</sup>一种开源方法，旨在使任何具备足够专业知识的人都能开发自己的大型语言模型。它建立在由志愿者策划的训练数据之上。
- **通义千问**:<sup>15</sup>阿里巴巴推出的大型语言模型，可以用英语或中文回应提示。目前正在被集成到阿里巴巴的商业工具套件中。
- **YouChat**:<sup>16</sup>结合了实时搜索功能的大型语言模型，可提供额外的上下文和见解，以生成更准确、更可靠的结果。

目前，市场上大多数ChatGPT替代工具都提供免费使用（在一定限制范围内），而其中一些则采用了开源模式。此外，许多其他产品正在基于这些大型语言模型之一进行开发和推广。以下是几个典型的例子：

- **ChatPDF**:<sup>17</sup>总结并回答与提交的PDF文档相关的问题。
- **Elicit AI研究助手**:<sup>18</sup>旨在自动化研究人员的部分工作流程，通过识别相关论文并总结关键信息。
- **Perplexity**:<sup>19</sup>为寻求快速、准确、量身定制答案的用户提供了一个“知识中心”。

除了独立的工具外，基于大型语言模型的功能也被嵌入到了其他产品中，如网络浏览器。以下是几个基于ChatGPT构建的Chrome浏览器扩展示例：

- **WebChatGPT**:<sup>20</sup>允许ChatGPT访问互联网，从而生成更准确、更新的对话内容。
- **Compose AI**:<sup>21</sup>在电子邮件和其他场景下自动完成句子。
- **TeamSmart AI**:<sup>22</sup>提供了“虚拟助手团队”。
- **Wiseone**:<sup>23</sup>简化在线信息。

此外，ChatGPT已被集成到了一些搜索引擎中，<sup>24</sup>并正在被引入到微软Word和Excel等大型生

产力工具组合中。这一趋势将使ChatGPT在全球办公室和教育机构中得到更广泛的应用和普及（Murphy Kelly, 2023）。

最后，作为向图像生成式人工智能过渡的一个有趣转变，OpenAI最新推出的GPT-4模型具备了在提示中接受图像和文本的能力，呈现出多模态的特性。这一变化引发了人们对“大型语言模型”这一名称是否还适用的讨论。事实上，这也是斯坦福大学研究人员提出“基础模型”（foundation model）这一替代术语的原因之一（Bommasani et al., 2021）。不过，这个新的术语目前尚未得到广泛采用和认可。

### 1.2.2 图像生成式人工智能模型如何运行？

图像生成式人工智能和音乐生成式人工智能通常采用一种称为生成对抗网络的特殊类型人工神经网络。这种网络也可以与变分自编码器（Variational Autoencoders）相结合使用。一些图像生成式人工智能模型，如Dall·E和Stable Diffusion，则使用了另一种名为扩散模型（Diffusion Models）的生成式人工神经网络。以生成对抗网络为例来解释图像生成式人工智能模型的工作原理：生成对抗网络由两个相互“对抗”的部分组成，即“生成器”（Generator）和“判别器”（Discriminator）。在图像生成的场景中，生成器根据给定的提示创建一个随机图像，而判别器则试图区分这个生成的图像和真实图像。随后，生成器利用判别器的反馈结果来调整自身的参数，以生成另一个图像。这个过程可能会重复数千次，使得生成器生成的图像越来越逼真，判别器也越来越难以将其与真实图像区分开来。举例来说，一个在数千张风景照片数据集上成功训练的生成对抗网络模型，可能会生成全新的、但并不真实存在的风景图像。这些生成的图像在视觉上几乎与真实照片无法区分。类似地，一个在流行音乐（甚至是单个艺术家的音乐）数据集上训练的生成对抗网络模型，则可能会生成遵循原始音乐结构和复杂性的全新音乐作品。

截至2023年7月，已有多款**图像生成式人工智能**模型可供使用，它们都能根据文本提示生成相应的图像。大多数模型在一定限度内提供免费使用。以下是一些主要的图像生成式人工智能工具：

- **Craiyon**：<sup>25</sup>以前被称为DALL·E mini。
- **DALL·E 2**：<sup>26</sup>由OpenAI开发的图像生成式人工智能工具。
- **DreamStudio**：<sup>27</sup>基于Stable Diffusion的图像生成式人工智能工具。
- **Fotor**：<sup>28</sup>在一系列图像编辑工具中整合了生成式人工智能技术。
- **Midjourney**：<sup>29</sup>一个独立的图像生成式人工智能工具。
- **NightCafe**：<sup>30</sup>提供Stable Diffusion和DALL·E 2的接口。
- **Photosonic**：<sup>31</sup>由WriteSonic推出的人工智能艺术生成器。

一些易于访问的**视频生成式人工智能**工具也已经面世，包括：

- **Elai**：<sup>32</sup>可以将演示文稿、网站和文本转换为视频。
- **GliaCloud**：<sup>33</sup>可以从新闻内容、社交媒体帖子、实时体育赛事和统计数据生成视频。
- **Pictory**：<sup>34</sup>可以自动从长篇内容创建短视频。
- **Runway**：<sup>35</sup>提供一系列视频(和图像)生成和编辑工具。

最后，这里是一些易于访问的**音乐生成式人工智能**示例：

- **Aiva**：<sup>36</sup>可以自动创建个性化的音轨。
- **Boomy**、<sup>37</sup>**Soundraw**、<sup>38</sup>和**Voicemod**：<sup>39</sup>无需音乐创作知识，即可从任何文本生成歌曲。

## 1.3 通过提示工程生成期望的输出结果

虽然使用生成式人工智能可以像输入问题或其他提示一样简单，但现实情况是，用户要获得完全符合自己期望的输出结果并非易事。以在美国科罗拉多州博览会上获奖的突破性人工智能作品《空间歌剧院》（Théâtre D'opéra Spatial）为例，创作者花费了数周时间编写提示，并对数百张图像进行微调，才最终生成了提交的获奖作品（Roose, 2022）。在文本生成式人工智能领域，编写有效提示所面临的类似

挑战已经导致招聘网站上出现越来越多的提示工程相关岗位（Popli, 2023）。“提示工程”是指编写输入的过程和技术，旨在生成更加接近用户预期目的的生成式人工智能输出结果。

当提示语能够围绕特定问题阐述一个连贯的推理链，或以合乎逻辑的顺序表达一系列思想时，其效果往往最为显著。具体建议包括：

- 提示语应当使用**简单、清晰、直白**的语言，避免复杂或模棱两可的措辞，以确保模型能够便于理解。
- 提示语中应当包含恰当的**示例**，以说明我们期望的响应或生成结果的格式。
- 为了使生成的结果更加相关和有意义，提示语应当提供足够的**背景信息和上下文支持**。
- 提示语应当根据实际效果和反馈不断进行**调整和迭代**，并尝试不同的变化和组合。
- 提示语要始终**遵守道德规范和伦理原则**，避免可能产生不当、偏颇或有害内容的提示。

同时，我们必须立即认识到，在未经批判性评估的情况下，不能完全依赖生成式人工智能的输出结果。正如OpenAI在介绍其最先进的GPT模型时所指出的那样<sup>40</sup>：

尽管GPT-4展现出了强大的能力，但它与早期的GPT模型仍然存在着类似的局限性。最为关键的是，GPT-4目前还不够可靠，有时会“幻想”事实并出现推理错误。因此，在利用语言模型的输出结果时，尤其是在高风险情境下，我们必须格外谨慎，并根据特定用例的需求采取精确的协议和措施（如人工审核、提供额外背景以厘清事实，或完全避免在高风险

鉴于生成式人工智能输出的质量，在大规模或高风险场景中采用这些工具之前，应当进行严格的用户测试和性能评估。这些测试应根据用户要求生成式人工智能提供输出的任务类型，设计出最为相关的性能指标。例如，对于解决数学问题，可以使用“准确性”作为主要指标，以量化生成式人工智能工具生成正确答案的频率；对于回答敏感问题，衡量性能的主要指标可能是“回答率”，即直接回答问题的频率；对于代码生成，指标可以是“直接可执行的生成代码的比例”，即评估生成的代码是否可以直接在编程环境中执行并通过单元测试；对于视觉推理，指标可以是“精确匹配”，即判断生成的视觉对象是否与真实情况完全一致（Chen et al., 2023）。

总的来说，从表面上看，生成式人工智能似乎易于使用；然而，要获得更加复杂的输出结果，仍然需要有经验的专业人员参与其中，并在使用前进行严格的评估和测试。



### 对教育和研究的影响

尽管生成式人工智能可能有助于为教师和研究人員提供有用的文本和其他输出结果，从而为其工作提供支持，但这一过程并非一蹴而就。事实上，为了获得理想的输出结果，可能需要经过多次迭代提示。令人担忧的是，由于年轻学习者确实在专业知识和经验方面尚不及教师，他们可能在缺乏充分认知和批判性思考的情况下，轻易接受生成式人工智能输出的肤浅、不准确，甚至有害的内容。

#### 1.4 新兴的教育生成式预训练转换器 (EdGPT) 及其影响

鉴于生成式人工智能模型可以作为开发更专业化或特定领域模型的基础或起点，一些研究者提出将GPT重新命名为“基础模型”(Foundation Models) (Bommasani et al., 2021)。在教育领域，开发人员和研究人员已经开始对基础模型进行微调，以开发教育生成式预训练转换器。<sup>41</sup>教育生成式预训练转换器模型利用特定的教育领域数据进行训练，以服务于教育目的。换言之，教育生成式预训练转换器旨在通过利用少量高质量的教育领域特定数据，对从海量通用训练数据中派生出的模型进行优化。

这一特性使教育生成式预训练转换器有可能为实现4.3节所列的教育变革提供更多支持。以针对课程共同设计的教育生成式预训练转换器模型为例，它可以让教育者和学习者生成与有效教学方法、特定课程目标和学习者的挑战水平紧密相关的适当教育材料，如教案、测验和互动活动。类似地，在1:1语言技能辅导的背景下，使用特定语言的恰当文本优化后的基础模型，可用于生成示例句子、段落或对话，供学习者练习。当学习者与该模型互动时，模型可以给出与其水平相符的相关且语法正确的文本响应。从理论上讲，教育生成式预训练转换器模型的输出可能包含比标准GPT更少的一般偏见或令人反感的内容，但仍可能产生错误。需

要特别指出的是，除非生成式人工智能的基础模型和方法发生重大变革，否则教育生成式预训练转换器仍可能存在错误并表现出其他局限性。因此，教育生成式预训练转换器的主要用户，尤其是教师和学习者，仍需要以批判的视角看待任何输出结果，这一点至关重要。

目前，针对在教育领域更有针对性地使用GPT而对基础模型进行优化的工作仍处于起步阶段。例如，华东师范大学开发了一个名为EduChat的基础模型，旨在为教学和学习提供服务。该模型的代码、数据和参数均以开源方式共享。<sup>42</sup>除EduChat外，好未来教育集团也正在开发一个名为MathGPT的大型语言模型。这个模型专注于为全球用户解决数学相关问题和提供数学讲解。<sup>43</sup>

然而，在取得重大进展之前，优化基础模型的工作尤为关键。这不仅需要在模型中添加学科知识和消除偏见，还要融入与学习方法相关的知识，并将其体现在算法和模型的设计中。在这一过程中，我们面临着诸多挑战。一是要确定教育生成式预训练转换器模型在掌握学科知识之外，还能在多大程度上以学生为中心，促进积极的师生互动。另一个挑战是确定在何种程度上可以合乎道德地收集和使用学习者和教师的数据，以优化教育生成式预训练转换器。最后，还需要开展严谨的研究，以确保教育生成式预训练转换器的应用不会侵犯学生的基本人权，也不会削弱教师的专业能力。

