高校中的人工智能的伦理与治理

Ethics and Governance of Artificial Intelligence in Universities

祝智庭
华东师范大学
Zhu Zhiting
East China Normal University
2023.12-07

报告要点：
Contents

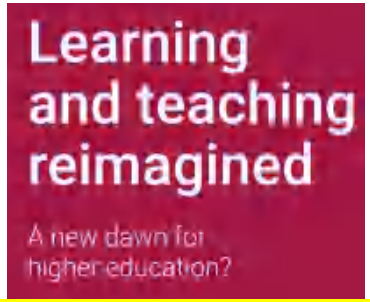
AI赋能高校教育创变的机遇与伦理风险

高校教育中AI伦理的治理生态框架

高校开展AI伦理治理的行动建议

文件案例：未来千变万化，大学应该如何应对?
Existing Documents: How should HEIs Respond?

Key opinions: - Put Digital Culture at the core of higher education culture
- Human in the loop of AI integration
- Setting guidelines and guarding fences
- Technology is to empower teaching personnel, rather than replacing them



《学习与教学再想象：高等教育的新黎明？》

《AI与未来的教与学报告》

《福布斯》：教育数字化转型六大趋势

- 将数字文化嵌入大学文化的核心。
- 长远的战略眼光。
- 探索新的经济模式。
- 混合式学习。
- 提高数字技能和信心。
- 加强对数字贫困的应对。

对教育者的七大行动建议：
1. 强调人在回路
2. 将AI模型与共同的教育愿景相结合
3. 使用现代学习原理设计
4. 优先考虑加强信任
5. 让教育者知情与介入
6. 集中于解决情境问题，增强信任和安全
7. 制定教育特定的准则和护栏

AI新技术和新学习模式为学生带来无限的可能

人工智能在教育中的应用包括个性化学习、评估课程和内容的质量以及使用智能辅导系统促进一对一辅导。

技术并不是要取代教师，而是要增强教师。

https://beta.jisc.ac.uk/reports/learning-and-teaching-reimag
https://tech.ed.gov/ai/
https://www.forbes.com/sites/danielnewman/2017/07/18/top-6-digital-transformation-trends-in-education/?sh=91b5cf12a9a2

高等教育中的人工智能：ChatGPT 文献综述和负责任实施指南

优点	缺点
学术/科学写作	信息不正确/不准确的风险
能够获得对研究的支持 - 数据收集、数据分析等。	降低用户的创造力、批判性思维、和解决问题的能力。
生成模型答案。	透明度问题
各个领域的教育效益	引文/参考文献不准确或不足引用
促进远程学习。	法律问题和版权问题
改善师生互动。	2021 年之前的受限知识
免费提供。	
提高学习参与度。	
促进与同龄人的协作学习。	
自动执行重复且耗时的任务。	
让研究人员腾出时间专注于更复杂和他们研究的重要方面。	
创建文学文本。	
有助于各种教育实践	
加强同伴沟通。	

- 提高对生成式人工智能工具的潜在用途和局限性的认识
- 使用 ChatGPT 作为辅助工具
- 将机考与亲身评估结合
- 人工智能难以复制的项目。
- 制定在高等教育中使用 ChatGPT 的道德原则和准则
- 提高作业的原创性和创造力
- 防止抄袭，要求学生带作业草稿备审
- 提供个性化反馈
- 使用形成性评估

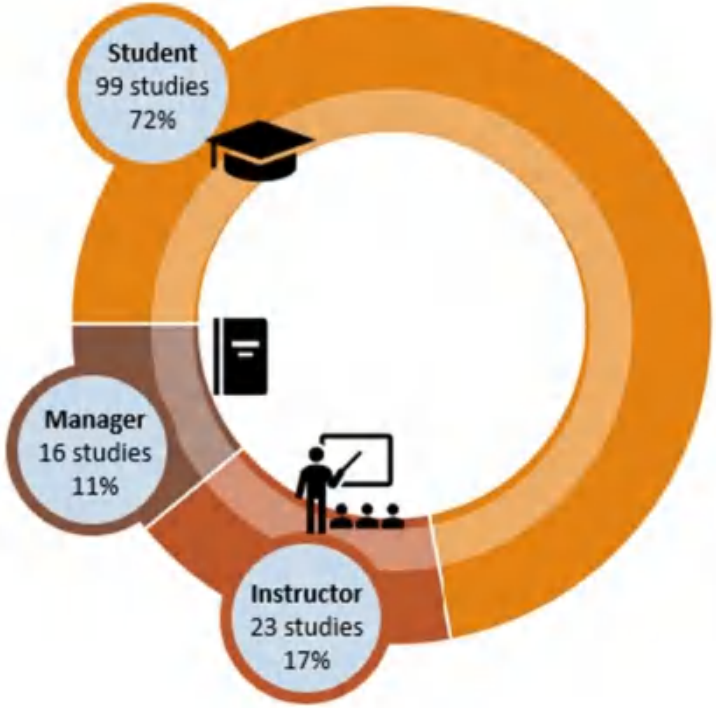
Review of ChatGPT and Guidelines for Responsible Implementation

- Knowing it’s potential and limitations
- Combine machine based assessment + experience-based assessment
- Identify projects that cannot be replicated by AI
- Make ethical guidelines and principles
- Promote originality and creativity in assignments
- Ask students to show thought process and evidences of ideation
- Use formative assessments

https://www.researchgate.net/publication/371956885_AI_in_Higher_Education_A_Literature_Review_of_ChatGPT_and_Guidelines_for_Responsible_Implementation

学者研究：AI在高等教育中的应用研究概貌

Overview of the current studies in the field of AI application in higher education



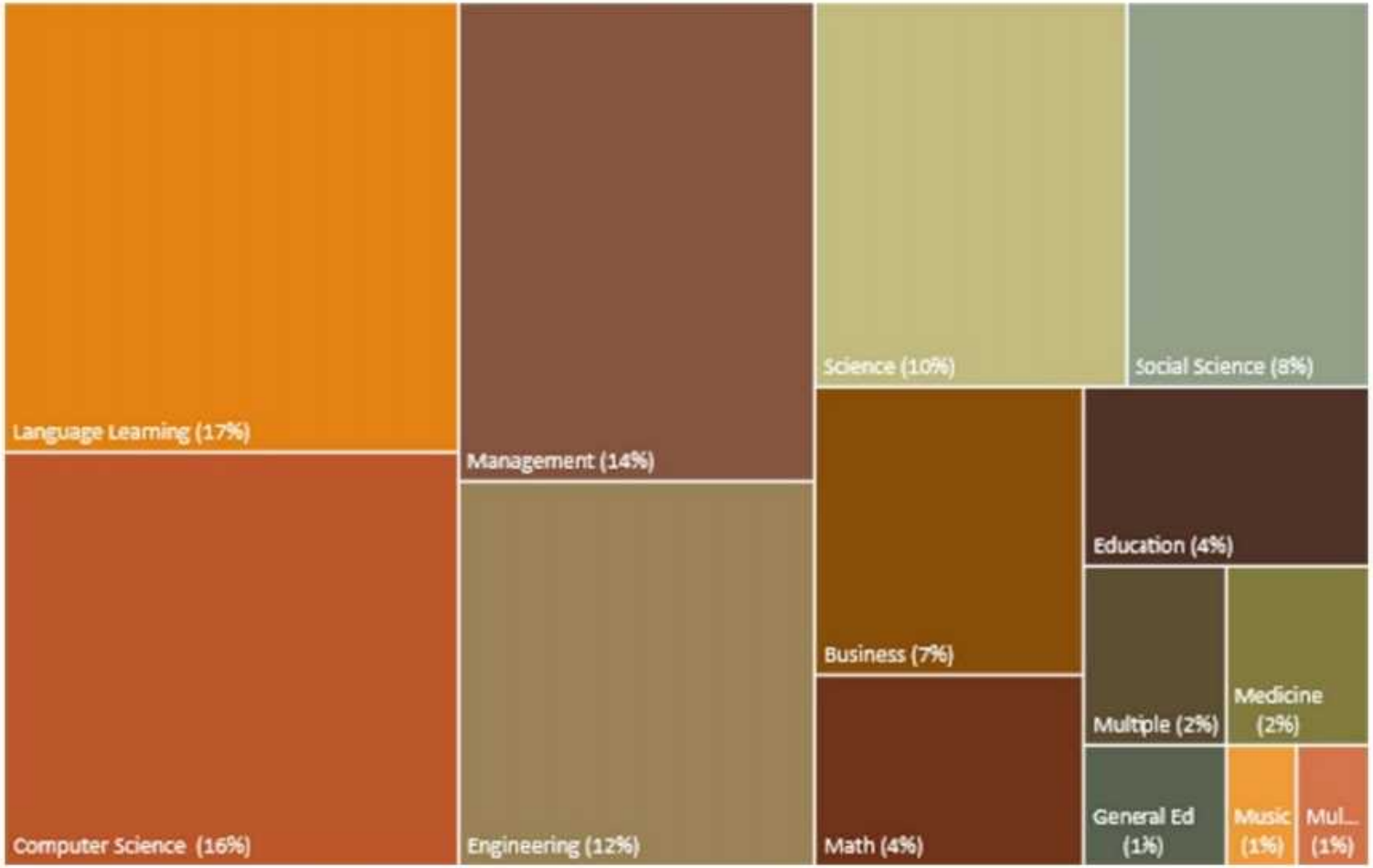


Fig. 6 Subject domains of AIEd in HE

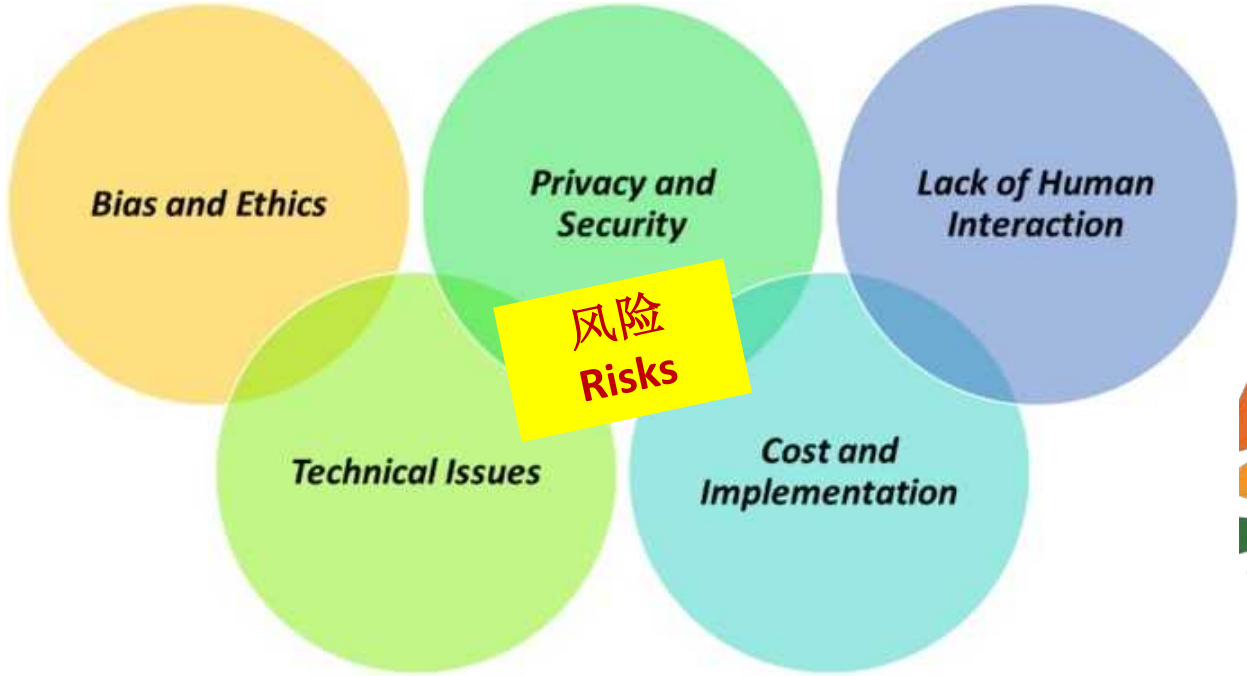
Artificial intelligence in higher education: the state of the field (springer.com)

https://usergeneratededucation.wordpress.com/2023/01/22/AIGC-with-key-students/

Education and Information Technologies (2023)

学者研究：ChatGPT赋能高教在数字时代的终身学习

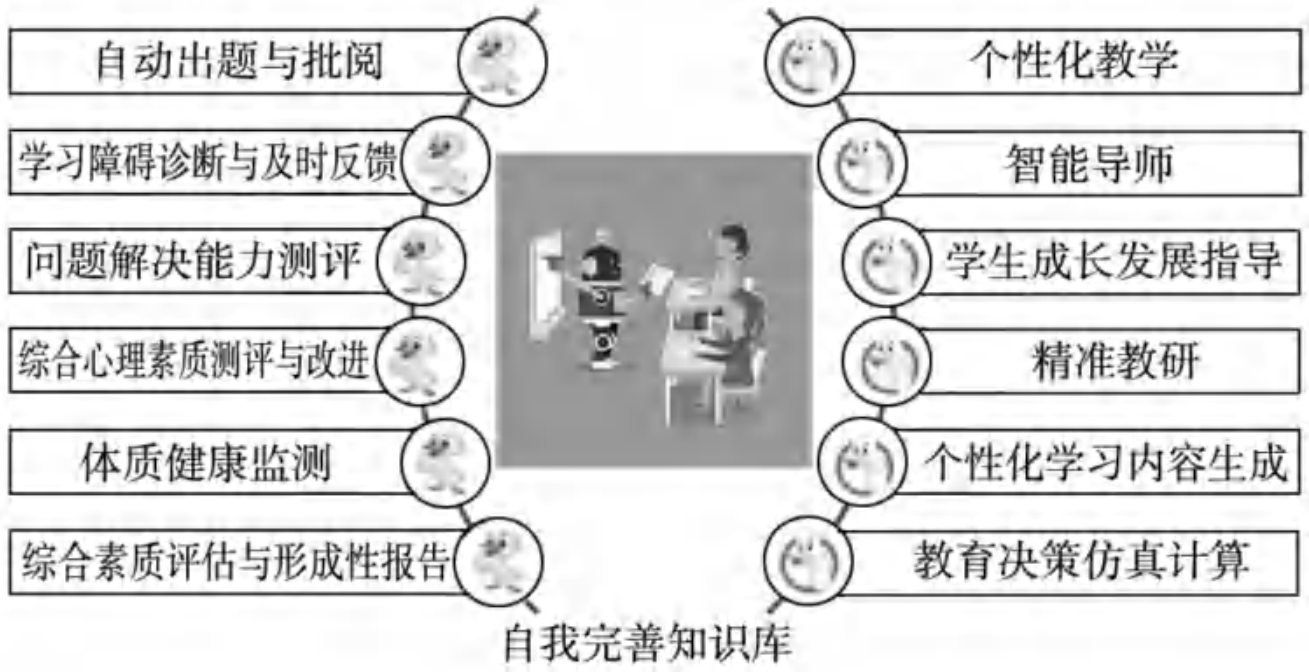
ChatGPT Empowers Lifelong Learning in the Digital Age of Higher Education



Rawas, S. ChatGPT: Empowering lifelong learning in the digital age of higher education. *Educ Inf Technol* (2023). https://doi.org/10.1007/s10639-023-12114-8

人工智能对教师的角色赋能作用：中国学者的见解

How AI empowers Teaching Personnel: YU Shengquan, 2018

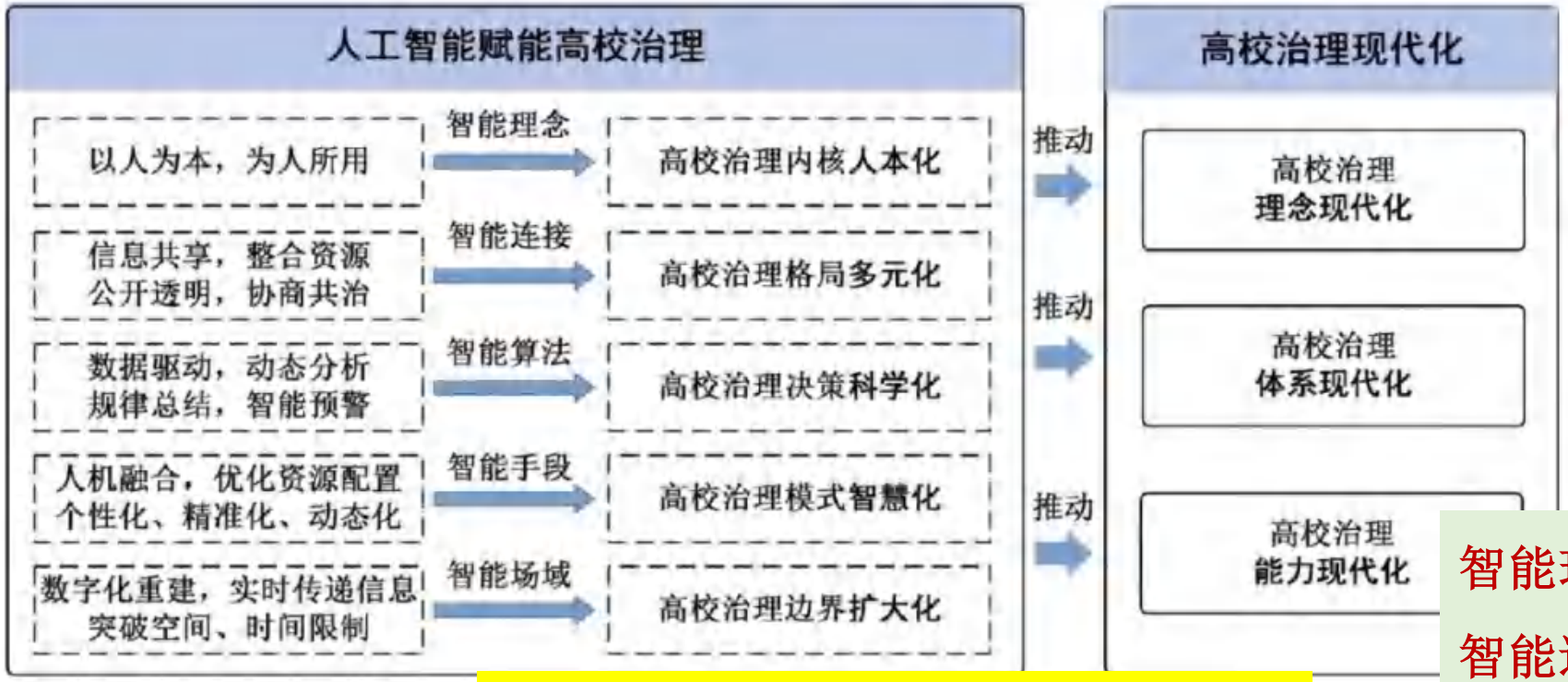


- Automated assessment questioning and evaluation
- Learning diagnosis
- Psychological tests and monitoring; physical fitness monitoring and evaluation
- Self-improving knowledge base
- Personalized learning
- Smart tutoring
- Learning analytics supported instructional research (precision teaching and research)
- Simulation on educational decisions

(余胜泉，2018)

人工智能赋能高校治理现代化

Artificial intelligence enables the modernization of university governance



人本、共治、善治

- **Human-centered**
- **Shared governance or co-governance**
- **Governance for goodness**

智能理念：以人为本的高校治理内核

智能连接：协商共治的高校治理格局

智能算法：数据驱动的高校治理决策

智能手段：人机融合的高校治理模式

智能场域：跨越时空的高校治理边界

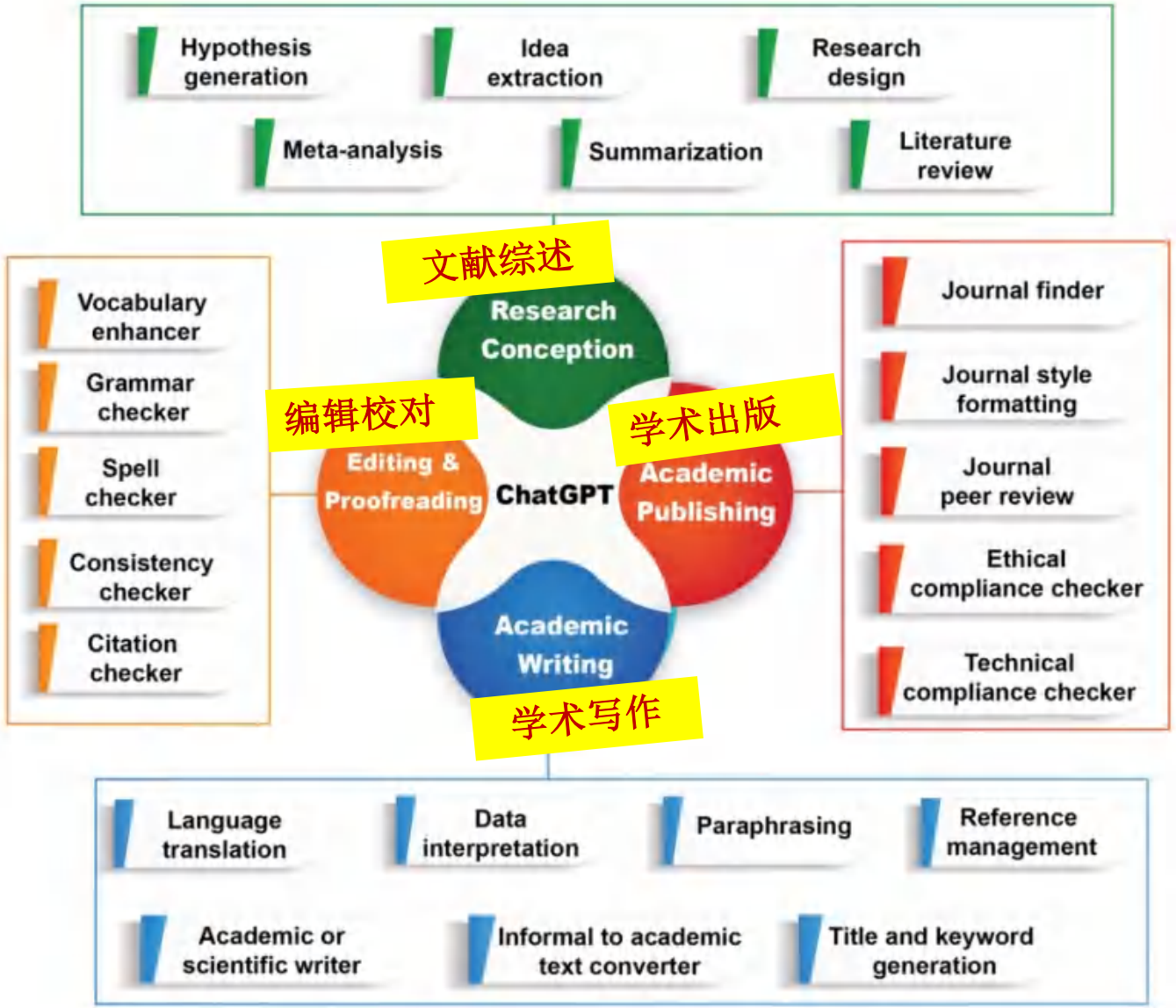
（申国昌,申慧宁，2023）

ChatGPT的角色和应用 UNESCO IESALC： ChatGPT, artificial intelligence and higher education

Role[6]	Description	Example of implementation
Possibility engine 或然引擎	AI generates alternative ways of expressing an idea	Students write queries in ChatGPT and use the Regenerate response function to examine alternative responses.
Socratic opponent 苏格拉底辩手	AI acts as an opponent to develop and argument	Students enter prompts into ChatGPT following the structure of a conversation or debate. Teachers can ask students to use ChatGPT to prepare for discussions.
Collaboration coach 协作教练	AI helps groups to research and solve problems together	Working in groups, students use ChatGPT to find out information to complete tasks and assignments.
Guide on the side 侧面导学	AI acts as a guide to navigate physical and conceptual spaces	Teachers use ChatGPT to generate content for classes/courses (e.g., discussion questions) and advice on how to support students in learning specific concepts.
Personal tutor 个人辅导	AI tutors each student and gives immediate feedback on progress	ChatGPT provides personalized feedback to students based on information provided by students or teachers (e.g., test scores).
Co-designer 协同设计者	AI assists throughout the design process	Teachers ask ChatGPT for ideas about designing or updating a curriculum (e.g., rubrics for assessment) and/or focus on specific goals (e.g., how to make the curriculum more accessible).
Exploratorium 科博馆	AI provides tools to play with, explore and interpret data	Teachers provide basic information to students who write different queries in ChatGPT to find out more. ChatGPT can be used to support language learning.
Study buddy 学习伙伴	AI helps the student reflect on learning material	Students explain their current level of understanding to ChatGPT and ask for ways to help them study the material. ChatGPT could also be used to help students prepare for other tasks (e.g., job interviews).
Motivator 激励者	AI offers games and challenges to extend learning	Teachers or students ask ChatGPT for ideas about how to extend students' learning after providing a summary of the current level of knowledge (e.g., quizzes, exercises).
Dynamic assessor 动态评价者	AI provides educators with a profile of each student's current knowledge	Students interact with ChatGPT in a tutorial-type dialogue and then ask ChatGPT to produce a summary of their current state of knowledge to share with their teacher/for assessment.

xploratorium

ChatGPT在研究过程中的可能用途
Possible use of ChatGPT in Research



人工智能赋能学术发表
AI empowers Academic Publishing

科学写作是一项需要清晰、准确和严谨的复杂任务。这不仅涉及大量的研究、分析和综合来自不同来源的信息，而且常常耗时且容易出错。然而，高级人工智能模型，例如 **ChatGPT**，能够简化学术写作和出版的过程。在学术和科学写作及出版领域，**ChatGPT** 有多种应用，包括生成假设、文献综述、安全建议、故障排除、写作技巧、改写与概括、编辑和校对、选择期刊、期刊风格格式化以及其他应用。这些工具的使用，可以大大提高科学写作的效率和质量。

Zohery, Medhat. (2023)

https://www.researchgate.net/publication/369817340_Chapter_2_ChatGPT_in_Academic_Writing_and_Publishing_A_Comprehensive_Guide

目前高等教育仍缺乏应对人工智能时代挑战的解决方案
HE still NEEDs solutions for the age of AI

- Educations needs to harness AI and identify it's potential in quality education
- Use AI as assistive tool, rather than a replacement of education
- Awareness of the risks
- Prohibiting AI is almost impossible, educators need to look into hiding reasons

- **教育工作者和机构不应该排斥或禁止人工智能的使用。相反，我们应该积极拥抱人工智能，并充分认识到其在提高教育效果方面的潜力。**人工智能有能力使教学过程更加高效，为学生提供量身定制的反馈，同时也能开启新的学习路径。然而，我们还必须确保人工智能在教育中的应用是审慎的，保障它不会对学生的批判性思维、创造力和想象力造成侵害。正确的做法是将人工智能作为一种辅助工具，以增强而非替代传统的教育方式。
- **在高等教育领域，将人工智能技术，如 ChatGPT，融入教学过程能显著促进教育效果的提升，并带来一系列挑战。认识到这些潜在的风险是至关重要的，**同时，人工智能也可作为一种强大的工具，用于增强学生和教师的学习体验。通过整合这些新兴技术和创新的教学方法，教育系统不仅能够保持其时代相关性，还能提高其教育效果，为学生应对未来挑战做好充分准备。
- **完全禁止生成式人工智能工具并非一个可行的解决方案。**如果学生感到需要作弊，这通常指向了技术本身之外的更深层次原因。因此，我们应更加关注于探究和解决导致学生不端行为的根本原因，而不是单纯限制这些技术的使用。

https://www.universityworldnews.com/post.php?story=20230321141632370

unesco

ChatGPT, artificial intelligence and higher education

AI在高教中的应用挑战及伦理影响

Academic integrity 学术诚信

The main concern that has been expressed about ChatGPT in higher education relates to academic integrity.[11] HEIs and educators have sounded alarm bells about the increased risk of plagiarism and cheating if students use ChatGPT to prepare or write essays and exams. This may have deeper implications for subjects that rely more on written inputs or information recall, areas that ChatGPT can better support.

There are also concerns that existing tools to detect plagiarism may not be effective in the face of writing done by ChatGPT. This has already led to the development of other applications that can detect whether AI has been used in writing. In the meantime, multiple HEIs around the world have banned ChatGPT due to concerns around academic integrity and others have updated or changed the way they do assessments, basing them instead on in-class or non-written assignments.

Lack of regulation 缺乏监管

ChatGPT is not currently regulated, a concern addressed by the UNESCO Recommendation on the Ethics of AI (see next section). The extremely rapid development of ChatGPT has caused apprehension for many, leading a group of over 1,000 academics and private sector leaders to publish an open letter calling for a pause on the development of training powerful AI systems.[12] This cessation would allow time for potential risks to be investigated and better understood and for shared protocols to be developed.