## 2. 关于生成式人工智能及其对教育影响的争议

基于前一章节对生成式人工智能的概念及工作机制的讨论，本章将进一步考察生成式人工智能系统引发的争议和伦理风险，并考虑其对教育的影响。

### 2.1 数字贫困的加剧

正如前文所述，生成式人工智能的发展依赖庞大的训练数据、海量的算力资源以及不断迭代创新的人工智能架构和训练方法。这些资源大多只能被国际科技巨擘和少数几个经济体（主要是美国、中国，以及欧洲较为集中的几个国家）所掌握。这意味着，创造和控制生成式人工智能的可能性超出了大多数公司和国家的能力范围，尤其是对“全球南方”（Global South）而言。

随着数据获取对国家经济发展和个人数字机遇变得愈发重要，无法获取或负担足够数据的国家和个人则面临着“数据贫困”（Data Poverty）的境地（Marwala, 2023）。在算力获取方面的情况也类似。生成式人工智能在技术先进的国家和地区的快速普及指数级加速了数据的生成和处理，并同时加剧了人工智能财富在“全球北方”（Global North）的集中。作为直接结果，数据贫困地区进一步被排除在外，并长期面临被GPT模型中嵌入的标准殖民的风险。目前的ChatGPT模型是基于在线用户数据进行训练的，这些数据反映了“全球北方”的价值观和规范。因此，此类模型并不适用于“全球南方”的许多数据贫困社区或“全球北方”更弱势社区的本地化相关人工智能算法。



#### 对教育和研究的影响

对根植于生成式人工智能训练模型中的价值取向、文化标准和社会习俗，研究人员、教师和学习者应保持批判性的态度。面对训练和控制生成式人工智能模型的差距扩大而加剧的不平等问题，政策制定者也应对此有所警觉，并采取积极措施、制定解决方案。

### 2.2 国家监管调控的滞后

除了日益加剧的数字贫困问题，主流的生成式人工智能供应方也因拒绝接受严格的独立学术审查而受到批评（Dwivedi et al., 2023）。<sup>44</sup>企业的生成式人工智能基础技术往往受企业知识产权的保护。与此同时，许多开始使用生成式人工智能的公司也发现，维护自身的系统安全变得越来越具有挑战性（Lin, 2023）。此外，虽然人工智能行业本身呼吁对其进行监管，<sup>45</sup>但是起草关于创建和使用各类人工智能（包括生成式人工智能）的法规，往往跟不上技术快速发展的脚步。这在一定程度上解释了国家或地方机构在理解和治理法律及伦理问题时所面临的挑战。<sup>46</sup>

尽管生成式人工智能能够赋能人类解决特定的工作任务，但对这些掌控生成式人工智能的公司实施公共监管仍然十分有限。这一现象同时引发了对监管的质疑，特别是关于访问和使用国内数据的问题，包括地方机构和个人的数据，以及在国家领土范围内生成的数据。因此，适切法律规范的重要性日益凸显，以便地方政府机构可以在生成式人工智能的激增浪潮中获得一定控制权，确保这一技术服务公共利益的治理。



### 对教育和研究的影响

研究人员、教师和学习者应当意识到，对于维护国内机构和个人的数字所有权，保障国内生成式人工智能用户的合法权益，以及应对生成式人工智能引发的一系列法律问题，目前仍然缺乏适当的法规条例。

## 2.3 未经授权内容的使用

正如前文所述，生成式人工智能模型是基于大量数据（例如文本、声音、代码和图像）构建的。然而，这些数据通常直接从互联网抓取，且往往没有得到任何所有者的许可。许多图像和代码生成式人工智能系统因此被指控侵犯知识产权。在撰写本文时，已有若干与此问题相关的国际法律案件正在进行中。

此外，也有人指出，GPTs可能违反法律，例如欧盟的《通用数据保护条例》（European Union, 2016）。特别是人们的被遗忘权，因为目前一旦模型训练完成，便无法从GPT模型中删除个人的数据（或该数据的结果）。



### 对教育和研究的影响

- 研究人员、教师和学习者需要尊重数据所有者的权利，并应检查自己使用的生成式人工智能工具是否违反任何现有法规。
- 研究人员、教师和学习者还应意识到，使用生成式人工智能创建的图像或代码存在侵犯他人的知识产权的风险，并且自己在互联网上创建和分享的图像、声音或代码也可能被其他生成式人工智能所利用。

## 2.4 生成模型的不可解释性

长期以来，神经网络普遍被视为“黑箱”，即其内部工作机制难以检验。因此，神经网络并不“透明”或“可解释”，也无法确定其输出结果是如何得出的。

尽管生成式人工智能整体的方法取向和算法使用通常是可解释的，但特定模型及其参数（包括模型的权重）却无法被检查，生成的具体结果也因此无法被解释。以GPT-4为代表的模型中有数十亿个参数及权重（见表2），正是这些权重协同预测了模型从数据中学习的模式，以最终生成输出结果。由于在神经网络中参数和权重并不透明（见表1），因此生成式人工智能模型生成的具体结果也无法被精确解释。

首先，随着生成式人工智能变得越来越复杂（见表2），其缺乏透明性和可解释性的问题日益严重，并常常导致意料之外或违背期望的结果。此外，生成式人工智能模型继承并延续了其训练数据中存在的偏见，并且由于模型的不透明性，这些偏见难以被发现和解决。最后，这种不透明性也是带来生成式人工智能信任问题的关键原因之一（Nazaretsky et al., 2022a）。如果无法得知生成式人工智能系统输出特定结果的机制，用户就会降低采纳或使用这一技术的意愿（Nazaretsky et al., 2022b）。



### 对教育和研究的影响

研究人员、教师和学习者应当意识到，生成式人工智能系统的运作方式类似于黑箱，难以反映具体内容生成的原因。对输出生成机制缺乏可解释性，往往会使用户陷入由生成式人工智能系统内置参数所界定的逻辑。这些参数可能反映特定的文化或商业价值观与规范，致使在生成内容中引入潜在的偏见。

## 2.5 人工智能生成内容对网络的污染

由于GPT的训练数据通常来源于互联网，而互联网中经常包含歧视性和其他不可接受的语言，因此人工智能开发者已不得不实施“防护措施”（Guardrails），以防止GPT输出具有攻击性等其他有悖伦理的内容。然而，由于缺乏严格的监管和有效的监控机制，生成式人工智能所产生的歧视性内容正日益在互联网上传播，污染了全球大多数学习者获取内容或知识的主要来源。对于这一点的认知尤为重要，因为生成式人工智能所生成的内容表面上可能显得相当准确和令人信服，但实际上往往包含错误和偏见的观点。如果年轻学习者对所询问的主题缺乏坚实的知识基础，生成式人工智能的这一特性将带来较高风险。同时，这也给未来期望基于互联网抓取文本进行训练的GPT模型造成了递归隐患，因为这些生成式人工智能生成的文本中也包含了模型中的偏见和错误。



### 对教育和研究的影响

- 研究人员、教师和学习者需要意识到，生成式人工智能系统可能会输出冒犯性和不道德的材料。
- 他们还需要了解到，当未来的GPT模型基于先前GPT模型生成的文本进行训练时，知识的可靠性可能会出现长期问题。

## 2.6 真实世界理解的缺失

文本GPTs有时被蔑称为“随机鹦鹉”（Stochastic Parrots），因为正如前文提到的，尽管它们可以生成看似令人信服的文本，但这些文本往往包含错误，甚至可能包含有害的陈述（Bender et al., 2021）。出现这一问题的原因

在于，GPTs只能够重复它从训练数据中识别的语言模式（通常是来自互联网的文本），始于随机（或“随机生成”）的模式，却不能理解背后的含义——就像鹦鹉可以模仿声音但无法真正理解它所说的内容一样。

生成式人工智能模型“看似”理解它们使用和生成的文本，但在实际上并不能理解语言和现实世界，这二者之间的脱节可能导致教师和学生对其输出的内容寄予不应有的信任。这对未来的教育构成了严重风险。实际上，生成式人工智能并不能通过观察现实世界或使用其他科学方法的关键要素来获取信息，也不能与人类或社会价值观保持一致。正因如此，它无法生成关于现实世界、物体及其关系、人与社会关系、人一物关系或人一技关系的真正新颖内容。关于生成式人工智能模型生成的表面上新颖的内容是否可以被认定为科学知识仍存在争议。

如前所述，GPTs经常会生成不准确或不可靠的文本。实际上，众所周知，GPTs会编造一些在现实生活中不存在的事物。尽管使用这种拟人化且因此具有误导性的术语受到了部分批评，有些人称GPTs的这一现象为“幻觉”（Hallucination）。生成式人工智能供应公司也承认了这一问题。例如，ChatGPT公开界面底部的声明表示：“ChatGPT可能会产生关于人、地点或事实的不准确信息”。

一些倡导者还建议，生成式人工智能代表了面向通用人工智能迈出的重要一步。其中，通用人工智能指的是一种比人类更智能的人工智能类别。然而，这一观点一直受到批评，且有观点认为，至少在基于知识的人工智能（也称为符号或基于规则的人工智能）与基于数据的人工智能（也称为机器学习）能够通过某种方式进行融合之前，人工智能不会朝向通用人工智能进展（Marcus, 2022）。此外，通用人工智能或人工智能具备感知的主张也使我们无法更仔细地考虑现有人工智能所造成的伤害，例如对已经受到歧视群体的隐性偏见（Metz, 2021）。



### 对教育和研究的影响

- 文本生成式人工智能所输出的内容在表面上与人类生成的文本极为相似，仿佛能够理解生成的文本。然而，生成式人工智能并不理解任何内容。相反，这些工具只是根据互联网常见的方式将单词组合在一起。因此，生成的文本也可能是错误的。
- 研究人员、教师和学习者需要意识到生成式人工智能不理解其生成的文本；它有可能并且经常生成不正确的表述；因此，他们需要对生成式人工智能生成的一切内容采取批判性的态度。

## 2.7 观点多样性的减少及少数者声音的进一步边缘化

ChatGPT及类似的工具往往只能够输出标准答案，这些答案继承了用于训练模型的数据的所有者或创建者的价值观。实际上，类似一些常见且达成共识的主题，以及主流的或主导性的观念，如果一系列词语在训练数据中频繁出现，则GPT在输出中很可能会重复这些单词序列。

这一风险可能会限制和削弱多元化观点的发展和多样化思想的表达。数据贫困的人群，包括“全球北方”的边缘化社区，在网络上只有极小或非常有限的数字存在感。他们的观点因此没有被接受，他们的关切在用于训练GPTs的数据中也没有得到体现，因而很少出现在输出内容中。出于这些原因，基于来自互联网网页和社交媒体对话数据进行预训练的方法，使得GPT模型可能进一步边缘化已经处于不利地位的人群。



### 对教育和研究的影响

- 虽然生成式人工智能模型的开发者和供应方对持续解决这些模型的数据集和输出中的偏见负有主要责任，但用户端的研究人员、教师和学习者需要知道，文本生成式人工智能的输出结果仅代表其训练数据产生时世界上最常见或主导的观点，并且其中一些内容是有问题或存在偏见的（例如，刻板印象中的性别角色）。
- 学习者、教师和研究人員不应轻率地将生成式人工智能提供的信息视为绝对真理，而应始终对其进行批判性评估。
- 研究人员、教师和学习者还必须意识到，在生成式人工智能输出的结果中，少数群体的声音可能会被忽视，因为根据定义，少数群体的声音在训练数据中本身就较少。

## 2.8 更加深度伪造内容的生成

除了所有生成式人工智能所共有的争议之外，基于生成对抗网络的生成式人工智能可被利用去篡改现有的图像或视频，从而生成难以与真实图像区分的伪造图像。生成式人工智能使得创建这些“深度伪造”和所谓的“假新闻”变得越来越简单。换句话说，生成式人工智能使某些行为者更容易实施不道德、不正当和犯罪的行为，例如传播虚假信息、宣传仇恨言论，以及在未经他人知情或同意的情况下，将人们的面孔融入完全虚假的、有时甚至是侵害名誉的影片中。



### 对教育和研究的影响

虽然生成式人工智能供应商有责任保护用户的版权和肖像权，但研究人员、教师和学习者也需要意识到，他们在互联网上分享的任何图像都可能被纳入生成式人工智能的训练数据，并可能以不道德的方式被操纵和使用。

## 3. 生成式人工智能在教育领域的规范应用

为了应对生成式人工智能引发的争议，并充分发挥其在教育领域的潜在优势，首先需要对其进行合理的规范。在教育中，生成式人工智能的应用需要采取以人为本的政策措施与行动步骤，以确保其使用符合伦理、安全、公平，并具有实际意义。

### 3.1 以人为本的人工智能方法

2021年，联合国教科文组织发布了《人工智能伦理建议书》，为解决生成式人工智能引发的多重争议提供了必要的规范框架，这些争议包括教育及研究的相关问题。该框架基于以人为本的人工智能方法，倡导人工智能的应用需服务于人类能力的发展，以实现包容、公正和可持续的未来。这一方法必须遵循人权原则，并致力于保护人的尊严和定义知识共享中的文化多样性。在治理层面，以人为本的方法需要适当的规制，确保人类的能动性、信息透明度和公共问责制。

2019年发布的《北京共识——人工智能与教育》进一步阐明了在教育背景下以人为本的方法内涵。该共识强调，人工智能技术在教育中的应用需致力于增强人类在可持续发展中能力，以及在生活、学习和工作中有效的人机协作能力。此外，该共识呼吁采取进一步措施，确保弱势群体能够公平获取人工智能资源，解决不平等问题，并在此过程中促进语言和文化多样性。该共识建议，教育领域的人工智能政策规划应采取全政府、多部门协作以及多利益相关方参与的方式。

《人工智能与教育：政策制定者指南》（UNESCO，2022）在审视了人工智能在教育中的利弊，及教育在发展人工智能能力中的作用之后，进一步突出了人工智能以人为本的意义。为引导人工智能的应用，该指南提出了具体政策建议：(1) 促进包容性的学习机会，特别是为残障学习者等弱势群体提供平等的学习机会；(2) 支持个性化和开放的学习选择；(3) 改善

基于数据的服务和管理，扩大学习机会并提升质量；(4) 监测学习过程，协助教师对潜在的教学风险进行预警；以及(5) 促进对人工智能的伦理和有效使用的理解与技能的培养。

### 3.2 规范生成式人工智能在教育应用中的步骤

ChatGPT发布以前，各国政府已在多个领域（包括教育）制定或调整框架，以规范数据的收集和使用，以及人工智能系统的部署，为新兴生成式人工智能的应用提供必要的立法和政策背景。自2022年11月起，随着多个具有竞争力的生成式人工智能模型相继面世，各国政府采取了不同的政策应对措施——从禁止生成式人工智能的使用，到评估现有框架调整的必要性，再到紧急制定新法规。

2023年4月，联合国教科文组织对各国政府在规范和促进生成式人工智能创造性应用的策略进行了梳理和评估（UNESCO，2023）。该评估提出了七个步骤，供政府机构参考，以规范生成式人工智能的应用，并重新确立公共控制，从而最大限度地发挥其在各领域（包括教育领域）的潜力。

#### 步骤一：批准国际或地区性的通用数据保护条例，或制定国家级数据保护条例

生成式人工智能模型的训练，会涉及收集和处理来自多个国家公民的在线数据。在未经同意的情况下使用这些数据和内容，进一步加剧了数据保护方面的挑战。

以2018年实施的欧盟《通用数据保护条例》为代表的通用数据保护法规，为生成式人工智能提供者提供了规范个人数据收集和处理的必要法律框架。根据联合国贸易和发展会议（UNCTAD）数据保护和隐私立法世界概览门户网站的数据，194个国家中已有137个颁布了数据保护和隐私方面的法律。<sup>48</sup>

然而，这些框架在各国的具体落实情况仍不明确。因此，确保这些法规得到执行，尤其是对生成式人工智能系统的运行进行定期监测，显得尤为重要。同时，对于尚未制定通用数据保护法律的国家，加快立法进程已刻不容缓。

## 步骤二：制定/修订并资助全政府的人工智能战略

对生成式人工智能的规范，须融入更广泛的国家人工智能战略中，以确保包括教育在内的各个发展领域能够安全、公平地使用人工智能。制定、批准、资助和实施国家人工智能战略需要全政府的协同，如此才能确保多部门的协调行动，以应对新兴挑战。

截至2023年初，已有约67个国家<sup>49</sup>制定或计划制定国家人工智能战略，其中61个国家制定了独立的人工智能战略，7个国家将人工智能纳入到更广泛的国家信息通信技术或数字化战略中。然而，鉴于生成式人工智能的创新性，截至本指南撰写时，尚无国家在其人工智能战略中将生成式人工智能作为特定问题加以涵盖。

各国应修订现有的国家人工智能战略或制定新的战略，确保在各领域（包括教育）规范人工智能应用合乎伦理。

## 步骤三：巩固并实施关于人工智能伦理规范的具体法规

为应对人工智能使用带来的伦理挑战，各国政府需要制定具体的法规。

2023年，联合国教科文组织对各国的人工智能战略做了一项调查，调查结果显示，仅有约40个国家的人工智能战略中涵盖了伦理问题识别和指导原则制定。<sup>50</sup> 即便如此，这些伦理原则仍需要转化为可执行的法律或法规。这种转化在各国中并不普遍。事实上，只有约20个国家在其国家人工智能战略或其他政策中明确规定了关于人工智能伦理的具体条款，其中也包括教育领域的伦理问题。有趣的是，尽管约45个国家的人工智能战略中有提到教育这一政策领域，<sup>51</sup> 但大多数讨论更多集中在人工智能技能和人才培养以提升国家竞争力方面，较少关注伦理问题。

尚未制定人工智能伦理法规的国家必须加快制定并实施相关法规。

## 步骤四：调整或实施现有版权法，规范人工智能生成的内容

生成式人工智能的日益普及对版权法提出了新挑战，这不仅涉及模型训练所使用的受版权保护的作品，还涉及人工智能生成的“非人类”知识产出的版权状态。

目前，只有中国、欧盟和美国对版权法进行了调整，以应对生成式人工智能带来的影响。例如，美国版权局裁定，像ChatGPT这样的生成式人工智能系统的输出不受美国版权法保护，理由是“版权只能保护人类创造的作品”（US Copyright Office, 2023）。与此同时，欧盟拟议的《人工智能法》要求人工智能工具开发者披露其在构建系统时所使用的受版权保护的材料（European Commission, 2021）。中国则通过2023年7月发布的生成式人工智能法规，要求将生成式人工智能的输出标记为人工智能生成的内容，并仅将其视为数字合成的产物。

对生成式人工智能模型训练中使用的版权材料进行规范，并明确其输出的版权状态，已成为版权法面临的新责任。因此，调整现有法律以应对此问题已迫在眉睫。

## 步骤五：完善生成式人工智能的监管框架

人工智能技术的迅速发展迫使各国及地方治理机构加快更新相关法规。截至2023年7月，只有中国发布了针对生成式人工智能的具体官方监管规定。2023年7月13日，中国国家网信办发布了《生成式人工智能服务管理暂行办法》（Cyberspace Administration of China, 2023a），要求生成式人工智能系统的提供者依法对人工智能生成的内容、图像和视频进行正确标注，并遵循现行《互联网信息服务深度合成管理规定》。各国亟需基于对现行地方性法律法规的差距评估，制定更多针对生成式人工智能的专门框架。

### 步骤六：提升在教育和研究中正确应用生成式人工智能的能力

学校和其他教育机构需要提升对人工智能（包括生成式人工智能）在教育中的潜在益处和 risks 的理解。只有在这一理解的基础上，他们才能有效评估人工智能工具的应用。此外，还需要为教师和研究人員提供支持，通过培训和持续的辅导，帮助他们提升正确使用生成式人工智能的能力。许多国家（如新加坡）已经启动了此类能力建设计划。新加坡通过其人工智能政府云集群（其中包含一个专门的GPT模型库）为教育机构的人工智能能力发展提供了专门平台（Ocampo, 2023）。

### 步骤七：反思生成式人工智能对教育和研究的长期影响

当前版本的生成式人工智能的影响才刚刚显现，尚未全面探索和理解其对教育的影响。同时，更强大的生成式人工智能及其他类型的人工智能工具也在持续开发和部署。关于生成式人工智能在知识创造、传播和验证方面的深远影响，尤其是其对教学与学习、课程设计与评估，以及科研与版权的影响，仍存在诸多关键问题悬而未决。大多数国家在教育中采用生成式人工智能尚处于早期阶段，对其长期影响仍未全面了解。为了确保人工智能的应用以人为本，亟需开展关于其长期影响的公开辩论和政策对话。这些辩论应当包容政府、私营部门和其他合作伙伴的意见，并为法规与政策的迭代更新提供见解和建议。

## 3.3 生成式人工智能监管：关键要素

所有国家都需要妥善监管生成式人工智能，以确保其为教育及其他领域的发展带来积极影响。本节提出的关键行动要素包括：(1) 政府监管机构，(2) 生成式人工智能工具的提供者，(3) 机构用户，以及 (4) 个人用户。尽管该框架中的许多要素具有跨国性质，但也应结合各国的具体情况进行考虑，即现有的教育系统和通用监管框架。

### 3.3.1 政府监管机构

政府需要采取全政府的方式，协调生成式人工智能监管框架的设计、对接和实施。以下是建议的七个关键要素和行动：

- **部门间协调：**建立一个国家机构，负责主导生成式人工智能的全政府协调，并在不同部门间进行合作。
- **立法对接：**将现有框架与每个国家的相关立法和监管背景对接，例如与通用数据保护法、网络安全法规、公民数据安全法及其他相关法律和常规做法。评估现有法规的适用性，并根据生成式人工智能带来的新问题进行调整。
- **监管与推动创新的平衡：**推动公司、组织、教育和科研机构以及相关公共机构之间的跨部门合作，共同开发值得信赖的模式；鼓励构建开源生态系统，促进超级计算资源和高质量预训练数据集的共享；并推动生成式人工智能在各领域的实际应用，创造对公众有益的高质量内容。
- **潜在风险的评估和分类：**在生成式人工智能服务部署之前及其生命周期内，建立评估其有效性、安全性和保障性的原则和流程。根据生成式人工智能对公民可能带来的风险水平，制定分类机制，将应用程序分为严格监管（如禁止具有不可接受风险的应用程序或系统）、针对高风险应用的特殊监管，以及针对非高风险应用的通用监管。可参照欧盟《人工智能法》草案中的做法。