Cognitive bias 认知偏差

It is important to note that ChatGPT is not governed by ethical principles and cannot distinguish between right and wrong, true and false. This tool only collects information from the databases and texts it processes on the internet, so it also learns any cognitive bias found in that information. It is therefore essential to critically analyse the results it provides and compare them with other sources of information.

Gender and diversity 种族性别歧视

Concerns about gender and other forms of discrimination are not unique to ChatGPT but to all forms of AI.[14] On the one hand, this reflects the lack of female participation in subjects related to AI and in research/development on AI and on the other hand, the power of generative AI to produce and disseminate content that discriminates or reinforces gendered and other stereotypes.[15]

Accessibility 通达性

There are two main concerns around the accessibility of ChatGPT. The first is the lack of availability of the tool in some countries due to government regulations, censorship, or other restrictions on the internet. The second concern relates to broader issues of access and equity in terms of the uneven distribution of internet availability, cost and speed. In connection, teaching and research/development on AI has also not been evenly spread around the world, with some regions far less likely to have been able to develop knowledge or resources on this topic.

Privacy concerns 个人隐私

In April 2023, Italy became the first country to block ChatGPT due to privacy related concerns.[13] The country's data protection authority said that there was no legal basis for the collection and storage of personal data used to train ChatGPT. The authority also raised ethical concerns around the tool's inability to determine a user's age, meaning minors may be exposed to age-inappropriate responses. This example highlights wider issues relating to what data is being collected, by whom, and how it is applied in AI.

Commercialization 商业化

ChatGPT was created by a private company, OpenAI. Whilst the company has pledged to maintain a free version of ChatGPT, it has launched a subscription option (currently US$20/month) that offers greater reliability and faster access to new versions of the tool. The involvement of private entities in higher education is not new and calls for care and regulation if selecting AI and other tools that are run by companies dependent on making profit, may not be open source (and therefore more equitable and available), and which may be extracting data for commercial purposes.

What is Ethics in Artificial Intelligence?

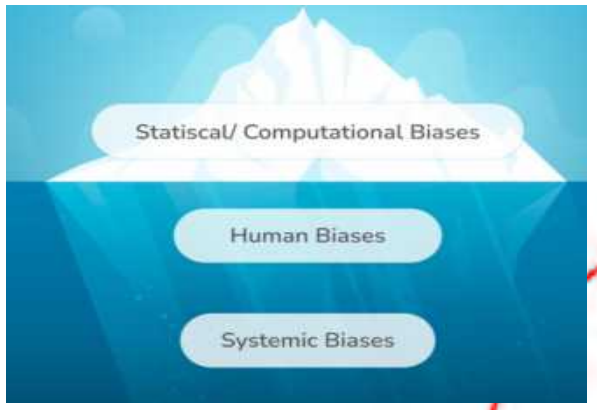
什么是人工智能的伦理?

- The notion of “artificial intelligence” (AI) is understood broadly as any kind of artificial computational system that shows intelligent behaviour, complex behaviour that is conducive to reaching goals.
- Striving to achieve goals requires AI systems to make decisions which impact human lives. E.g., autonomous cars. Taking such decisions requires rational as well as emotional understanding of how humans think and what are the values, they base their decision on.
- AI ethics **is a system of moral principles** and **techniques intended to help this system take informed decisions** ethically acceptable along with being logically optimal.

- 伦理是一系列道德原则，旨在帮助我们明辨是非。
- AI 伦理是一系列指导方针，旨在为人工智能的设计和结果提供建议。

Ethical issues are universally found in digital intelligent technologies

数智科技中伦理问题的普遍性



https://www.mdpi.com/2078-2489/13/9/440/html

独家见解 Insights：

（1）技术本身被认为是中性的，但如何使用技术一直存在伦理问题；

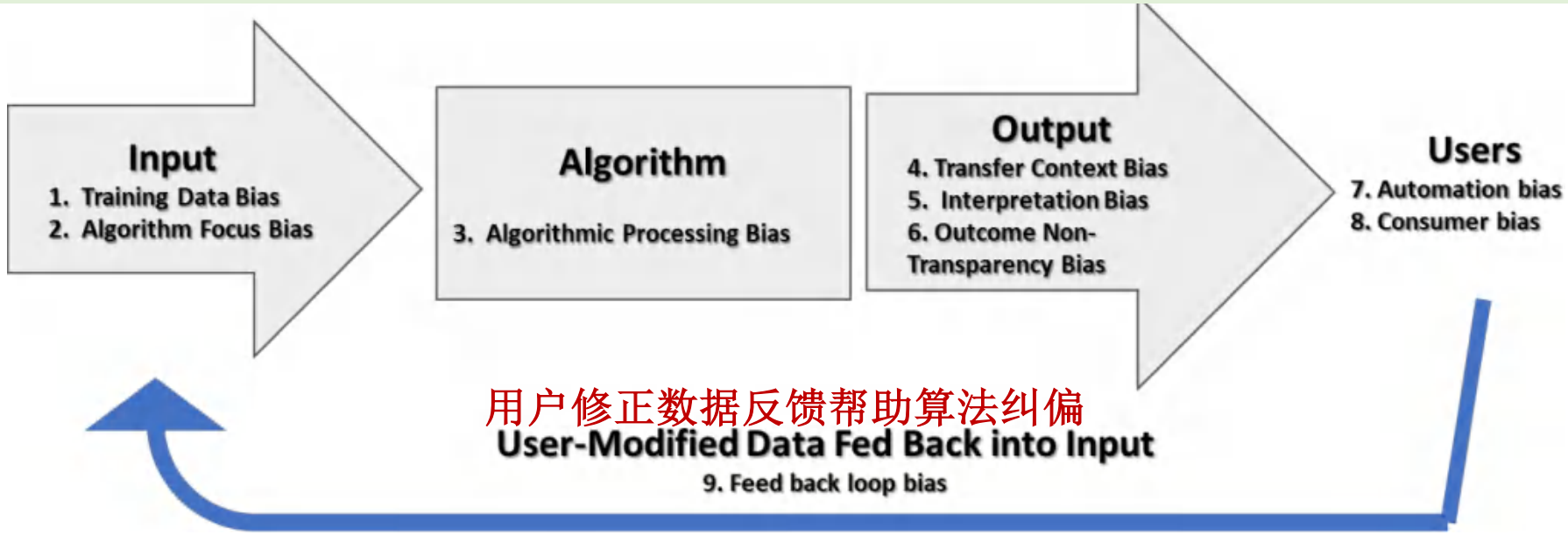
（2）软件智能算法的设计充满“心计”，所以智能技术本身就带有伦理问题。

（3）现在大多数字产品都带有智能软件，所以应该普遍关注数字道德伦理问题。

- Technology is neutral, but how to use it is problematic
- The design of algorithm was conducted with an intention, so it’s almost impossible to avoid bias in AI
- AI is becoming ubiquitous, calling for awareness the ethics of AI by general public

数字平台中的偏心算法的伦理问题不容忽视
Biased Algorithms

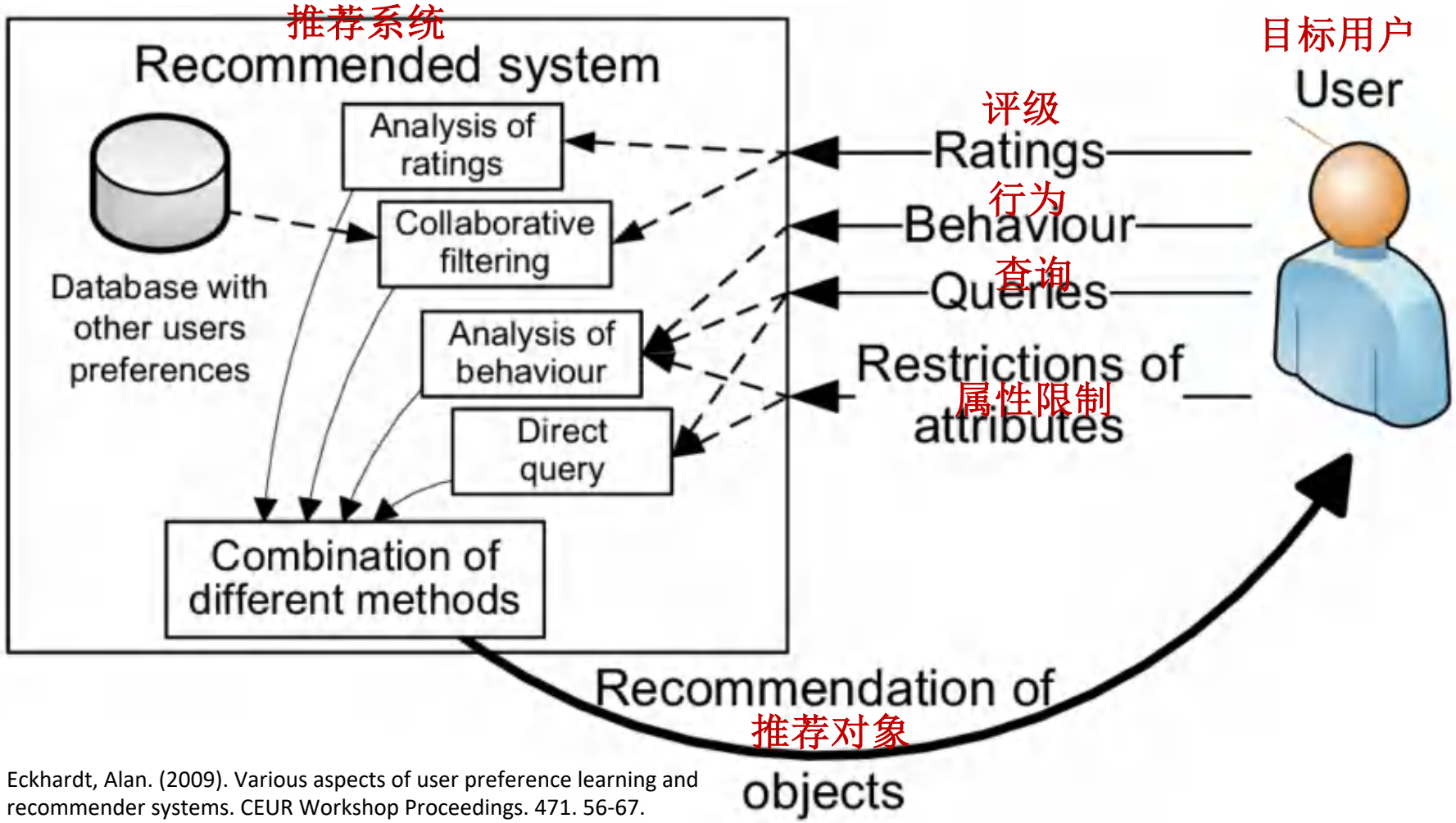
数字平台由软件中执行的算法组成。在执行这些功能时，程序“代码”就像法律一样起着构建人类活动的作用。算法和在线平台不是中立的；它们被构建为框架和驱动动作。算法“机器”是用关于人与思想中事物之间对应关系的特定理论构建的。随着机器学习等技术的广泛应用，人们的担忧越来越尖锐。对于工程师和政策制定者来说，了解算法过程中的偏差是如何发生的以及在何处发生，有助于解决这一问题。



Selena, S., & Kenney, M. (2019). Algorithms, Platforms, and Ethnic Bias: A Diagnostic Model.

推荐算法专心过度的副作用：造成认知茧房现象

Over-focused Algorithm vs Information Cocoon and Echo Chamber Effect

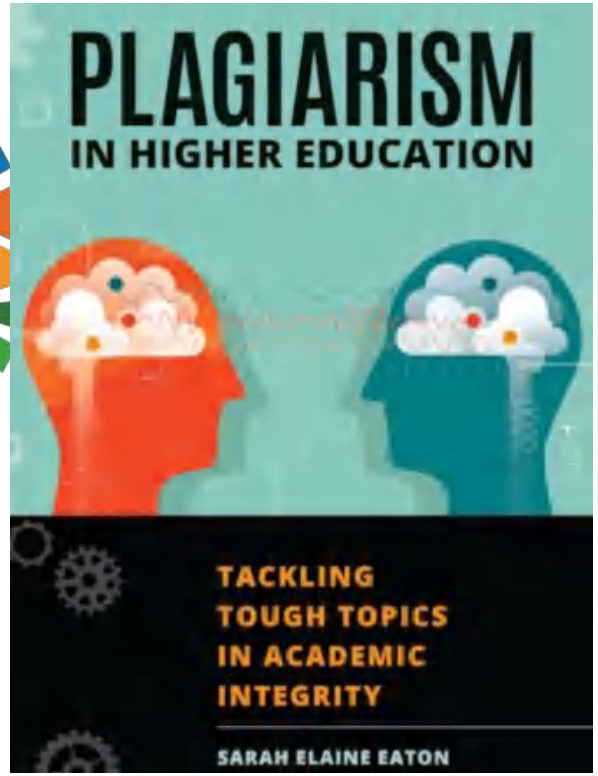


信息茧房与回声室效应

Eckhardt, Alan. (2009). Various aspects of user preference learning and recommender systems. CEUR Workshop Proceedings. 471. 56-67.

人工智能与学术诚信——后剽窃时代

Artificial intelligence and academic integrity, post-plagiarism



剽窃被定义为在缺乏适当归属的情况下复制他人的作品或思想。这通常被视为人类行为。然而，在人工智能的领域，特别是与大型语言模型如 ChatGPT 相关的工具，其生成的文本不符合传统剽窃的定义。这些人工智能程序产生的内容，尽管可能基于从多种在线资源聚合的信息，通常被视为原创。

我们正在进入一个后剽窃时代。

在这个时代，人类和技术协作创作文本成为常态。这种协作所产生的是人类与人工智能应用程序结合的创作成果。在这个时代，使用人工智能工具来增强和提升创造性成果将成为日常生活的一部分。

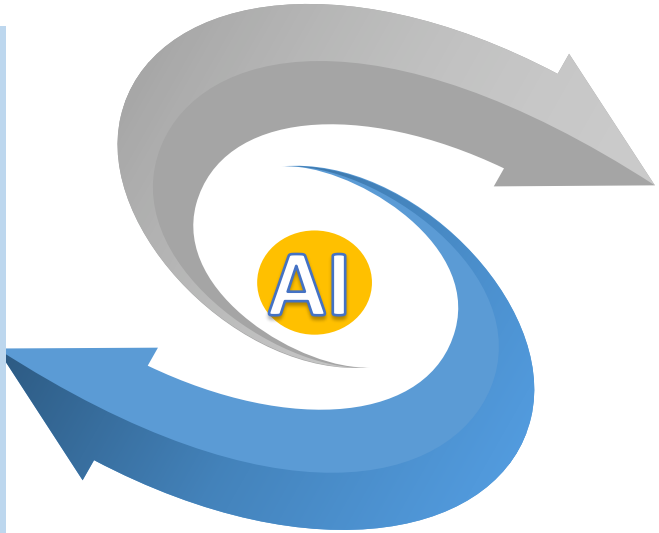
区分人类创作的文本和由机器生成的文本变得越来越困难，因为二者的输出开始交织且难以区别。这种趋势对教育领域提出了新的挑战和机遇，要求我们重新思考关于原创性和剽窃的定义。

https://www.universityworldnews.com/post.php?story=20230228133041549

人工智能在高等教育中的机遇和伦理挑战
Opportunities and Challenges for the AI ethics in HE

机遇 Opportunities

- 带来人机协同的学习和教学的新范式
- 自适应学习系统或教学平台提升学习和教学的效率
- 教师的专业发展和师生数字素养的提升
- 赋能终身教育和终身学习
- 实现高校治理的现代化



伦理挑战 Challenges

- 数据安全问题
- 算法偏见和歧视
- 知识产权问题
- 写作剽窃的认定
- 资本市场逻辑对核心价值观的冲击
- 学生创新能力的退化

- New paradigm of human-machine coordinated T&L
- Adaptive learning system and efficacy in T&L
- Teaching professional development and Digital Literacy
- Lifelong learning
- Modern governance in HEI

- Data security
- Algorithmic discrimination
- Intellectual property ownership
- Identifying plagiarisms
- Impact of capital market on core values
- Student innovation at risk

报告要点：
Contents

AI赋能高校教育创变的机遇与伦理风险

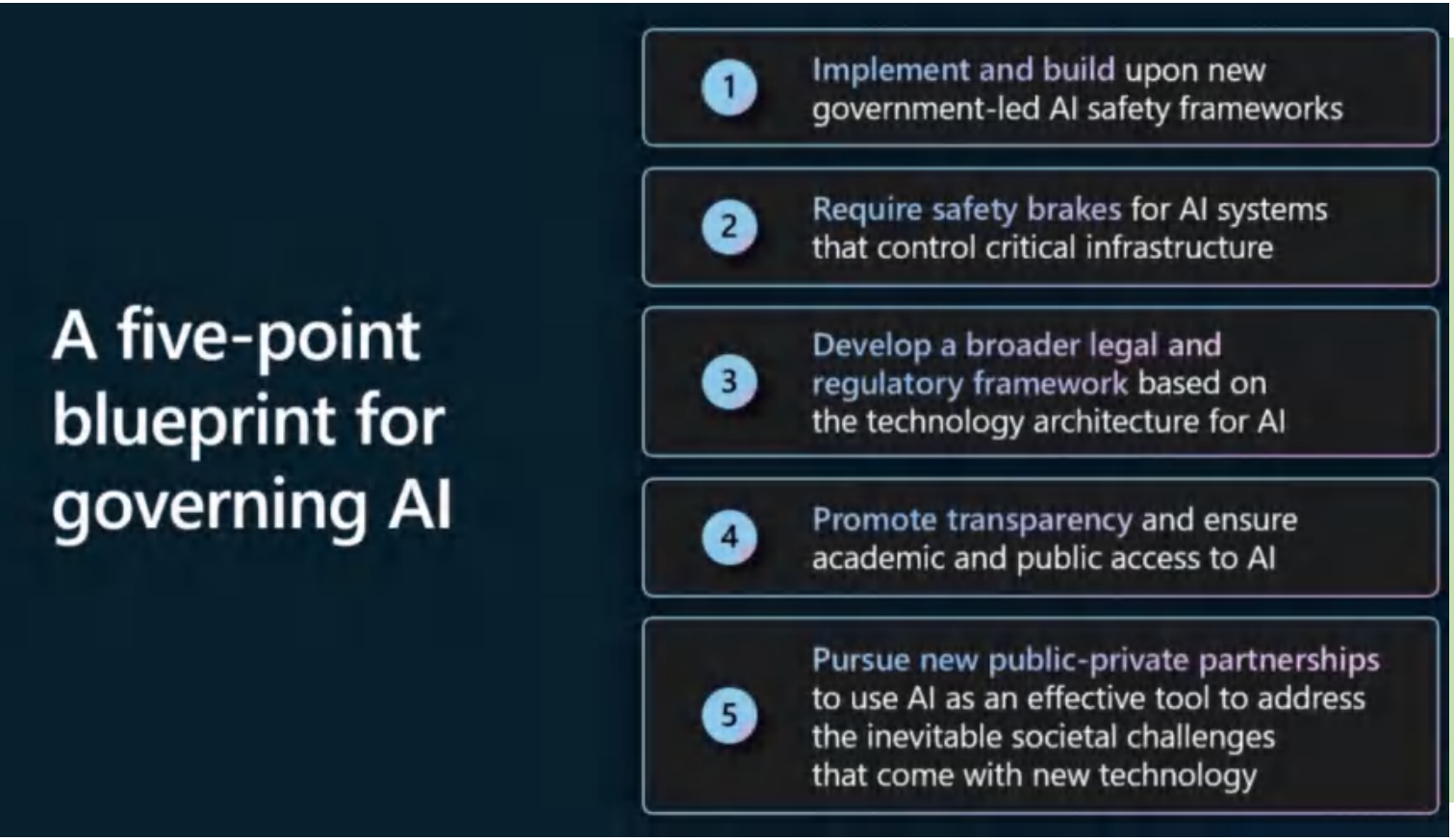
高校教育中AI伦理的治理生态框架

高校开展AI伦理治理的行动建议

有效治理人工智能：构建最佳治理框架
Building an effective governance framework for AI

《Governing AI: Blueprint for the Future》（《治理人工智能：未来蓝图》）

--Brad Smith, Microsoft



微软副总倡议AI治理五点蓝图：

1. 构建并实施以政府为主导的人工智能安全框架。
2. 必须对控制关键基础设施的人工智能系统设置有效的安全制动机制。
3. 基于人工智能技术架构制定广泛适用的法律和监管框架。
4. 提高人工智能的透明度，确保学术界和非营利组织能够获得人工智能资源。
5. 寻求新型公-私伙伴关系，用AI作为解决伴随新技术而来的不可避免的社会挑战。

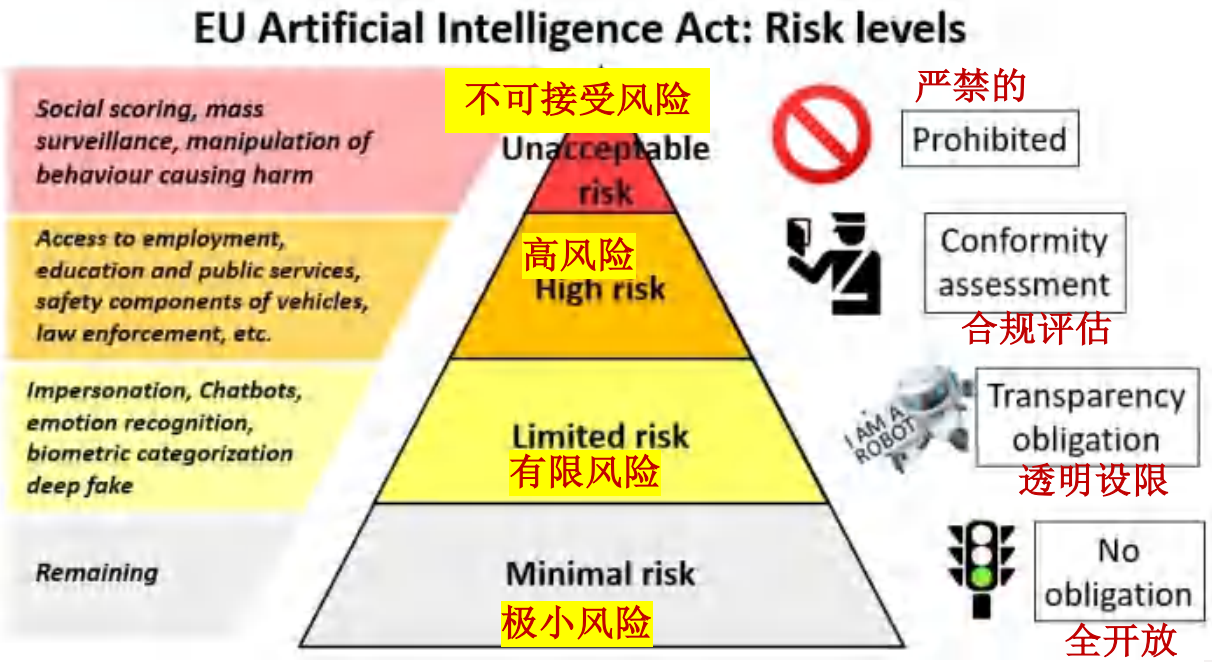
在教育实践与科研中应用AI的 SWOT 分析
SWOT Analysis for using AI in Education and Research

	对于达成目标的好处 **Helpful** to achieve goals	对于达成目标的坏处 **Harmful** to achiev goals
Internal factors 内因	**Strengths** 优势 生成似真答案 Generating plausible responses 自我改进能力 Self-improving capability 提供个性化应答 Providing personalised responses 提供实时应答 Providing real-time responses	**Weaknesses** 劣势 缺乏深度理解 Lack of deep understanding 难以评价应答质量 Difficulty in evaluating the quality of responses 存在偏差与歧视风险 The risk of biases and discrimination 缺乏高阶思维技能 Lack of higher-order thinking skills
External factors 外因	**Opportunities** 机会 Increasing accessibility of information 增加信息获取能力 Facilitating personalised learning 促进个性化学习 Facilitating complex learning 促进复杂问题解决 Decreasing teaching workload 降低教学负担	**Threats** 威胁 Lack of understanding of the context Threatening academic integrity Perpetuating discrimination in education Democratisation of plagiarism in education Declining in high-order cognitive skills

缺乏情境理解力
威胁到学术诚信
教育歧视问题固化
教育作弊问题泛化
高阶思维技能退化

依据风险等级确定监管措施

Regulatory measures are determined according to the level of risk



https://www.telefonica.com/en/communication-room/blog/a-fit-for-purpose-and-borderless-european-artificial-intelligence-regulation/

风险程度	描述	监管措施
不可接受风险	• 威胁人的安全、生计和权利，包括违背自由意志操纵人类行为的人工智能系统（如鼓励未成年人危险行为）和允许政府使用社会信用评分的系统	• 禁止； • 若违反，处以前一财年全球营业额最高6%的罚款
高风险	• 重要基础设施（如交通），可能威胁人的生命和健康； • 教育或职业培训，可能决定某人受教育的机会（如考试评分）； • 产品的安全零件（如人工智能在机器人辅助手术中的应用）； • 就业、员工管理（如招聘软件）； • 基本的私人和公共服务（如信用评分剥夺公民获得贷款的机会）； • 可能干涉人的基本权利的执法（如评判证据可靠性的系统）； • 移民、庇护和边境控制（如合适旅行文件真实性的系统）； • 运用于司法和民主程序的人工智能系统	前置审查： • 完备的风险评估系统； • 向系统提供高质量的数据集，最小化风险和歧视性结果； • 留存系统日志以确保结果的可追溯性； • 提供有关系统及其目的的所有必要信息，以供政府评估其合规性； • 向用户提供明确、充分的信息； • 采取适当的人为监督措施最小化风险（如停止按钮）； • 高水平的安全性和准确性。 全过程监督和合规评估 严格执法和处罚
有限风险	• 使用人工智能（如聊天机器人）时，使用者能意识到在与机器互动进而作出明智决定	• 实现透明公开
最低风险	• 允许自由使用人工智能的电子游戏或垃圾邮件过滤器等应用	• 不作干预

https://www.pishu.com.cn/skwx_ps/multimedia/ImageDetail?SiteID=14&type=ImageTable&ID=14573194&ContentType=MultimediaImageContentType&isHost=r

Ethical Principles for Artificial Intelligence in Education

不同权威文件中的人工智能教育伦理原则

Ethical Principles for AIED	Code	联合国教科文组织道德人工智能2020（草案）	教科文组织教育与人工智能(2021)	北京共识(2019)	经合组织(2021)	欧盟委员会(2019)	欧洲议会报告人工智能教育(2021)
	General	2020 (Draft)	Education & AI (2021)	Consensus (2019)	(2021)	Commission (2019)	Report AI Education (2021)
Governance & Stewardship	Governance & Stewardship	✔	✔	✔	✔	✔	✔
	Multistakeholder	✔	✔		✔	✔	
	Interdisciplinary planning		✔				
	International Cooperation	✔			✔		
	Monitoring & Evaluation	✔	✔		✔	✔	✔
Transparency & Accountability	Transparency	✔	✔	✔	✔	✔	✔
	Explicability					✔	
	Explainability	✔	✔		✔	✔	✔
	Responsibility	✔					
	Accountability	✔				✔	
	Auditability	✔	✔	✔		✔	✔
Sustainability & Proportionality	Sustainability	✔	✔		✔	✔	✔
	Environment	✔			✔	✔	
	Local Alignment	✔	✔	✔			
	Proportionality	✔	✔	✔		✔	
	Economy & Labour	✔	✔		✔	✔	✔
	Lifelong learning		✔	✔			
Privacy	Data Privacy	✔	✔	✔	✔	✔	✔
	Children's Privacy		✔				✔
Security & Safety	Data governance					✔	✔
	Safety	✔			✔	✔	✔
	Robustness			✔	✔	✔	✔
	Prevention of harm	✔	✔	✔		✔	✔
	Security				✔	✔	✔
Inclusiveness	Inclusiveness	✔	✔	✔	✔	✔	✔
	Accessibility		✔			✔	✔
	Diversity	✔				✔	✔
	Integrity of data					✔	✔
	Non-discriminate Data	✔				✔	✔
	Algorithms biases		✔			✔	✔
	Fairness	✔	✔	✔	✔	✔	✔
	Gender equality	✔	✔	✔		✔	✔
Human-centered AIED	Human Oversight	✔	✔		✔	✔	✔
	Human-centric/centered		✔	✔	✔	✔	✔
	Human Rights	✔				✔	
	Human Dignity	✔				✔	
	Human (Learner) Agency		✔	✔		✔	

治理与管理

透明度和问责制

可持续性

隐私

安全

包容性

以人为本

五个最常见的主要原则：

1. 透明度； **Transparency**
2. 正义与公平； **Just**
3. 非恶意（即不造成伤害）； **Harmless**
4. 负责任； **Responsible**
5. 隐私安全。 **Privacy Security**

Comprehending Ethical AI Challenges and it's Solutions

人工智能伦理挑战的解决方案

What are the Principles of Ethical AI?
Ethical AI should follow principles such as fairness, reliability, safety, privacy, security, and inclusiveness. It should provide transparency and accountability.