- **数据隐私保护：**考虑到生成式人工智能的使用几乎总是涉及用户向提供者分享数据，应强制要求制定和实施保护用户个人信息的法律，并识别和打击非法的数据存储、分析及共享行为。
- **生成式人工智能的年龄限制的定義与实施：**大多数生成式人工智能应用，主要面向成人用户，但这些应用对儿童却带来了潜在风险，包括接触不当内容及操控风险。考虑到这些风险及生成式人工智能应用的不确定性，强烈建议为通用人工智能技术设立年龄限制，以保护儿童的权利和福祉。

目前，ChatGPT的使用条款要求用户必须年满13岁，且18岁以下的用户必须获得父母或法定监护人的许可才能使用这些服务。<sup>52</sup> 这些年龄限制源自美国《儿童在线隐私保护法》（Federal Trade Commission, 1998），该法律在社交媒体尚未广泛使用、且在像ChatGPT这样的易用且强大的生成式人工智能应用程序问世之前通过。法律规定，未经父母许可，组织或个人社交媒体提供商不得为13岁以下儿童提供服务。

许多评论人士认为这一年龄门槛过低，并呼吁立法将年龄提高到16岁。欧盟《通用数据保护条例》规定，用户必须年满16岁才能在未经父母许可的情况下使用社交媒体服务。（European Union, 2016）。

随着各类生成式人工智能聊天机器人的出现，各国需要仔细考虑并公开讨论与生成式人工智能平台进行独立对话的适当年龄门槛，最低应设为13岁。此外，各国还需决定是否继续采用自我报告年龄作为年龄验证的方式，并强制规定生成式人工智能提供者在年龄验证方面的责任，以及父母或监护人在监控未成年儿童独立对话中的责任。

- **国家数据所有权与数据贫困的风险：**应采取立法措施保护国家数据所有权，并监管在其境内运营的生成式人工智能提供商。对于由公民生成且用于商业目的的数据集，应建立法规以促进互利合作，确保这类数据不会被大科技公司独占，从而流失于国家。

### 3.3.2 生成式人工智能工具提供者

生成式人工智能的提供者包括负责开发和提供生成式人工智能工具的组织和个人，以及利用生成式人工智能技术通过可编程应用程序接口（API）提供服务的实体。大多数具有影响力的生成式人工智能工具提供者都是资金雄厚的公司。必须明确指出，生成式人工智能的提供者对“设计中的伦理”负有责任，包括实施法规中规定的伦理原则。以下10个责任类别应予以涵盖：

- **人类责任：**生成式人工智能提供者应对确保遵循核心价值观和合法目的负责，尊重知识产权，践行伦理规范，同时防止虚假信息和仇恨言论的传播。
- **可信的数据与模型：**生成式人工智能提供者应被要求证明其使用的数据来源和模型方法的可信性和伦理性。他们必须采用有合法来源的数据和基础模型，并遵守相关知识产权法。此外，当模型需要使用个人信息时，必须在信息所有者知情且明确同意的情况下进行信息收集。
- **非歧视性内容生成：**生成式人工智能提供者必须禁止设计和部署基于种族、国籍、性别等受保护特征而生成偏见或歧视性内容的系统。他们应确保构建有效的“防护机制”，以阻止生成式人工智能生成冒犯性、带有偏见或虚假的内容，同时保障参与防护机制设计的人员得到保护，避免被剥削。

- **生成式人工智能模型的可解释性与透明度：**提供者应向公共治理机构解释其模型使用的数据来源、规模和类型，其预训练数据标注规则，模型用于生成内容或回应的方法或算法，以及生成式人工智能工具所提供的服务。必要时，他们应向治理机构提供支持，帮助其理解相关技术和数据。生成式人工智能生成的内容可能包含错误或有争议的回应，这应对用户透明公开。
- **生成式人工智能内容的标识：**根据人工智能辅助在线信息生成相关的法律法规，提供者需要适当且合法地标识生成式人工智能生成的论文、报告、图像和视频。例如，生成式人工智能输出的内容应明确标注为机器生成。
- **安全保障原则：**生成式人工智能提供者应确保其服务在系统生命周期内的安全、稳健和可持续性。
- **访问和使用的适用性规定：**生成式人工智能提供者应明确其服务的适用对象、使用场景和用途，帮助生成式人工智能工具的用户做出理性且负责的决策。
- **承认局限性与防止可预见风险：**生成式人工智能提供者应明确展示其系统使用方法及其输出的局限性。他们需开发技术，确保输入数据、方法和输出不会对用户造成可预见的伤害，并在发生不可预见伤害时采取应对措施。同时，必须提供指南，帮助用户根据伦理原则理解生成式人工智能生成的内容，防止用户过度依赖或沉迷于生成内容。
- **投诉与补救机制：**生成式人工智能提供者需要建立投诉收集机制和渠道，及时受理用户及公众的投诉。
- **监测和报告非法使用：**生成式人工智能提供者应与公共治理机构合作，以便于监测和报告非法使用情况。这包括人们以非法或违反伦理及社会价值观的方式使用生成式人工智能产品，如传播虚假信息或仇恨言论、生成垃圾邮件或编写恶意软件等。

### 3.3.3 机构用户

机构用户包括教育主管部门及各类院校，如大学和中小学等。这些机构负责决定是否采用生成式人工智能工具，以及应采购和部署哪些类型的生成式人工智能工具。

- **机构对生成式人工智能算法、数据和输出的审计：**实施相关机制，以尽可能监控生成式人工智能工具使用的算法、数据及其所产生的输出。这包括定期进行审计和评估，保护用户数据，以及自动过滤不当内容。
- **验证相称性并保障用户福祉：**实施国家层面的分类机制或构建机构性政策框架，以便对生成式人工智能系统及应用进行分类与验证。确保机构所采用的生成式人工智能系统符合经本地验证的伦理准则，且不会对机构的目标用户群体，特别是儿童及弱势群体，造成可预见的伤害。
- **评估并解决长期影响：**随着时间的推移，教育中对生成式人工智能工具或内容的依赖可能会对人类能力的发展产生深远影响，比如批判性思维技能和创造力。这些潜在影响应得到仔细评估并妥善处理。
- **年龄适用性：**考虑在机构内部对独立使用生成式人工智能设定最低年龄限制。

### 3.3.4 个人用户

个人用户可能包括全球所有能够使用互联网且至少能使用一种生成式人工智能工具的人。此处所使用的“个人用户”一词，主要指在正规教育机构中的个体教师、研究人员和学习者，或参与非正式学习项目的人员。

- **对生成式人工智能使用条款的认知：**用户在签署或表示同意服务协议时，应了解协议中遵守条款的义务以及协议背后的法律法规。
- **在伦理范畴内应用生成式人工智能：**用户应负责任地使用生成式人工智能，避免以可能损害他人声誉和合法权益的方式加以利用。
- **监测和报告生成式人工智能的非法应用：**当发现违反一项或多项法规的生成式人工智能应用时，用户应通知政府监管机构。

# 4. 为在教育和使用生成式人工智能制定政策框架

我们需要制定恰当的政策以规范生成式人工智能的使用，赋能教育和使用。上文所引2023年的调研数据表明，只有少数国家采用了在教育中使用人工智能的具体政策或计划。本报告上一章节概述了生成式人工智能在教育中的愿景、采用生成式人工智能所需步骤以及各利益相关方可以采取的关键要素和行动。本章节进一步提出我们可以采取的措施，以制定一致、全面的生成式人工智能教育政策框架，规范生成式人工智能在教育和使用中的使用。

《2022年人工智能与教育：决策者指南》是规范教育中使用生成式人工智能的起点（UNESCO，2022b）。该指南提出了一套全方位建议，以指导各国制定和实施关于人工智能和教育的全部政策，重点强调使用人工智能以促进优质教育、社会公平和包容。该指南中大多数建议仍然适用于在教育中使用生成式人工智能的场景，并可进一步调整，以指导制定教育中使用生成式人工智能的具体政策。为对该指南进行补充，针对规划教育和使用生成式人工智能政策，本报告提出以下八项具体措施。

## 4.1 促进包容、公平以及语言和文化多样性

在生成式人工智能的整个生命周期中，各国必须认识到包容性的重要性并确保生成式人工智能的包容性。具体而言，除非生成式人工智能工具能够被无障碍获取（不论性别、种族、特殊教育需求、社会经济地位、地理位置、流离失所状况等），而且在设计上能促进语言多样性和文化多元化（甚至促进公平），否则这些工具将不会对教育的根本挑战或实现可持续发展目标4的承诺产生帮助。为此，建议采取以下三项政策措施：

- 各国应遴选出无法负担互联网连接或数据成本，甚至是无法连接互联网的人

群，采取行动促进该人群的互联网全覆盖和数字素养提升，以消除该人群平等和无障碍地获取人工智能应用的障碍。各国应建立可持续资助机制，为残疾或有特殊需求的学习者开发和提供人工智能工具。各国应倡导使用生成式人工智能以为所有年龄、地点和背景的终身学习者提供支持。

- 各国应制定生成式人工智能系统验证标准，防止数据或算法中含有性别偏见、对边缘群体的歧视或仇恨言论。
- 各国应为生成式人工智能系统制定和实施包容性的规范，并在教育和使用领域大规模部署生成式人工智能时，实施保护语言和文化多样性的制度措施。相关规范应要求生成式人工智能提供商在训练生成式预训练转换器时纳入多语言数据，特别是当地方言或土著语言，以提高生成式人工智能处理和生成多语言文本的能力。规范和制度措施应严格禁止人工智能提供商有意或无意间删除少数民族语言或歧视土著语言使用者，并要求提供商停止推广主流语言或文化规范的系统。

## 4.2 保护人的能动性

随着生成式人工智能技术的持续发展，其对人类能动性的削弱是一个核心危机。随着使用生成式人工智能辅助写作或其他创造性活动的人群日渐庞大，该人群可能会在无意中依赖生成式人工智能，影响其心智技能的发展。虽然生成式人工智能可以用来挑战和扩展人类思维，但我们不应允许它篡夺人类思维。在我们设计和采用生成式人工智能时，保护和赋能人类始终是核心考虑因素，包括以下7个方面：

- 我们应让学习者了解生成式人工智能可能从他们那里收集的数据类型、这些数据的被使用方式，以及该使用方式对教育和更广泛的生活产生的潜在影响。
- 我们应保护学习者作为个体成长和学习的内在动力。在生成式人工智能系统日益成熟复杂的背景下，我们应加强人类对自己的研究、教学和学习方法的主权。
- 如果生成式人工智能的使用会剥夺学习者通过观察现实世界、实验等实证实践、与其他人讨论和独立的逻辑推理来发展认知能力和社交技能的机会，则我们应禁止使用生成式人工智能。
- 我们应确保学习者充分与社会互动和适当接触人类创造性产出的机会，防止学习者沉迷或依赖生成式人工智能。
- 我们应使用生成式人工智能工具以尽量减轻作业和考试的压力，而不是加剧这种压力。
- 我们应向研究人员、教师和学习者咨询他们对生成式人工智能的看法，并利用其反馈意见来决定是否以及如何在教育科研机构层面部署特定的生成式人工智能工具。我们应鼓励学习者、教师和研究人员对人工智能系统背后的方法论、输出内容的准确性以及它们可能带来的规范或教学法提出批评和质疑。
- 我们应在做出高风险决策时，主动承担责任，防止将人类责任转移给生成式人工智能系统。

### 4.3 监控并验证用于教育的生成式人工智能系统

如上所述，生成式人工智能教育系统的开发和部署在设计上应符合伦理。随后，一旦生成式人工智能系统投入使用，我们需要在其整个使用生命周期中对其进行仔细的监控和验证

证——包括其伦理风险、其教学的适当性和严谨性，以及对学生、教师和课堂/学校关系的影响。在这方面，建议采取以下五项行动：

- 我们应建立验证机制，以测试教育和研究中使用的生成式人工智能没有偏见（尤其是性别偏见）以及它们是否基于代表多样性的数据进行训练（包括性别、残疾、社会和经济地位、种族和文化背景以及地理位置方面）。
- 我们应解决知情同意这个复杂问题，特别是考虑到儿童或其他弱势学习者没有能力真正地达到与成人同样的知情同意的情况。
- 我们应审核生成式人工智能的输出是否包括虚假图片、虚假（不准确或错误）新闻或仇恨言论。如果生成式人工智能被发现产生不适当的内容，教育机构和教育工作者应该愿意并能够采取迅速而有力的行动来减轻或消除这个问题。
- 在教育或研究机构正式采用生成式人工智能应用之前，我们应对其进行严格的伦理验证（即采用设计伦理方法）。
- 我们应在教育科研机构层面决定统一采用生成式人工智能应用之前，确保这些应用不会对学生造成可预见性伤害，确保这些应用对目标学习者的年龄和能力具有教育效果和有效性，并且符合合理的教学原则（即基于相关知识领域和预期的学习成果和价值观发展）。

### 4.4 培养学习者包括生成式人工智能相关技能在内的人工智能素养

培养学习者的人工智能素养是在教育及其他领域安全、合乎道德和有意义地使用人工智能的关键。然而，根据教科文组织的数据，截至2022年初，只有约15个国家制定并实施了或正在制定政府认可的学校人工智能课程（UNESCO, 2022c）。生成式人工智能的最新发展进一步使得人们达到适当人工智能相关素养

水平成为当务之急，包括人工智能的人文和技术两方面。人们应从广义上理解人工智能是如何运作的以及生成式人工智能的具体影响。为此，我们当下迫切需要采取以下五项行动：

- 我们应致力于为学校教育、技术和职业教育与培训以及终身学习提供政府认可的人工智能课程。人工智能课程应涵盖人工智能对我们生活的影响，包括其引发的伦理问题，以及对算法和数据的适龄理解，以及正确和创造性地使用人工智能工具（包括生成式人工智能应用）的技能；
- 我们应支持高等教育和研究机构加强培养本地人工智能人才的计划；
- 我们应在培养高阶人工智能能力方面促进性别平等，建立性别平衡的人工智能专业人才网；
- 我们应对最新的生成式人工智能自动化所引起的国家和全球工作岗位变化进行跨部门预测，并根据需求的预期变化，在各级教育和终身学习系统中提高面向未来的技能；
- 我们应为可能需要学习新技能和适应新环境的老年工人和公民提供特别项目。

#### 4.5 培养教师和研究人員正确使用生成式人工智能的能力

根据2023年关于政府将人工智能用于教育的调查数据（UNESCO, 2023c），只有7个国家（中国、芬兰、格鲁吉亚、卡塔尔、西班牙、泰国和土耳其）明确表明已经或正在为教师制定人工智能框架或培训计划。只有新加坡教育部报告其建立了一个在线资源库，其核心内容是在教学中使用ChatGPT。这表明，大多数国家的教师都无法获得关于在教育中使用人工智能的结构清晰合理的培训，尤其是关于使用生成式人工智能的培训。

为了培养教师负责任地有效使用生成式人工智能，各国需要采取以下四项行动：

- 各国应根据当地验证、制定或调整指导意见，帮助研究人员和教师驾驭广泛可用的生成式人工智能工具，并指导设计新的特定领域人工智能应用。
- 各国应在使用生成式人工智能时，保护教师和研究人员的权利及其实践的价值。具体而言，各国应分析教师在促进高阶思维、组织人类互动和培养人类价值观方面的独特作用。
- 各国应确定教师所需的价值取向、知识和技能，以便有效和合乎道德地理解和使用生成式人工智能系统。使教师能够创建基于生成式人工智能的具体工具，以促进课堂学习和自身专业发展。
- 各国应动态审查教师在教学、学习和专业发展中理解和使用人工智能所需的能力，并将人工智能方面新出现的价值观、理解和技能纳入在职和职前教师培训的能力框架和方案中。

#### 4.6 促进多元意见和多元观点表达

如前所述，生成式人工智能既不理解用户提示，也不理解其所生成的回复。相反，其回答基于模型训练阶段所用（来自互联网的）数据中发现的语言模式的概率。为了解决其输出的一些基本问题，目前学界正在研究新的方法，例如将生成式人工智能与知识数据库和推理引擎连接起来。尽管如此，由于生成式人工智能的工作方式、源材料及其开发者的默认观点，生成式人工智能从本质来看在其输出中复制了主流世界观，并破坏了少数和多元意见。因此，如果人类文明要蓬勃发展，我们就必须认识到，无论生成式人工智能涉及什么主题，它都不可能成为权威的知识来源。因此，用户需要批判性地看待生成式人工智能的产出。

特别地：

- 用户应明白生成式人工智能是一个快速但经常不可靠的信息来源。虽然前文提到的一些插件和基于大语言模型的工具

旨在满足用户获取有效和最新信息的需求，但几乎没有有力的证据表明当下工具有效满足该需求。

- 我们应鼓励学习者和研究人员批判生成式人工智能所提供的回答。认识到生成式人工智能通常只是重复既定或标准的观点，其可能削弱了多元化和少数人的观点以及多元化的思想表达。
- 我们应为学习者提供充分的机会，让他们从试错、实证实验和对现实世界的观察中学习。

#### 4.7 测试本地相关的应用模型并建立不断迭代的证据库

迄今为止，生成式人工智能模型主要由来自“全球北方”的信息主导，对来自“全球南方”和当地社区的声音代表不足。只有我们通过不懈的努力，例如利用合成数据（Marwala, 2023），生成式人工智能工具才能对当地社区，尤其是全球南方社区的背景和需求保持敏感。为了探索与当地需求相关的方法，同时开展更广泛的学术合作，建议采取以下八项行动：

- 我们应确保对生成式人工智能的设计和采用进行战略规划，以主动地在教育关键环节采购和采用生成式人工智能，而非被技术倒逼，在教育非关键环节采购和采用该技术。
- 我们应激励生成式人工智能的设计者瞄准为学习者提供开放式、探索性和多样化的学习选择。
- 我们应以教育优先事项为原则来测试和推广在教育和研究中应用人工智能的循证用例，而不应追热点、神化技术或炒作技术。
- 我们应引导使用生成式人工智能来激发研究创新，包括通过利用计算能力、大规模数据和生成式人工智能的产出，为改进研究方法提供信息和启示。

- 我们应审查在研究过程中纳入生成式人工智能的社会和伦理影响。
- 我们应根据有实证依据的教学研究和方法制定具体标准，并为生成式人工智能在支持提供包容性学习机会、达成学习和研究目标以及促进语言和文化多样性方面的有效性建立实证基础。
- 有关生成式人工智能的社会和伦理影响，我们应采取逐步迭代的步骤来加强循证。
- 我们应分析大规模利用人工智能技术的环境成本（如培训GPT模型所需的能源和资源），并制定人工智能提供商应达到的可持续目标，以避免加剧气候变化。

#### 4.8 以跨部门和跨学科的方式审查长期影响

跨部门和跨学科方法对于在教育和研究中有效和合乎道德地使用生成式人工智能至关重要。只有利用各种专业知识，同时汇集多个利益相关方，才能及时发现并有效应对关键挑战，最大限度地减少长期负面影响，同时利用持续和累积的效益。因此，建议各国采取以下三项行动：

- 各国应与人工智能提供商、教育工作者、研究人员以及家长和学生代表合作，计划对课程框架和评估方法进行全系统调整，以充分利用生成式人工智能的潜力并降低其在教育和研究方面的风险。
- 各国应汇集跨部门和跨学科的专业知识，包括教育工作者、研究人员、学习科学家、人工智能工程师和其他利益相关者的代表，探索生成式人工智能对学习和知识生产、研究和版权、课程和评估以及人类协作和社会动态的长期影响。
- 各国应及时提供建议，为法规 and 政策的迭代更新提供信息。

## 5. 促进创造性地将生成式人工智能应用在教育与研究中

自ChatGPT首次推出以来，全球教育工作者既看到了其在文本生成方面的潜力，也对其可能引发的学生作弊问题表达了担忧。近年来，包括一些世界顶尖大学在内的机构纷纷表示“技术的发展已无法逆转”，认为像ChatGPT这样的工具将会持续存在，并有可能在教育环境中得到有效应用。同时，互联网上不断涌现出关于如何在教育和研究中使用生成式人工智能的建议。例如，利用生成式人工智能来产生新想法、生成多视角示例、制定课程计划和演示文稿、总结现有材料，以及激发图像创作。尽管几乎每天都有新方法在网络上涌现，但研究人员和教育工作者仍在探索生成式人工智能对教学、学习和研究的实际意义。尤其是，许多提出这些建议的人可能未能充分考虑伦理原则，还有一些想法更多是基于技术的潜力，而非真正为了满足研究人员、教师或学习者的需求。因此，本节将概述促进生成式人工智能在教育与研究中创造性应用的方法。

### 5.1 促进负责任和创造性使用生成式人工智能的制度战略

正如前文所述，为满足教学、学习和研究的需求，教育和研究机构应制定、实施并验证适当的策略和伦理框架，以负责任且合乎伦理地引导生成式人工智能系统的使用及其应用。这可以通过以下四种战略来实现：

- **培养设计生成式人工智能提示工程的能力：**除了特定的学科知识外，研究人员和教师还需要具备工程方面以及对生成式人工智能生成的提示进行批判性评估的专业知识。鉴于生成式人工智能带来的挑战是复杂的，研究人员和教师必须接受高质量的培训和支持来应对这些挑战。
- **识别书面作业中基于生成式人工智能的剽窃行为：**生成式人工智能的使用可能会导致学生把不是自己写的文本冒充为自己的成果，这是一种新型的“剽窃”手段。生成式人工智能供应方被要求使用“由人工智能生成”这样的水印来标记输出内容。同时，相关人员正在开发工具来识别由人工智能生成的材料。然而，目前很少有证据表明这些措施或工具是有效的。最直接的策略设计是通过人类的严格检测来维护学术诚信和加强问责制。长期的策略设计是让机构和教育工作者重新思考书面作业的设计，这样他们就不需要评估生成式人工智能工具比人类学习者做得更好的任务。相反，他们应该关注人类能做到而生成式人工智能和其他人工智能工具无法做到的事情，诸如同情心和创造力等人类的价值应用于复杂的现实世界挑战中。
- **伦理原则的战略设计：**确保研究人员、教师和学习者能够负责任且合乎伦理地使用生成式人工智能工具，并对其输出的准确性和有效性保持批判态度。
- **指导和培训：**为研究人员、教师和学习者提供有关生成式人工智能工具的指导和培训，确保他们理解数据标签和算法中的偏见等伦理问题，遵守关于数据隐私和知识产权的相关法规。