Social Well-Being
Fairness
Privacy Protection and Security
Reliability and Safety
Transparency and Explainability
Accountability
Value of Alignment
Governable
Human-Centered

道德人工智能应遵循公平、可靠、安全、隐私、安全、包容等原则。它应该提供透明度和问责制。

社会福祉
公平
隐私保护和安全
可靠性和安全性
问责制
可治理
以人为本

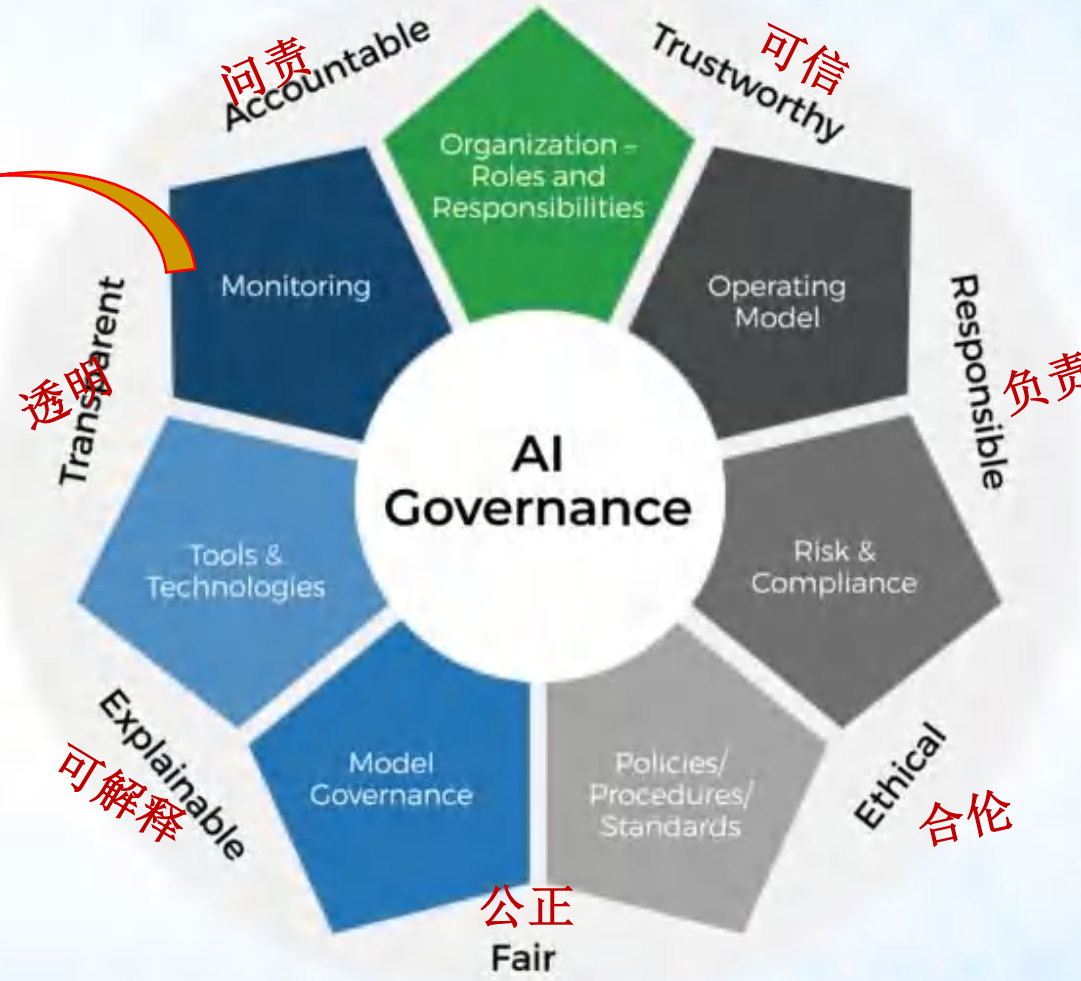
INFO-TECH RESEARCH GROUP

AI Governance

A framework for building responsible, ethical, fair, and transparent AI.

人工智能治理：构建负责任、道德、公平和透明人工智能的框架

AI Governance Framework



监督Monitoring: Monitoring compliance and risk of AI/ML systems/models in production

工具与技术Tools & Technologies: Tools and technologies to support AI governance framework implementation

模型治理Model Governance: Ensures accountability and traceability for AI/ML models

组织 **Organization:** Structure, roles, and responsibilities of the AI governance organization

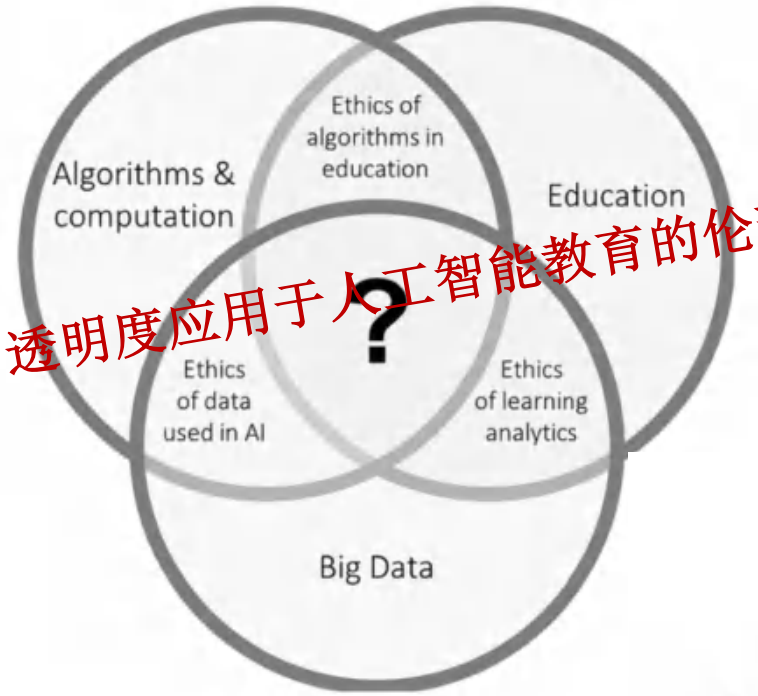
运作模型 **Operating Model:** How AI governance operates and works with other organizational structures to deliver value

风险与对标 **Risk and Compliance:** Alignment with corporate risk management and ensuring compliance with regulations and assessment frameworks

政策/过程/标准 **Policies/Procedures/ Standards:** Policies and procedures to support implementation of AI governance

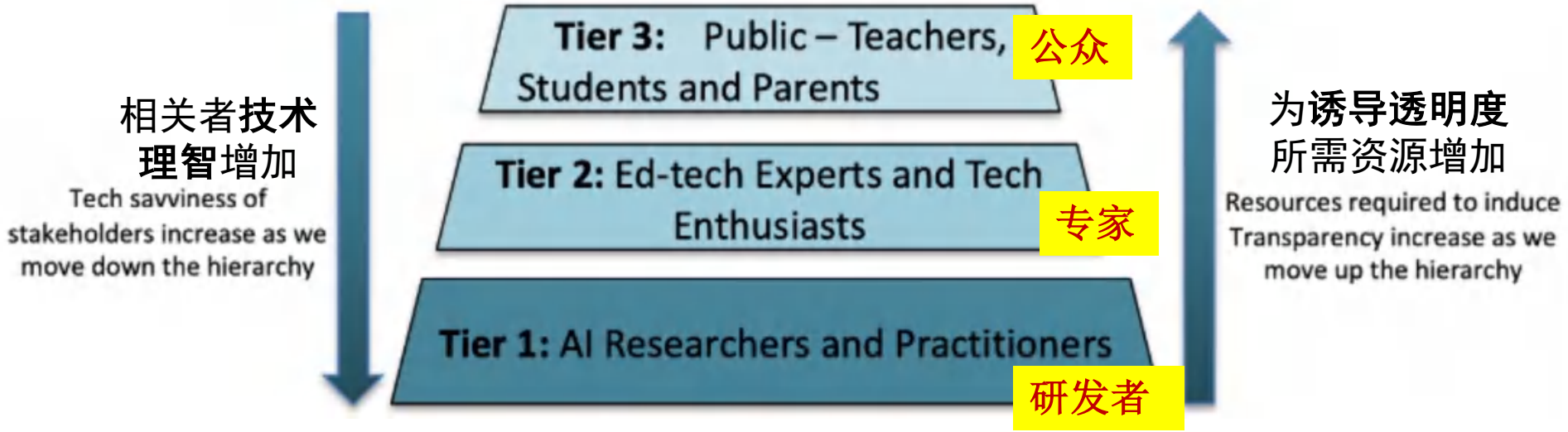
A Transparency Index Framework for AI in Education

教育领域人工智能透明度指数



透明度应用于人工智能教育的伦理框架

人工智能教育的三层透明度

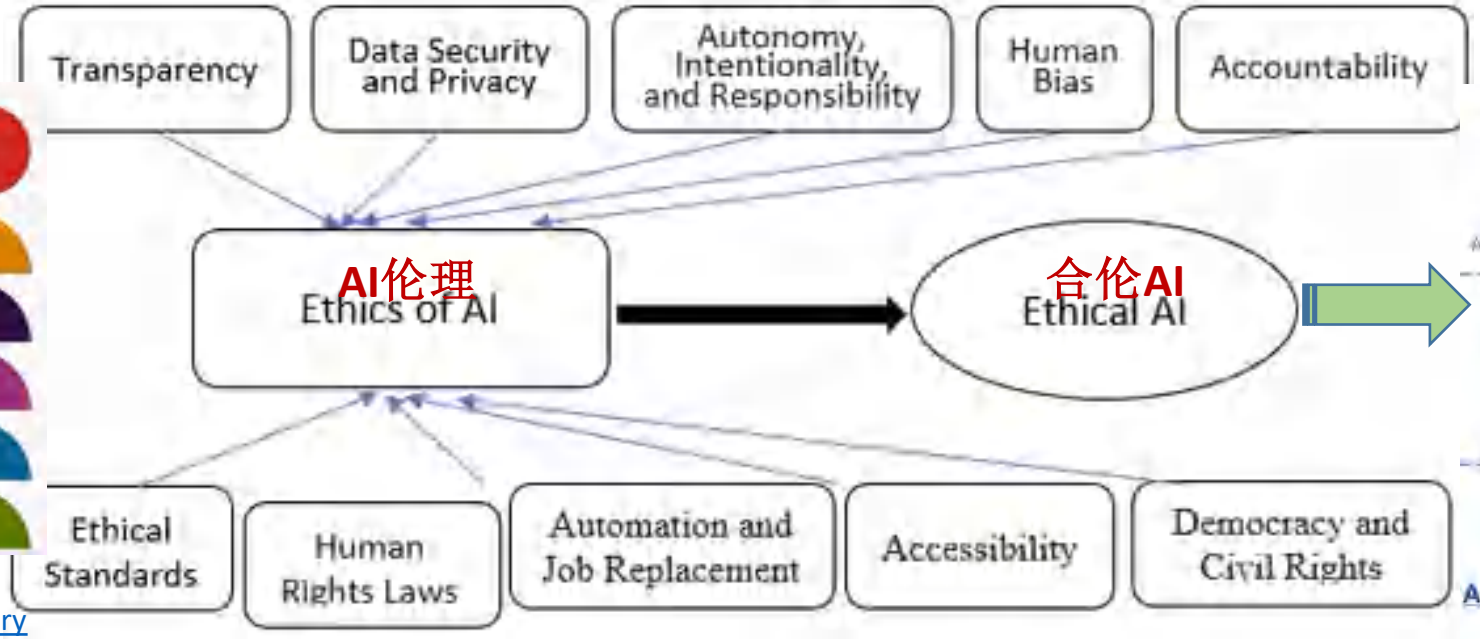


https://www.google.com.hk/url?sa=i&url=https%3A%2F%2Fosf.io%2Fbstcf%2Fdownload&psig=AOvVaw2jr8UY2e0Tvm_JbJLqpnmZ&ust=1699177482567000&source=images&cd=vfe&opi=89978449&ved=0CBEQjhxqFwoTCPinjMWHqoIDFQAAAAAdAAAAABAX

AI Ethics: Framework of building ethical AI

AI伦理建设框架：设计AI伦理导致合伦AI

[Navigating AI Ethics in the 21st Century Tech Landscape (aiworldschool.com)]

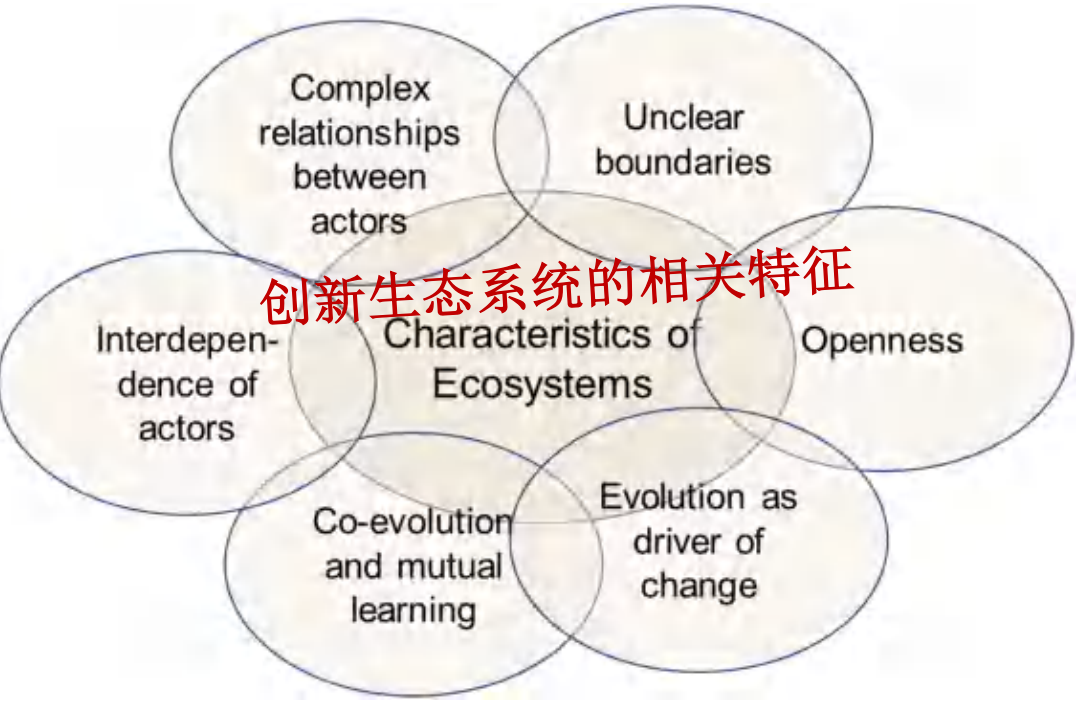


https://www.euriun.com/ai-guidelines/

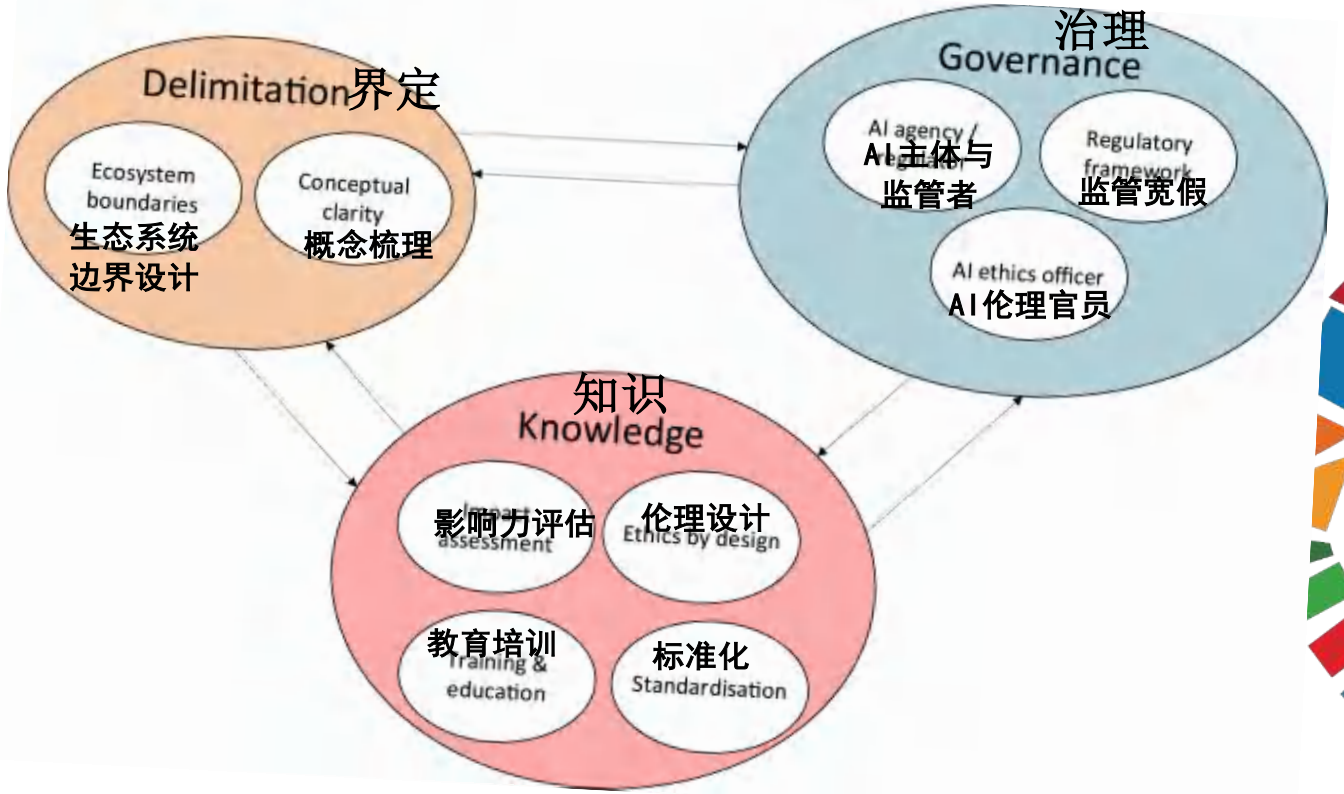
Siau, Keng & Wang, Weiyu. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. Journal of Database Management. 31. 74-87. 10.4018/JDM.2020040105.

Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence

负责任的创新生态系统：生态协同概念用于AI的合伦意蕴



人工智能创新生态系统可采取的干预措施

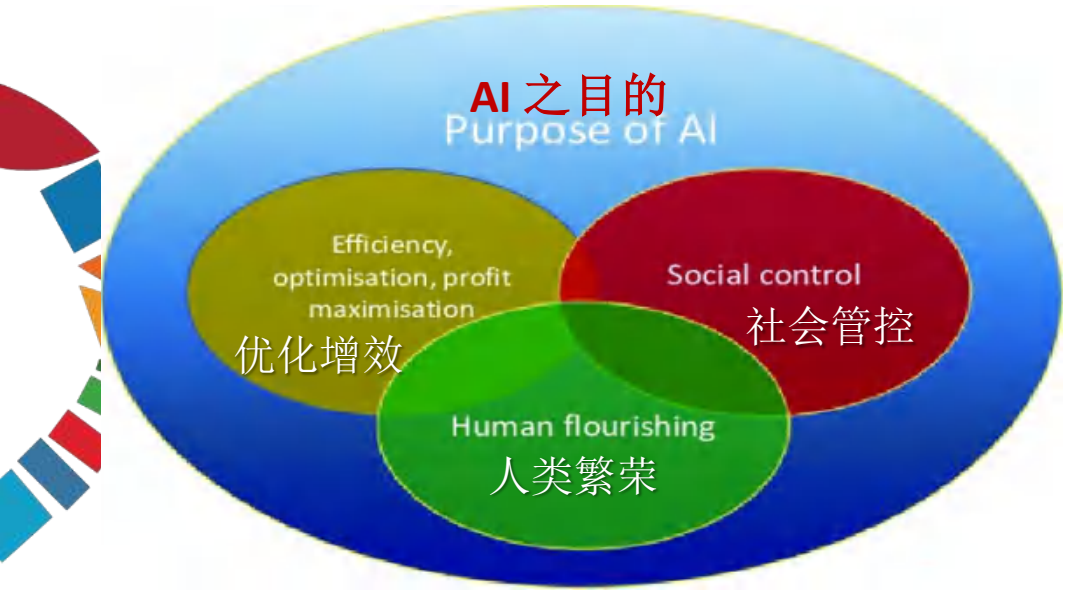


linkinghub.elsevier.com/retrieve/pii/S0268401221001341

https://www.sciencedirect.com/science/article/pii/S0268401221001341

Artificial Intelligence for a Better Future：
An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies

AI使未来更美好：AI和新兴数字技术伦理的生态系统视角



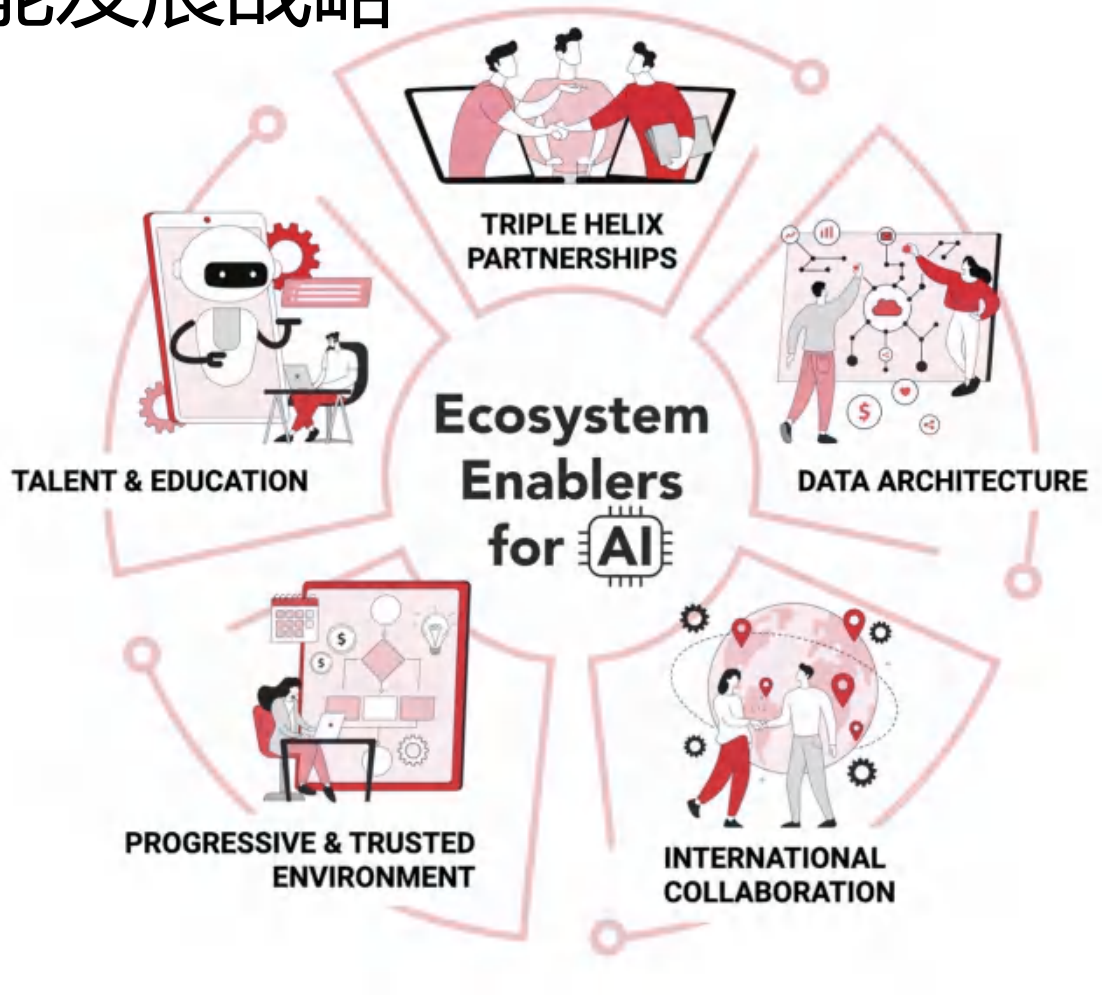
Stahl, Bernd. (2021). Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies.

Singapore： National Artificial Intelligence Strategy

新加坡生态化人工智能发展战略

1.Triple helix partnerships between the research community, industry and Government enables the rapid commercialisa of fundamental research and deployment of AI solutions
2.Talent and ed
homegrown ta
roles and helps
economy.
3.Data architec
quality dataset
4.A **progressive**
bedding, devel
5.International
development of AI with multi-nationals researchers, businesses and governments.

1.研究界、行业和政府之间的**三螺旋合作伙伴关系**使**基础研究和人工智能解决方案的部署能够快速商业化。**
2.人才和教育满足了在与人工智能相关的整个工作岗位上培养本土人才的需求，并帮助新加坡人为未来的人工智能经济做好准备。
3.数据架构可以快速、安全地访问各个部门的高质量数据集。
4.渐进**且值得信赖的环境**对于测试、开发和部署人工智能解决方案非常重要。
5.与跨国研究人员、企业和政府开展**国际合作，推动和支持人工智能的可持续发展。**



National AI Strategy (smartnation.gov.sg)

报告要点：
Contents

AI赋能高校教育创变的机遇与伦理风险

高校教育中AI伦理的治理生态框架

高校开展AI伦理治理的行动建议

United Nations
Educational, Scientific and
Cultural Organization

Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence

UNESCO: 关于人工智能伦理治理的建议初稿

Distribution: limited

SHS/BIO/AHEG-AI/2020/4 REV.2
Paris, 7 September 2020
Original: English

OUTCOME DOCUMENT:

FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

针对人工智能设计、开发和应用，专家组初步讨论了以普世伦理准则和以人权为基础的一套价值观、基本原则和推荐的政策行动。专家们还强调了以下问题：

• 低收入国家的关切；	- Concerning Low-income countries
• 当代和后世的福祉；	- Well-being of present and future generations
• 人工智能对环境的影响；	- AI and environmental impact
• 《2030年可持续发展议程》；	- SDG and 2030 agenda
• 性别和其他偏见；	- Gender and other biases
• 国家间和国家内部的不平等；	- inequity within- and/or inter-country
• 不让任何人掉队。	- No one left behind

https://unesdoc.unesco.org/ark:/48223/pf0000373434

中国网信办发布《全球人工智能治理倡议》2023.10.18.

Cyberspace Administration of China releases Global Artificial Intelligence Governance Initiative

- 发展人工智能应坚持“以人为本”理念
- 尊重他国主权，严格遵守他国法律，接受他国法律管辖。
- 发展人工智能应坚持“智能向善”的宗旨
- 发展人工智能应坚持相互尊重、平等互利的原则
- 推动建立风险等级测试评估体系，实施敏捷治理，分类分级管理，快速有效响应。

- 逐步建立健全法律和规章制度，保障人工智能研发和应用中的个人隐私与数据安全
- 坚持公平性和非歧视性原则坚持伦理先行，建立并完善人工智能伦理准则、规范及问责机制
- 坚持广泛参与、协商一致、循序渐进的原则，
- 积极发展用于人工智能治理的相关技术开发与应用
- 增强发展中国家在人工智能全球治理中的代表性和发言权

- Human-centered
- Sovereignty
- Intelligence for good
- Monitoring and evaluation, prompt response
- Scaled and classified actions
- Privacy Security
- Ethical terms and accountability mechanism
- Extensive consultation and cooperation
- Development of technology for governance
- Raising representations of developing countries

https://baike.baidu.com/item/%E5%85%A8%E7%90%83%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD%E6%B2%BB%E7%90%86%E5%80%A1%E8%AE%AE/63608167?fr=ge_ala

中华人民共和国中央人民政府

中国政府发布《生成式人工智能服务管理暂行办法》

Interim Measures for the administration of generative artificial intelligence services

国家互联网信息办公室
中华人民共和国国家发展和改革委员会
中华人民共和国教育部
中华人民共和国科学技术部
中华人民共和国工业和信息
中华人民共和国公安部
国家广播电视总局
令
第15号

第四章 监督检查和法律责任

第十六条 网信、发展改革、教育、科技、工业和信息化、公安、广播电视、新闻出版等部门，依据各自职责依法加强对生成式人工智能服务的管理。

国家有关主管部门针对生成式人工智能技术特点及其在有关行业和领域的服务应用，完善与创新发展相适应的科学监管方式，制定相应的分类分级监管规则或者指引。

第十七条 提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。

第十八条 使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报。

第十九条 有关主管部门依据职责对生成式人工智能服务开展监督检查，提供者应当依法予以配合，按要求对训练数据来源、规模、类型、标注规则、算法机制机理等予以说明，并提供必要的技术、数据等支持和协助。