## 5.2 融合“以人为本和适切教学互动”的方法

在决定何时以及如何使用生成式人工智能时，研究人员和教育工作者应优先考虑人类的能动性，以及人类与人工智能工具之间负责任的、适切教学互动的方法。这包括以下五个考虑因素：

- 工具的使用应当有助于人类的需要，使学习或研究比无技术或其他替代方法更有效；
- 教育者和学习者对工具的使用应建立在他们的内在动机上；
- 使用工具的过程应由人类教育者、学习者或研究人员控制；
- 工具的选择和组织及其生成的内容应符合学习者的年龄范围、预期结果和目标知识类型（如事实性、概念性、程序性或元认知性）或目标问题类型（如结构良好或结构不良）；
- 使用工具的过程应确保人类与生成式人工智能进行互动并促进高阶思维的提升，同时人类需要对人工智能生成内容的准确性，教学或研究策略，以及对人类行为的影响等相关的决策负责。

## 5.3 共同设计生成式人工智能在教育与研究中的应用

在教育和研究中应用生成式人工智能既不应采取自上而下的推进方式，也不应受到商业的吹捧和裹挟。相反，其安全有效的使用应该由教师、学习者和研究人员共同设计。此外，还需要一个强有力的试点和评估过程，以测试不同应用的有效性和长期影响。

为了促进共同设计，加强教学上适当的互动并优先考虑人的能动性，指南提出了一个包含以下六个维度的框架：

- 适当的知识或问题领域；
- 预期结果；
- 适当的生成式人工智能工具和比较优势；
- 用户需求；
- 所需的人性化教学方法和实例提示；
- 伦理风险。

本节提供了在教师、学习者和研究人员共同设计过程中如何利用生成式人工智能来指导研究实践、协助教学、为自主学习基础技能提供辅导、促进高阶思维，并支持有特殊需求的学习者的示例。这些示例仅是生成式人工智能潜在应用领域中日益增长的冰山一角。

### 5.3.1 用于研究的生成式人工智能

生成式人工智能模型已经展示了在研究中扩展研究纲要、丰富数据挖掘并进行文献综述方面的潜力（见表3）。尽管生成式人工智能可能会生成各种使用场景，但仍需创新性地探索有潜力的研究问题及进一步挖掘相关研究成果，不断展示在研究中使用生成式人工智能的有效性和准确性。同时，研究者在研究真实世界时，必须确保使用人工智能工具不会削弱人类的能动性。

表3. 生成式人工智能在研究中应用的设计

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法 and 实例提示	可能的风险
研究大纲的人工智能助手	适用于结构良好的研究问题领域	帮助开发和回答研究问题，建议合适的方法。  <b>潜在的方向：用于研究计划的一对一教练</b>	以章节1.2提到的列表为起点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	研究人员需对相关主题有基本的理解。  研究人员应具备验证信息的能力，特别是检测虚构的研究文献引用的能力。	研究问题定义的基本思路（如目标受众、问题、背景），以及方法、预期结果和形式的设定。  示例提示：  为[主题x]写出10个潜在的研究问题，并对[研究领域y]的重要性进行排名。	需要警惕生成式人工智能编造信息（如不存在的研究出版物）的高风险，以及用户可能会因诱惑而复制粘贴人工智能生成的研究大纲，这可能会减少初级研究人员从试错中学习的机会。
生成式数据探索和文献综述助手	适用于结构不良的研究问题领域	自动收集信息，广泛探索数据，提出文献综述草稿，并自动化部分数据的解读过程。  <b>潜在的方向：用于数据探索和文献审阅的人工智能训练器</b>	以章节1.2提到的列表为起点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够恰当地满足特定的人类需求。	研究人员必须对数据分析方法和技术有扎实的知识。	对问题的逐步定义、数据范围和文献来源的界定、用于数据探索和文献综述的方法论，以及预期结果及其形式的设定。	需要警惕生成式人工智能伪造的信息、数据处理不当、可能侵犯的隐私、未经授权的分析和性别偏见。  需要警惕主流的社会规范的传播及其对替代规范和多元观点的威胁。

5.3.2 促进教学的生成式人工智能

通用生成式人工智能平台和特定教育生成式人工智能工具的使用均应通过教师与人工智能共同设计课程计划、课程包或整个课程大纲等方式来提升教师的学科理解和教学方法知识。生成式人工智能辅助的对话式教师助教或“数字教师孪生体”<sup>53</sup>是根据经验丰富的教师和图书馆数据进行预训练得到的，这些工具已经在一些教育机构中进行了测试，可能蕴含未知的潜力和未探索的伦理风险。这些模型的实际应用过程和进一步迭代仍需通过本指南中推荐的框架进行仔细审查，并在表4所示的情况下得到人类监督的保障。

表4. 生成式人工智能支持教师 and 教学应用 的共同设计

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法和实例提示	可能的风险
课程大纲或课程的设计者	关于特定教学主题的概念性知识和关于教学方法的程序性知识	协助课程体系和课堂设计过程，包括概述或扩展对目标主题关键领域的看法，定义课程结构。它还可以通过提供问题示例和评价标准（评分细则）帮助教师准备测试和考试。  <b>潜在的方向：人工智能生成的课程</b>	以章节1.2提到的列表为出发点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	无论是希望涉及程序性知识还是概念性知识，教师必须理解并明确他们希望课程、教材、课堂或测试涵盖和想要的内容，以及他们希望应用的教学理论。	向生成式人工智能提出的问题包括：建议主题的结构和事实性知识示例，针对主题或问题建议的教学方法和流程，或基于主题和格式创建课程包或课堂计划。  人类课程设计师需要验证事实知识，并检查建议的课程包是否合适。	生成式人工智能强加主导规范和教学方法的风险很高。  它可能无意中使用排斥性做法，偏向已经拥有丰富数据的群体，进一步加剧获取高质量教育机会的不平等现状，导致数据贫乏的群体处于不利地位。
生成式聊天机器人作为助教	在结构良好的问题中，跨多个领域的概念性知识	提供个性化的支持，回答问题和识别资源。  <b>潜在的方向：数字教师孪生体</b>	以章节1.2提到的列表为出发点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	它支持教师，但直接面向学习者，因此这要求学习者具备足够的先验知识、能力和元认知技能，以便验证通用人工智能的输出并识别错误信息。因此，这可能更适合高等教育阶段的学习者。	要求教师清楚地理解问题，监控对话，并帮助学习者验证生成式人工智能提供的可疑答案。	基于目前生成式人工智能模型的能力，教育机构需要保证人类监督生成式人工智能工具提供的响应，并警惕错误信息的风险。  这也可能限制学习者获取人类指导和支持的机会，妨碍师生关系的建立，这对于儿童尤其令人担忧。

5.3.3 作为一对一教练促进个性化基本技能学习的生成式人工智能

虽然在定义学习成果时，高阶思维和创造力越来越受到关注，但不可否认的是，基础技能在儿童心理发展和能力进步中仍然至关重要。在众多能力中，这些基础技能包括听力、发音、母语或外语写作，以及基本的算术、艺

术和编程技能。“操练和练习”不应被视为过时的教学方法；相反，应通过生成式人工智能技术重新改进和更新，以促进学习者个性化地巩固基础技能。如果遵循伦理和教学原则，生成式人工智能工具有潜力成为这种促进个性化练习的一对一教练，如表5所示。

表5. 生成式人工智能作为一对一教练实现自主习得语言和艺术基本技能的共同设计

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法和实例提示	可能的风险
一对一语言技能教练	包括会话练习的语言学习	通过提供反馈、纠正、母语或外语的示范，让学习者参与对话练习，提高听、说、写的能力。同时帮助学习者提升写作技能。  <b>潜在的方向：初学者一对一的语言辅导</b>	以章节1.2提到的列表为出发点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	考虑到生成式人工智能系统可能生成文化不敏感或不适合特定年龄的内容，独立对话可能需要设定年龄限制。  学习者必须具备与人工智能系统进行对话的内在动机。  学习者应能够对生成式人工智能的建议采取批判性态度，并检查其是否准确。	在使用通用的生成式人工智能平台时，教师可以引导学习者使用这些工具来请求改进建议、纠正发音或提供写作示例。  例如：  用[X]语言与我对话，帮助我不断提升。  给我一些关于[X]主题的写作建议。	需要警惕文化不敏感或语境不准确的语言，以及无意中延续的刻板印象或文化偏见。  如果没有适当的教学策略来激发学习者的内在动机，可能会限制儿童的创造力和独创性，导致程式化的写作。  同时，这也可能减少现实生活中的互动、多元观点产生、多样化表达和提升批判性思维的机会。
一对一艺术教练	在艺术领域的技术技能，如音乐和绘画。	提供艺术技巧的建议（例如，透视和色彩的提示）或音乐创作的建议（例如旋律和和弦进行）。  <b>潜在的方向：初学者的一对一美术教师</b>	以章节1.2提到的列表为出发点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	学习者必须对创作艺术或音乐有一些初步目标，具备艺术或音乐领域的关键要素的基础理解，以及分析艺术作品或音乐作品的基本能力。	人类教师应要求学习者将人工智能工具的艺术技巧与他们自己的作品进行比较。人类教师或教练必须鼓励学习者发展和运用他们的想象力和创造力，这是通用生成式人工智能无法替代的。  示例提示：  请给我一些灵感，帮助我创作关于[主题/想法]的图像。	可能会使儿童接触不当的或冒犯性的内容，这可能侵犯他们的保护权和福祉权。  生成式人工智能工具增加了抑制学习者发展想象力和创造力的风险。
一对一的编码或算术教练	具有入门级的概念性编程知识和技能。这也可能适用于基础数学的学习。	支持基础编码知识和技能自主学习，帮助学习者发现代码中的错误并提供即时反馈，以及根据问题量身定制的答案。  <b>潜在的方向：入门级的一对一编程教师</b>	以章节1.3提到的列表为出发点，评估生成式人工智能工具是否在本地可访问、是否开源、以及是否经过权威机构的严格测试或验证。  此外，进一步考虑任何特定生成式人工智能工具的优势和挑战，并确保其能够妥善满足特定的人类需求。	发现和定义问题，并设计算法来解决问题，仍然是学习编码和编程的核心内容。学习者必须具备使用编码的内在动机，并具备使用编程语言的一些基础知识和技能。	人类教师和教练应教授基础知识和技能，并激励学习者运用计算思维、编程、协作编码等方式来解决问题。  示例提示：  建议一些不寻常的编码想法。	反馈和建议的准确性仍然是一个问题，因为生成式人工智能并不总是正确的。  生成式人工智能工具存在很高的风险，可能阻碍学习者发展计算思维技能和寻找、定义有意义编程问题的能力。

5.3.4 促进探究性学习或项目式学习的生成式人工智能

如果不以促进高阶思维或创造力为目的，使用生成式人工智能工具往往会助长剽窃或产生浅层的“随机模仿”式输出。然而，鉴于生

成式人工智能模型是基于大规模数据进行训练的，它们有潜在在苏格拉底式对话中扮演一个对手的角色，或在项目式学习中担任研究助理。然而，只有通过激发高阶思维的教学/学习设计过程，才能充分发挥这些潜力，如表6所示。

表6. 生成式人工智能促进探究式学习或项目式学习应用的设计

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法和实例提示	可能的风险
苏格拉底式挑战者	劣构问题	针对先验知识，引导学习者参与苏格拉底对话，引导其发现新知识或获得更深的理解。  <b>潜在的方向：一对一苏格拉底式辩论对手</b>	从章节1.3中的列表开始，评估特定的生成式人工智能工具是否可以在本地可访问、开源、是否经过当局的严格测试和验证。  进一步考虑任何特定的生成式人工智能工具的优势和挑战，并确保它正确地满足特定的人类需求。	学习者必须达到能够独立与生成式人工智能工具进行对话的年龄。学习者还需要具备先验的知识和能力，以检查所提出的论点和信息是否准确。	人类教师可以帮助准备一系列逐步深入的问题作为示例，供学习者改编成提示语。学习者也可以从一个广泛的提示开始，例如“与我进行苏格拉底式对话，以帮助我对[主题x]采取批判性视角”，然后通过逐渐优化的提示语深化对话。	当前的生成式人工智能工具可能会生成相似或标准化的回答，限制学习者接触多样化的观点和替代性观点，从而导致回音室效应，并阻碍独立思考能力的发展。
项目式学习顾问	科学或社会研究中劣构的研究问题。	通过帮助学习者进行项目式学习来支持知识创造。这包括让生成式人工智能扮演类似于表3中所描述的研究顾问的角色。  <b>潜在的方向：一对一的项目式学习教练</b>	从章节1.3中的列表开始，评估生成式人工智能工具是否可以在本地访问、是否可以开源、是否经过当局的严格测试或验证。  进一步考虑任何特定的生成式人工智能工具的优势和挑战，并确保它正确地满足特定的人类需求。	学习者可以作为初级研究人员参与规划和实施项目式学习。学习者必须年满足够的年龄，能够独立使用生成式人工智能平台。学习者还必须具备参与自主项目式学习活动的动机和能力，以避免被动地复制粘贴生成式人工智能工具提供的答案。	根据章节5.3.1的建议，人类教师应指导学习者向生成式人工智能请求提供研究问题定义的基本思路。  个体和小组学习者可以使用生成式人工智能工具进行文献综述、收集和处理数据，并撰写报告。	没有坚实的先验知识和验证答案准确性的必要能力，学习者可能会被生成式人工智能工具提供的信息误导。  这也可能限制学习者与同伴的讨论和互动，减少合作学习的机会，潜在地影响他们的社交发展。

5.3.5 支持有特殊需求学习者的生成式人工智能

生成式人工智能模型在理论上具有潜力帮助有听力或视力障碍的学习者。当前的实践包括为聋哑学习者提供生成式人工智能支持的字幕或说明文字，以及为视力障碍学习者提供生成的音频描述。此外，生成式人工智能可以实现文本转语音和语音转文本的功能，从而帮助视力、听力或语言障碍者获取内容，并通过提问与他人互动。然而，这些功能尚未得到广泛应用。根据联合国教科文组织在2023年进行的关于各国政府在教育中使用人工智能的调查，只有四个国家（中国、约旦、马来西亚和卡塔尔）报告其政府机构已验证并推荐人工智能辅助工具，以支持有障碍学习者的包容性访问（UNESCO, 2023c）。

另一种趋势是，生成式人工智能模型正被迭代训练，以支持学习者使用少数民族语言和本土语言进行学习和交流。例如，谷歌的下一

代大语言模型PaLM 2已在覆盖数百种语言的平行数据集上进行了训练，这些数据以源文本和目标文本对的形式存在。平行多语言数据的引入旨在提升模型对多语言文本的理解和生成能力（Google, 2023b）。通过提供实时翻译、释义和自动纠错功能，生成式人工智能工具有帮助使用少数民族语言的学习者更有效地表达想法的潜力，并促进他们与不同语言背景的同龄人协作。然而，这种潜力的实现并不会自然发生，只有通过精心设计的生成式人工智能工具，才能为边缘化群体创造更多表达的机会。

最后，有人提出，生成式人工智能系统可能具备基于对话进行诊断的潜力，能够识别心理或社会情感问题，或学习困难。然而，目前几乎没有证据表明这种方法有效或安全的，任何此类诊断都需要由具备专业技能的人员进行解释。

表7. 生成式人工智能支持有特殊需求学习者应用共同设计

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法和实例提示	可能的风险
学习困难的对话诊断	这可能对因心理、社会或情感问题而面临学习困难的学习者有帮助。	利用自然语言交互来确定心理、社会或情感问题或学习困难的学习者的需求，以便为他们提供相关的支持或指导。  <b>潜在方向：为有社会或情感问题或学习困难的学习者提供一对一初级顾问</b>	除了通用的生成式人工智能工具，还可以探索生成式人工智能驱动的聊天机器人。  评估它们是否可以在本地访问、是否可以开源、是否经过当局的严格测试或验证。  进一步考虑任何特定的生成式人工智能工具的优势和挑战，并确保它正确地满足特定的人类需求。	与这一群体的学习者一起合作的教师获专家需要确保生成式人工智能系统给出的主要建议是准确的。	教师或引导者需要提供舒适的环境，让学习者参与对话，以便诊断心理、社会或情感问题，或者学习困难。	可能会无意中对学习者的特定挑战作出错误诊断，从而导致提供错误的支持。

潜在但未经证实的用途	适当的知识或问题领域	预期结果	适当的生成式人工智能工具和比较优势	用户需求	所需的人性化教学方法和实例提示	可能的风险
人工智能驱动的可访问性工具	这些工具使有听力或视力障碍的学习者能够访问更广泛的内容，从而提高他们的学习质量。	<p>通过为音频或视频内容提供生成式人工智能支持的字幕和/或手语翻译，以及为文本或其他视觉材料提供音频描述，满足学习者的访问需求，并支持他们获得特定主题的知识。</p> <p><b>潜在转型：一对一个性化人工智能驱动的语言辅助工具</b></p>	<p>除了通用的生成式人工智能工具外，还可以寻找相关的和值得信赖的人工智能驱动的字幕和音频描述生成器。</p> <p>评估它们是否可以在本地访问、是否可以开源、是否经过当局的严格测试或验证。</p> <p>进一步考虑任何特定的生成式人工智能工具的优势和挑战，并确保它正确地满足特定的人类需求。</p>	教育工作者必须帮助学习者访问并掌握如何操作生成式人工智能工具，同时确保工具的输出能够真实支持学习者，而不会加剧他们的挑战或偏见。	<p>在使用平台或工具之前，需要测试平台或工具的可访问性，以识别和修复可访问性问题。生成式人工智能工具只能提供对内容的访问，因此教育工作者和引导者应该专注于提高他们的学习质量和社会福祉。</p> <p>教育工作者和引导者需要教导学习者根据他们的能力创建语音或文本提示。</p>	<p>由生成式人工智能平台生成的、并非专门为支持视力或听力而设计的字幕或音频描述往往不准确，可能会误导有特殊需求的学习者。</p> <p>这些工具可能会无意中加剧现有的偏见。</p>
帮助被边缘化学习群体的生成式“扩音器”	对于来自少数语言或文化背景的学习者来说，这可能有助于他们表达想法，参与在线活动，以及进行协作式的社会研究。	<p>提供实时翻译、释义和自动写作纠错，以支持边缘化群体学习者使用自己的语言与来自不同语言背景的同龄人进行交流。</p> <p><b>潜在转型：边缘化学习者的包容性大语言模型</b></p>	<p>一个可供考虑的具体例子是PaLM 2。</p> <p>评估它们是否可以在本地访问、是否可以开源、是否经过当局的严格测试或验证。</p> <p>进一步考虑任何特定的生成式人工智能工具的优势和挑战，并确保它正确地满足特定的人类需求。</p>	学习者应该对对话或协作学习的主题有一定的知识储备或有意义的意见。他们应能做出负责任、无歧视性的贡献，并避免发表仇恨言论。	教师或教育工作者应设计有关社会或文化主题的研究和写作任务，或者组织在线研讨会和跨文化合作，激发学习者思考并分享观点。	<p>需要识别和纠正人工智能翻译和释义中可能导致的跨文化误解错误。</p> <p>这种使用可以为边缘化的学习者提供表达想法的机会，但不会触及数据贫困的根本原因，因此无法使人智能工具去殖民化。</p>

## 6. 生成式人工智能与未来教育和研究的关系

虽然生成式人工智能仍在迅速迭代升级，但其对教育和研究可能产生的深远影响尚未得到深刻理解。因此，我们需要重视生成式人工智能在教育和研究领域的潜在影响，并进行深入探讨和梳理。

### 6.1 未知的伦理问题

伴随着生成式人工智能工具的不断成熟，它也引发了诸多伦理问题。生成式人工智能的持续迭代升级将会引发越来越多需要系统研究的伦理议题。除了本指南的第2章和第3章外，仍需要进行更深入、更具前瞻性的分析，至少从以下五个角度进行探索：

- **公平性：**在教育领域应用的生成式人工智能可能会加剧现有的技术不平等和教育资源分配不均，进而扩大不平等性。
- **人文性：**在教育领域应用的生成式人工智能可能会减少人与人之间的交互频率，导致学习者在社会情感学习方面的弱化。
- **智力发展：**在教育领域应用的生成式人工智能可能会通过提供预设方案或缩减学习体验过程而限制学习者的主体地位。因此，生成式人工智能对年轻学习者智力发展的长期影响有待调查。
- **心理影响：**模仿人类交互行为的生成式人工智能可能对学习者产生未知的心理影响，如认知发展、情绪健康以及可能存在的操控风险。
- **隐藏的偏见与歧视：**不断更新迭代的生成式人工智能在教育应用中，可能会基于新的训练数据和模型产生潜在的新形式偏见和歧视。

### 6.2 版权与知识产权

生成式人工智能的问世正在迅速改变科学、艺术和文学作品的创作、传播和消费方式。在未取得知识产权所有者许可的情况下复制、传播或使用版权作品均侵犯了其专有权，这需要承担相应的法律后果。如，在未被授权的情况下，借助相关数据训练生成式人工智能模型被指控侵犯了所有者的知识产权。最近，人工智能生成的歌曲作品因仿照“德克雷”和“威肯”（原名阿贝尔·特斯法耶）的风格而吸引了数百万听众，然而却因为版权纠纷而遭下线（Coscarelli, 2023）。诚然，新兴的监管框架也提出相应的举措：即要求生成式人工智能开发者在调用数据训练模型之时，注意识别并保护该数据所有者的知识产权。然而，想要确定大量生成作品的所有权和原创性变得越来越困难了。可追溯性不足的问题不仅引起了人们对创作者权益保护和知识公平补偿的关注，而且也给生成式人工智能在教育中的负责任应用带来了挑战。这值得教育和研究系统深思。

### 6.3 学习内容来源

生成式人工智能工具正在改变学习内容的生成与供给方式。未来，人机对话内容将可能成为知识生产的主要来源之一。如此一来，学习者直接参与基于人类创造和验证的学习资源、教科书和课程活动的机会可能会减少。披着“权威外衣”的生成式人工智能生成文本可能会误导那些没有足够先验知识的年轻学习者，导致他们对内容的真实性缺乏足够的判断，或者不敢对这些内容提出质疑。从这一角度出发，学习者接触这些未经验证的内容是否应被视为“学习”依然是值得商榷的。

过分关注综合性的二手信息可能会使学习者错失知识建构的机会，如直接感知真实世界、从错误中学习、开展实证研究以及常识训练。同时，还可能对社会知识构建和价值观形成构成威胁，因为知识构建和价值观形成通常需要通过真实的课堂协作才能进行培养。