参与生成式人工智能服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的国家秘密、商业秘密、个人隐私和个人信息应当依法予以保密，不得泄露或者非法向他人提供。

第二十条 对来源于中华人民共和国境外向境内提供生成式人工智能服务不符合法律、行政法规和本办法规定的，国家网信部门应当通知有关机构采取技术措施和其他必要措施予以处置。

第二十一条 提供者违反本办法规定的，由有关主管部门依照《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国科学技术进步法》等法律、行政法规的规定予以处罚；法律、行政法规没有规定的，由有关主管部门依据职责予以警告、通报批评，责令限期改正；拒不改正或者情节严重的，责令暂停提供相关服务。

构成违反治安管理行为的，依法给予治安管理处罚；构成犯罪的，依法追究刑事责任。

UNESCO's Recommendation on the Ethics of Artificial Intelligence

Actionable policies
可实行的政策

https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

关于高教AI伦理建设的行动建议
Actionable Suggestions for AI ethics Building in Higher Education

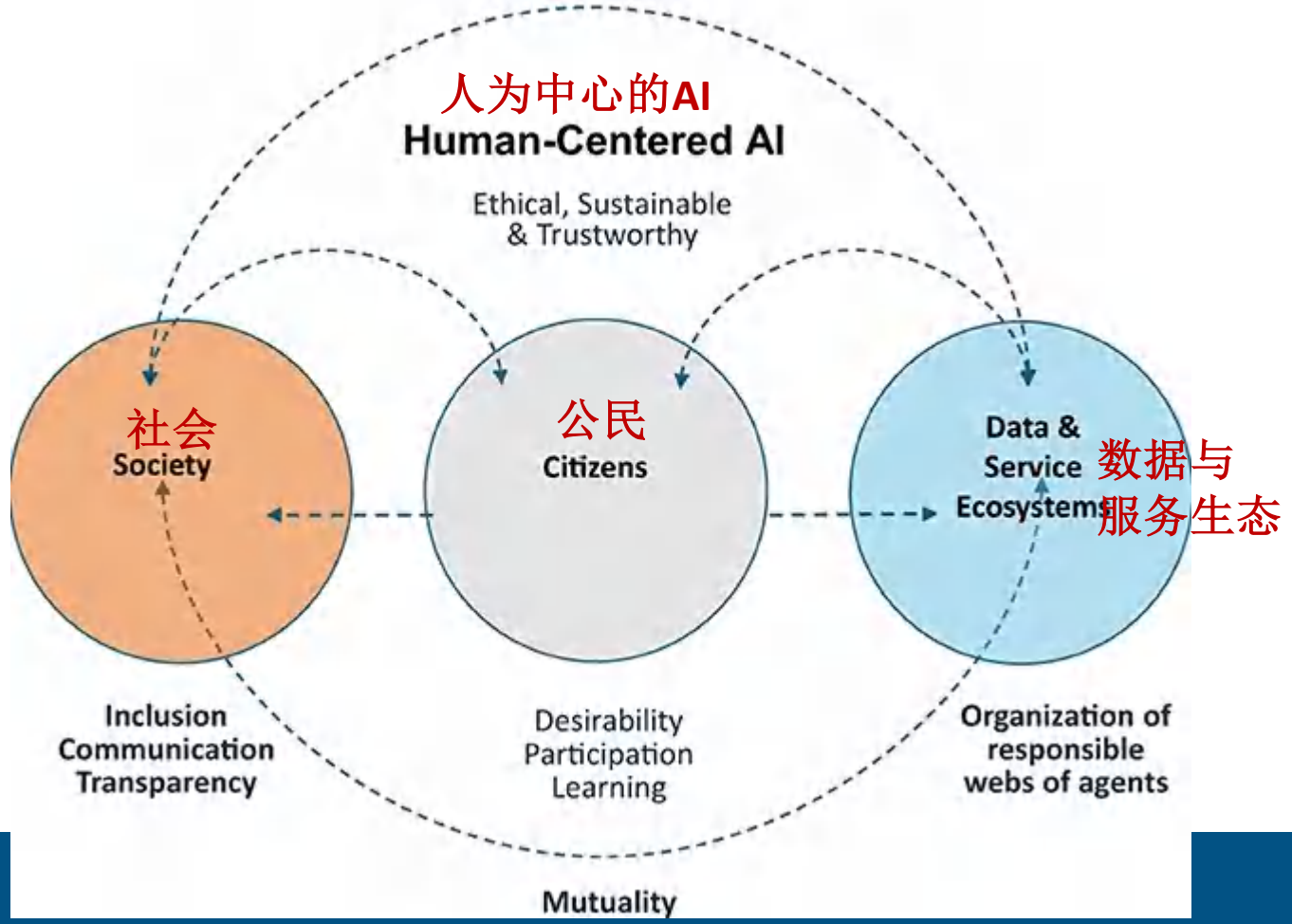
1. 坚持人本主义AI理念与实践
2. 建立AI渗透的高教伦理治理框架
3. 建立高教AI政策教育框架
4. 开展AI教育与提供AI伦理的微专业/微证书
5. 高教要为建设人技共善的AI伦理体系做贡献

1. Human-centered belief and practices
2. AI-integrated HE ethics and governance framework
3. AI policy guideline for HE
4. AI education, micro-degree and micro-certifications
5. HE's active participation in reciprocal AI ethics system

建议 1
Recommendation 1

Human-centricity in AI governance: A systemic approach

人工智能治理的系统性人本方略



https://www.frontiersin.org/articles/10.3389/frai.2023.976887/full

人本AI研究在行动

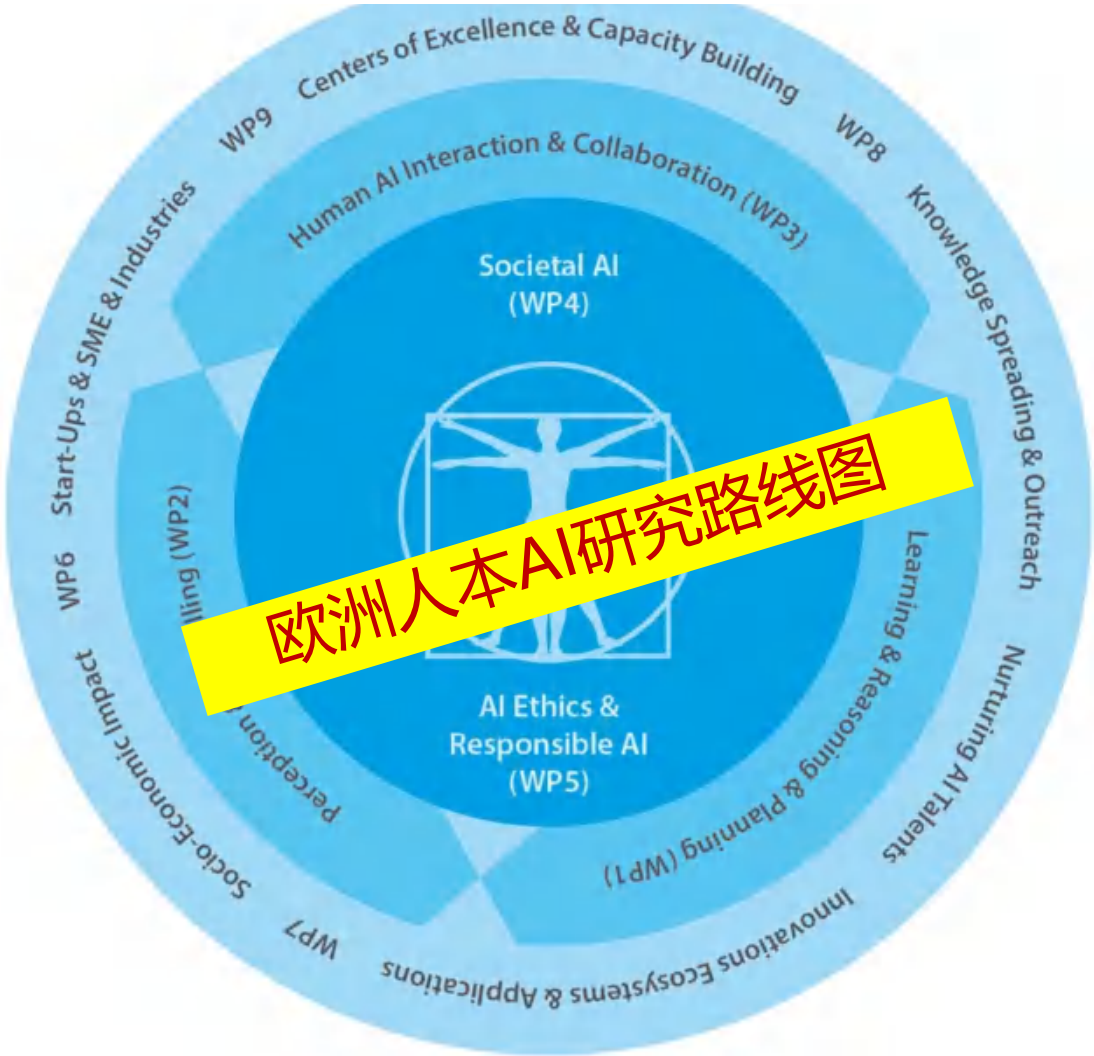
Stanford University
Human-Centered
Artificial Intelligence

Stanford Institute for Human-Centered Artificial Intelligence contributes to the following Sustainable Development Goals

斯坦福大学人本AI研究致力于
为UN2030可持续发展目标做贡献

https://securesustain.org/abstract/stanford-institute-for-human-centered-artificial-intelligence/

Research Roadmap for European Human-Centered AI



欧洲人本AI研究路线图

https://www.humane-ai.eu/research-roadmap/

What is Human-Centered AI and Its Design Principles?

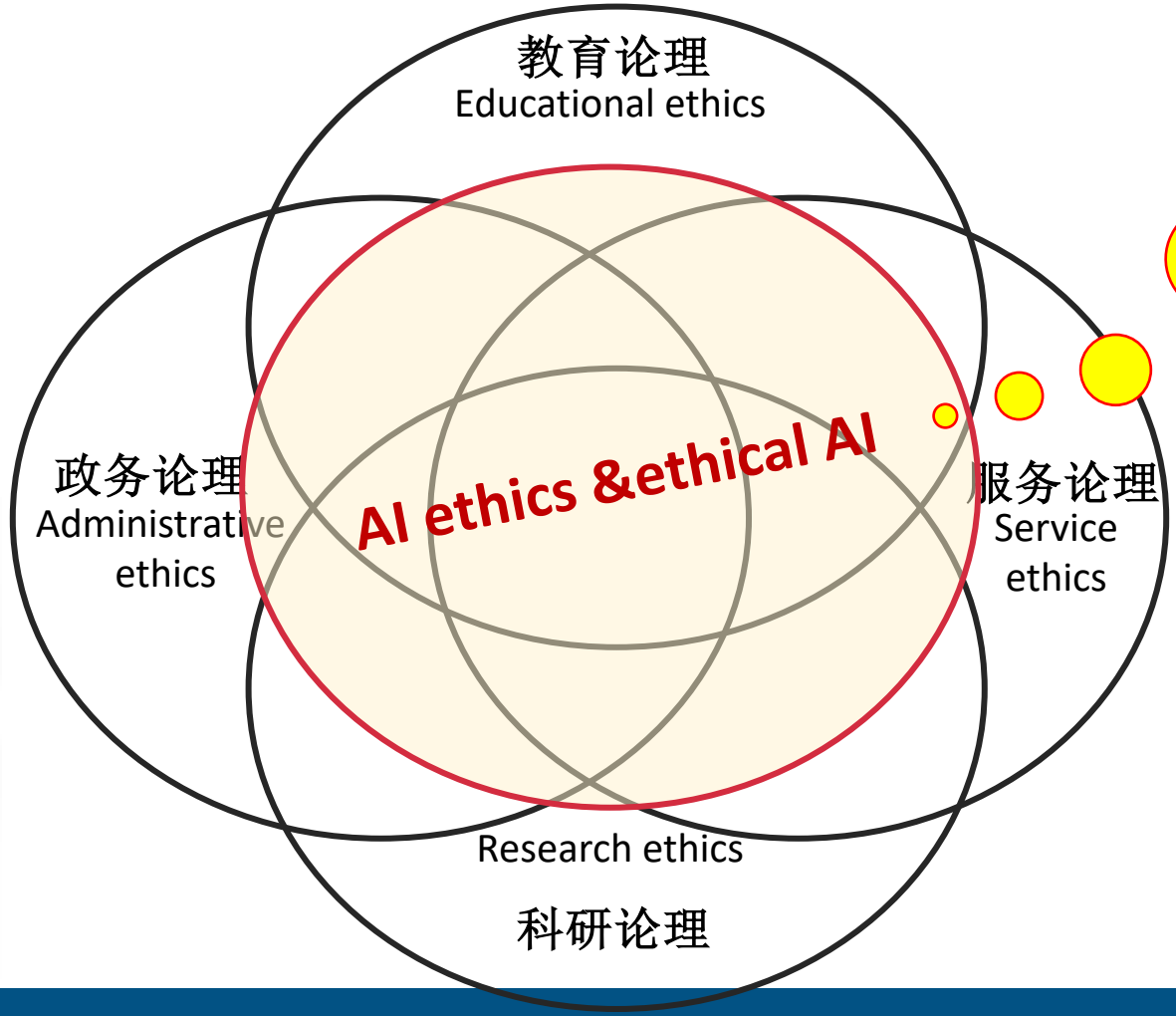
人本人工智能及其设计原理

https://www.xenonstack.com/blog/human-centred-ai

建议2
Recommendation 2

建立AI渗透的高教伦理治理框架

A Comprehensive Governance Framework of AI Ethics Integrated into Higher Education