## 6.4 同质化反应与多样化和创造性的产出

因为生成式人工智能的生成内容代表并强调主流观点的输出，缩小了多元叙述的范围，这种知识同质化限制了多元性和创造性思维的活跃度。随着教育领域的师生对生成式人工智能的依赖性增强，他们的思维方式也会逐渐趋同并固化，削弱独立思考和自主探究的自我价值。例如，在书面作品和艺术创作中，潜在同质化会限制学习者的想象力、创造力和表现力。

因此，生成式人工智能开发者和教育者需要通力合作，思考教育生成式预训练转换器的开发潜能并将其应用于创造性思维、协作思维、批判性思维以及其他更高阶思维的培养。

## 6.5 对评价方式和学习结果的再思考

生成式人工智能对评价方式的有利影响远远超出对学习者在书面作业中作弊的直接担忧。毋庸置疑，生成式人工智能能够生成结构良好的书面作品以及颇有特色的艺术作品，它甚至还可以通过某些学科领域的知识测试。因此，我们需要重新思考清楚以下问题，即我们应该学什么、为了什么而学，以及如何评估和验证所学到东西的有效性。

在教育工作者、政策制定者、学习者和其他利益攸关方的批判性讨论中，需要考虑以下四类学习产出：

**价值观：**坚持技术人本化设计的价值观，是重新思考数字时代学习成果及其评估的核心要素。在重新审视教育目标时，必须清晰阐述技术与教育相结合的价值。通过这种规范性视角，教育工作者需持续更新学习成果及其评估与验证方法，以适应人工智能等技术在社会中的日益普及。

**基础知识和技能：**尽管生成式人工智能在某些能力领域可能超越人类，但学习者仍需掌握坚实的基础知识和技能。基础的读写能力、计算技能和科学素养依旧是未来教育的核心。随着人工智能在我们生活中的日益丰富，这些基础技能的范围和特点需要不断更新和评估。

**高阶思维技能：**学习成果需要涵盖支持基于人机协作以及利用生成式人工智能的高阶思维和问题解决所需的技能。这些技能可能包括理解事实性知识和概念性知识在巩固高阶思维基础中的作用，以及对人工智能生成内容进行批判性评估的能力。

**与人工智能合作所需的职业技能：**在那些人工智能展现出超越人类能力并在实现任务自动化的领域，学习者需要培养新技能，以便他们能够开发、操作和有效利用生成式人工智能。为了适应这一变化，学习成果和教育评价方式的重新设计必须体现由人工智能所创造的新就业岗位所需的职业技能。

## 6.6 思维过程

生成式人工智能对教育研究深远影响的基本观点体现了人类与机器的互补性。核心议题在于探索人类是否可能将基础性思维和技能掌握过程交由人工智能来承担，从而将重心转向基于人工智能辅助的更高层次的思考能力培养上。

例如，写作通常与思维结构联系在一起。现在，有了生成式人工智能的辅助，人类可以从生成式人工智能所提供的结构良好的大纲开始，而不是从零开始苦想目标、范围和大纲。一些专家称这种方式为“无脑写作”（Chayka, 2023）。随着这些新兴生成式人工智能辅助实践被广泛使用，现有的写作技能以及评价方法需要进一步优化。例如，写作学习可能更集中在规划技能和提示工程技能训练上。同时，基于更高阶的思维对生成式人工智能输出的内容进行批判性评估的能力以及与生成式人工智能协作的能力也应该被重视。

## 结语

从以人为本的设计理念出发，人工智能的构建应旨在提升和增强人类智力与社会技能，而非破坏、冲突或取代人类的能力。长期以来，人们一直期望人工智能可以进一步集成并作为人类可用工具的一部分，以更包容和更可持续的发展态势支持人类的未来行动。

为了使人工智能成为人机协作行为中可靠的一员——不管是在个人、机构还是系统层面——2021年联合国教科文组织《人工智能伦理建议书》规定：以人为本理念将根据人工智能等新兴技术（如生成式人工智能）的具体特点加以具体规定和落实。只有这样，我们才能确保生成式人工智能成为研究人员、教师和学习者值得信赖的工具。

尽管生成式人工智能被认为是服务教育和研究领域的工具，我们也需要意识到它也可能重塑这些领域已有的体系及其根本。因此，如果生成式人工智能引发了系列变革，这些转变也应当经过严格的审视，并以人本主义的方法论来指导。唯有如此，我们才能确保人工智能及在教育领域应用的各类技术充分发挥其赋能人类的潜力，共同构建一个包容性强、惠及所有人的数字化未来。

## 参考文献

- 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.)
- Coscarella, 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., Gaud, 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-a2bdeedd4> (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/rq\\_ai\\_principles-final.pdf](https://russellgroup.ac.uk/media/6137/rq_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.)

# 尾注

- 1 生成式人工智能模型向研究人员和感兴趣群体开放的时间比聊天生成预训练模型（ChatGPT）更早。比如，早在2015年，谷歌就推出了一个叫做“深度梦境”（DeepDream）的生成式人工智能模型。（<https://en.wikipedia.org/wiki/DeepDream>）。
- 2 参见 <https://chat.openai.com>
- 3 有关人工智能技术和科技以及它们之间的关系的详细解释，参见UNESCO，2022b，pp.8-10。
- 4 请注意，由于生成式人工智能仍然相对较新，不同的公司常常以不同的方式使用这些术语，有时也会使用不同的词汇来指代同一个东西。
- 5 人们担心，用于训练OpenAI GPT模型的数据可能会包含大量由GPT早期版本生成的文本。这种自我引用的循环可能会污染训练数据，进而影响未来GPT模型的性能。
- 6 需要留意的是，开发GPT系列模型的OpenAI公司至今还未对外公布GPT-4的具体细节（The Verge，2023a）。实际上，有关参数数量的说法已经被OpenAI的首席执行官否认（The Verge，2023b）。尽管如此，仍有多来源报道了这些数字（比如，可以参考E2Analyst，2023）。不管怎样，最关键的一点是，GPT-4是建立在比GPT-3庞大得多的数据集之上，并且使用的参数数量也远超GPT-3。
- 7 参见 <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 8 参见 <https://bard.google.com>
- 9 参见 <https://writesonic.com/chat>
- 10 参见 <https://yiyan.baidu.com/welcome>
- 11 参见 <https://huggingface.co/chat>
- 12 参见 <https://www.jasper.ai>
- 13 参见 <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>
- 14 参见 <https://open-assistant.io>
- 15 参见 <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>
- 16 参见 <https://you.com>
- 17 参见 <https://www.chatpdf.com>
- 18 参见 <https://elicit.org>
- 19 参见 <https://www.perplexity.ai>
- 20 参见 <https://tools.zmo.ai/webchatgpt>
- 21 参见 <https://www.compose.ai>
- 22 参见 <https://www.teamsmart.ai>
- 23 参见 <https://wiseone.io>
- 24 参见 <https://www.microsoft.com/en-us/bing>
- 25 参见 <https://www.crayon.com>
- 26 参见 <https://openai.com/product/dall-e-2>
- 27 参见 <https://dream.ai/create>
- 28 参见 <https://www.fotor.com/features/ai-image-generator>
- 29 参见 <https://www.midjourney.com>
- 30 参见 <https://creator.nightcafe.studio>
- 31 参见 <https://writesonic.com/photosonic-ai-art-generator>
- 32 参见 <https://elai.io>
- 33 参见 <https://www.gliacloud.com>

- 34 参见 <https://pictory.ai>
- 35 参见 <https://runwayml.com>
- 36 参见 <https://www.aiva.ai>
- 37 参见 <https://boomy.com>
- 38 参见 <https://soundraw.io>
- 39 参见 <https://www.voicemod.net/text-to-song>
- 40 参见 <https://openai.com/research/gpt-4>
- 41 参见 <https://www.educhat.top> <https://www.mathgpt.com>
- 42 参见 <https://www.educhat.top>
- 43 参见 <https://www.mathgpt.com>
- 44 有一些例外，例如“拥抱脸”，这是一个致力于开源人工智能开发的团队。
- 45 例如，参见Google（2023a）和OpenAI（Bass and Metz, 2023）的呼吁。
- 46 参见欧盟委员会《人工智能法案》（2021）可以找到管理人工智能的相关项目。
- 47 这项审查是基于教科文组织进行的一项调查所收集的数据，该调查向其193个成员国分发，内容涉及各国政府在教育领域使用人工智能的情况（UNESCO, 2023c）。审查还参考了经合组织的人工智能政策观察站以及斯坦福大学发布的人工智能指数报告（Stanford University, 2023），并且包括了从一群国际专家那里直接获得的第一手信息。
- 48 参见 <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- 49 通过映射分析，截至2023年4月，已有以下国家发布了国家人工智能战略：阿根廷、澳大利亚、奥地利、比利时、贝宁、巴西、加拿大、保加利亚、智利、中国、哥伦比亚、塞浦路斯、捷克、丹麦、埃及、爱沙尼亚、芬兰、法国、德国、匈牙利、冰岛、印度、印度尼西亚、爱尔兰、意大利、日本、约旦、拉脱维亚、立陶宛、卢森堡、马来西亚、马耳他、毛里求斯、墨西哥、荷兰（王国）、挪威、新西兰、阿曼、秘鲁、波兰、葡萄牙、菲律宾、卡塔尔、韩国、罗马尼亚、俄罗斯联邦、沙特阿拉伯、塞尔维亚、新加坡、斯洛文尼亚、西班牙、瑞典、泰国、土耳其、突尼斯、阿拉伯联合酋长国、英国、美国、乌拉圭和越南。此外，一些国家已将人工智能战略纳入更广泛的信息通信技术或数字战略中，包括阿尔及利亚、博茨瓦纳、哈萨克斯坦、肯尼亚、塞拉利昂、斯洛伐克、瑞士和乌干达。
- 50 根据联合国教科文组织对各国人工智能战略的快速回顾（UNESCO, 2023b），超过40个国家在其战略文件中专门设有章节讨论伦理问题。
- 51 根据联合国教科文组织对各国人工智能战略进行的快速回顾（UNESCO, 2023b），约有45个国家在其战略规划中包含了专门针对教育领域的章节。
- 52 参见 <https://openai.com/policies/terms-of-use>
- 53 在某些国家，教师会配备一个助教助教，其职责是解答个别学生提出的问题并帮助他们理解课程内容。通过使用生成式人工智能，可以创建助教的“数字孪生”以辅助学生和教师的工作。然而，这种做法也可能带来一些负面效应（比如可能会影响到课堂中的社交关系等）。



unesco

联合国教育、  
科学及文化组织

# 生成式人工智能 教育与研究应用指南

本指南的目的是辅助规划合适的法规、政策和人才培养计划，确保生成式人工智能能够真正成为助力教师、学生和研究人员的有益工具。指南详细阐述了生成式人工智能所采用的人工智能技术，并梳理了一系列产品化的GPT模型，尤其是那些在开源许可下可以获取的模型。指南还探讨了教育生成式预训练转换器的兴起——这是一类专门为教育目的而用特定数据训练的生成式人工智能模型。此外，本指南还概述了围绕生成式人工智能的一些主要争议点，包括可能加剧的数字鸿沟、观点的单一化、深度伪造技术的深化以及版权问题。本着人文主义的理念，指南建议执行对生成式人工智能工具进行监管的关键措施，如强制执行数据隐私保护，以及设定与生成式人工智能平台独立交流的适龄限制。为了引导这些工具在教育和研究中的恰当应用，指南提出了一种结合人机互动和年龄适宜性的伦理审查及教学设计的方法论。



4

优质教育



可持续发展目标