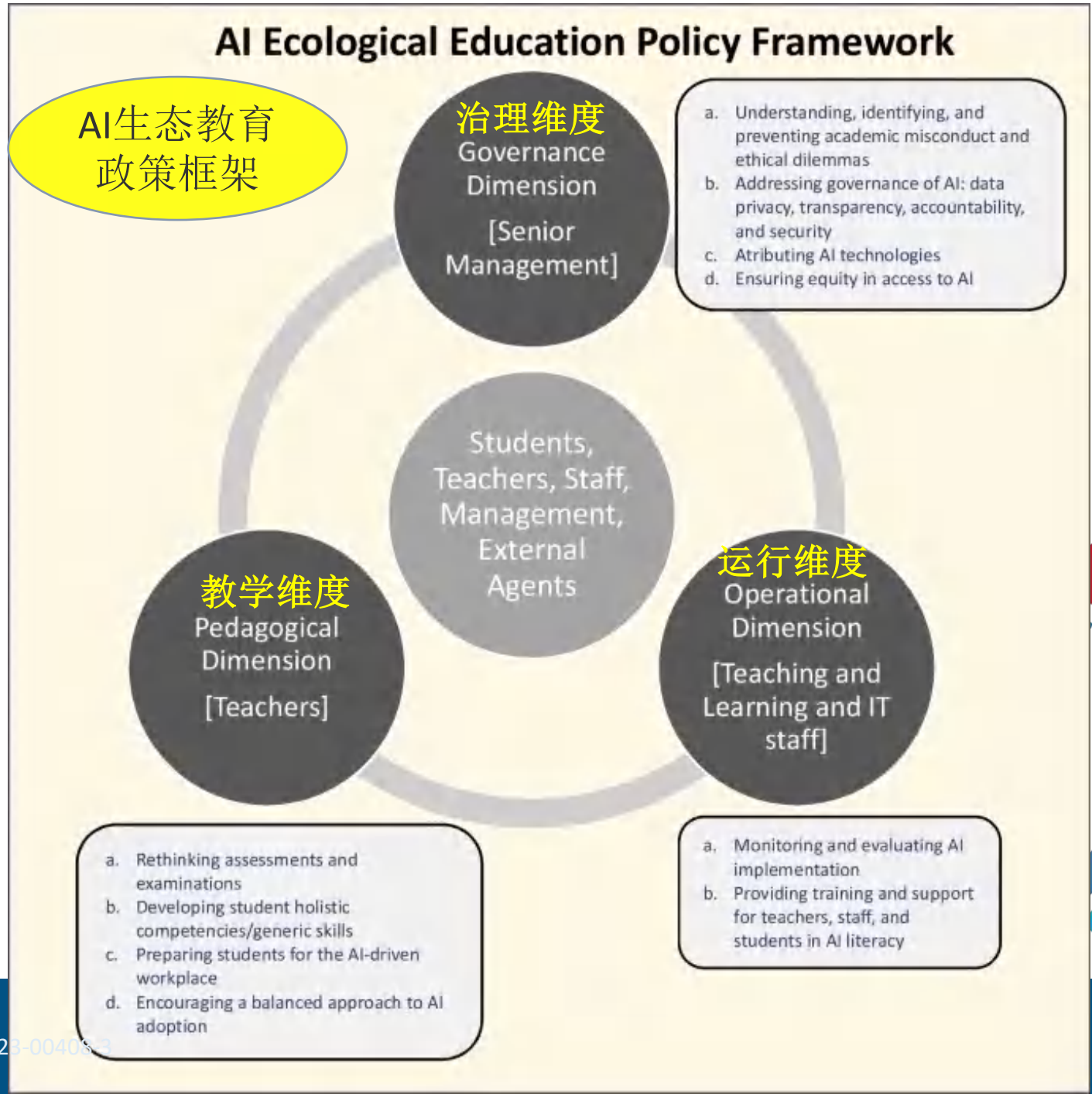
独家见解：因为AI在高教各方面具有渗透性，所以必须将AI伦理与合伦AI融合于高等教育全方位与全过程。

Comprehensive And holistic approach for AI Ethics

建议 3

A comprehensive AI policy education framework for university teaching and learning

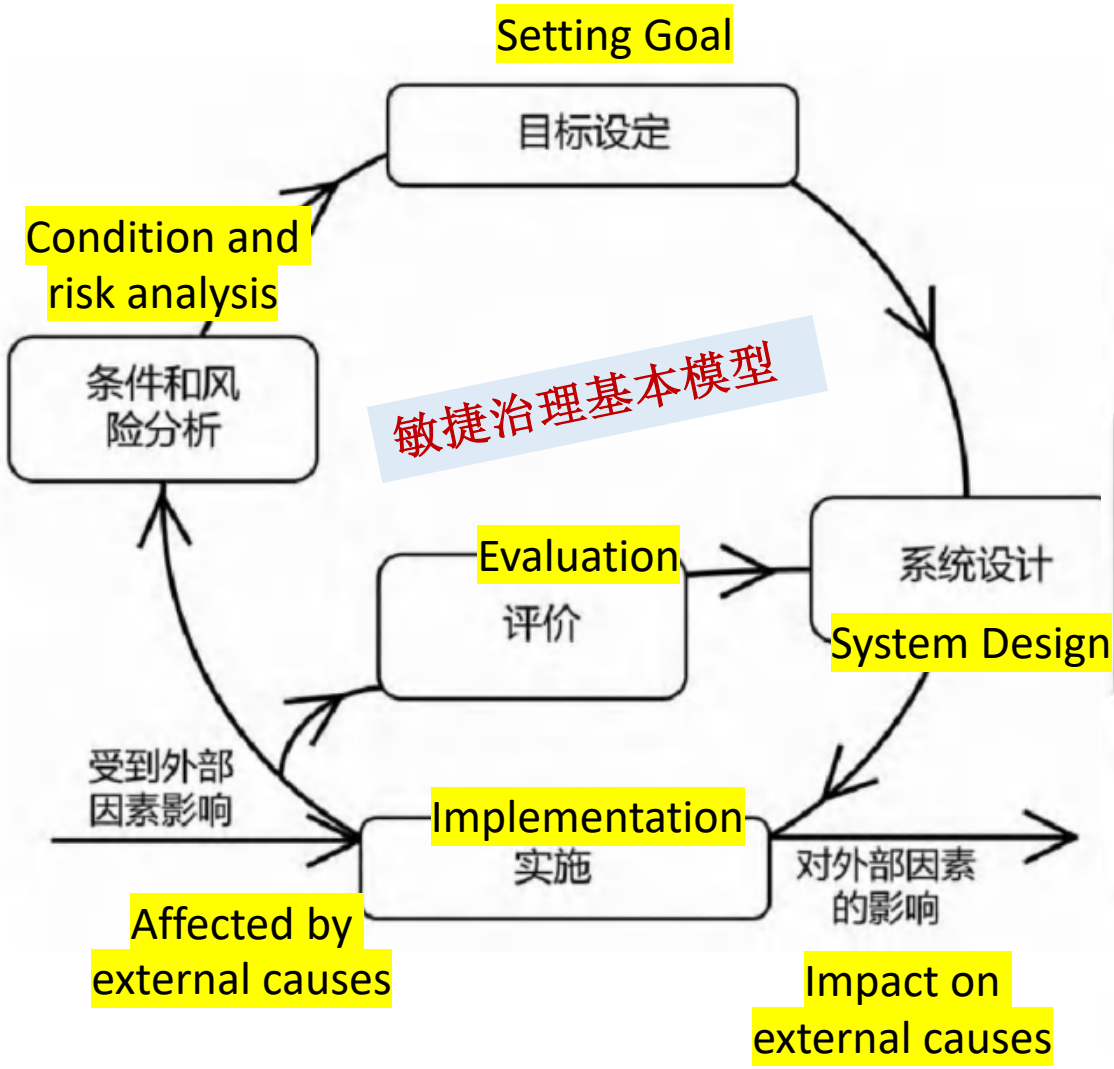
适用于大学教学的全面AI政策教育框架

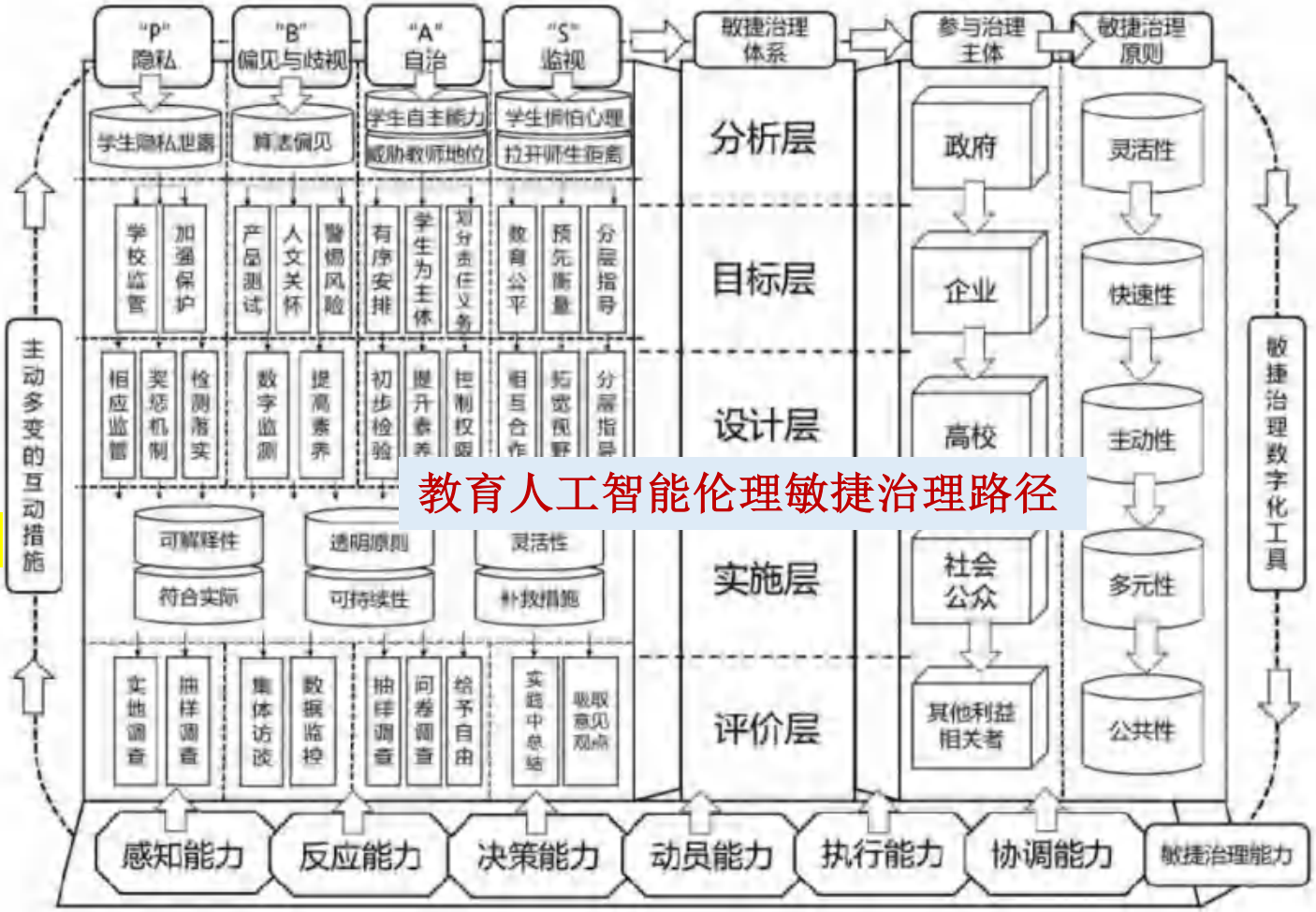


https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-023-00408-3

中国学者：教育人工智能伦理**敏捷治理**

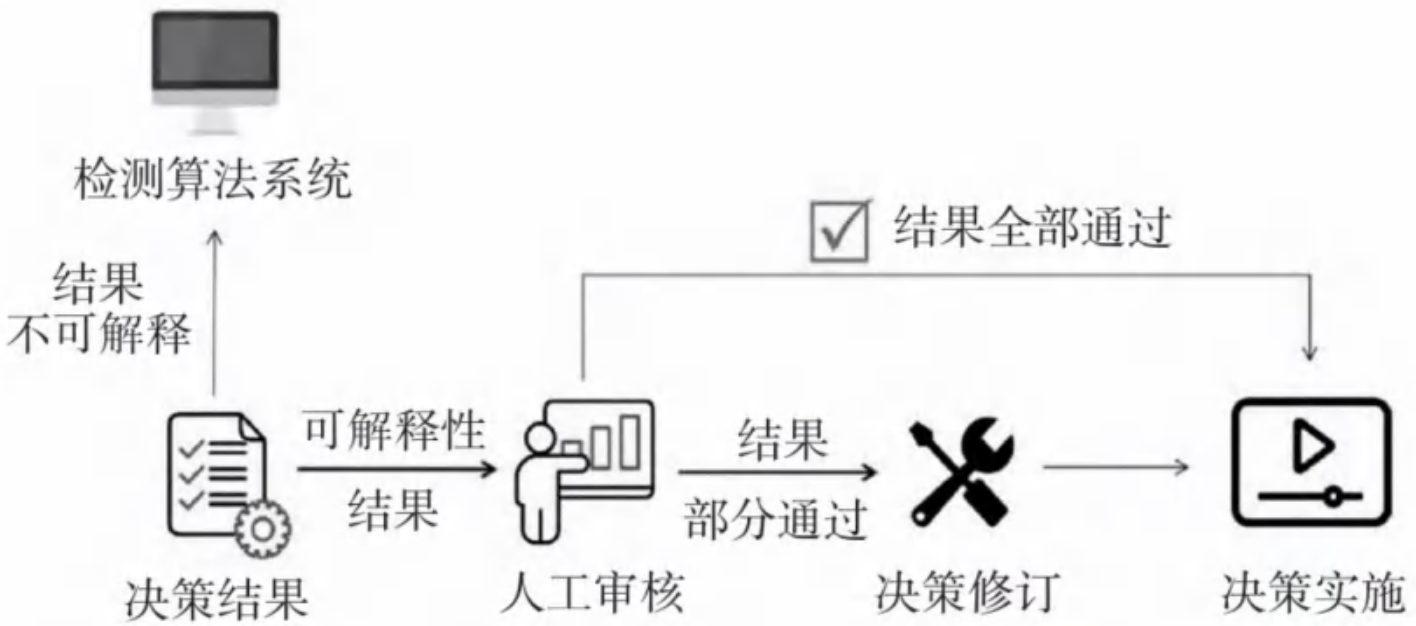
Agile Governance





（王佑镁,等，2023）

中国学者：人机协同教育治理



事前治理、事中治理、事后治理

Before- , In-progress- , and After- governance

推动人工智能技术回归辅助工具的身份，客观公正的看待智能教育系统给出的决策。

教师面对智能教育系统给出的决策结果应当判断其是否具有可解释性：

若结果可解释，则审核决策结果的具体内容，决定是否全盘接受决策结果或经过修订后接收算法结果；

若结果不可解释，教师可要求相关人员检测算法系统。

同时，我们也应当遵循“以人为本”的原则以及人才成长规律，提升学生的创新能力、思维能力以及批判能力；注重发挥教师在教育教学中人格塑造、情感交互等引导性作用，以此来推动学生全面健康的发展。

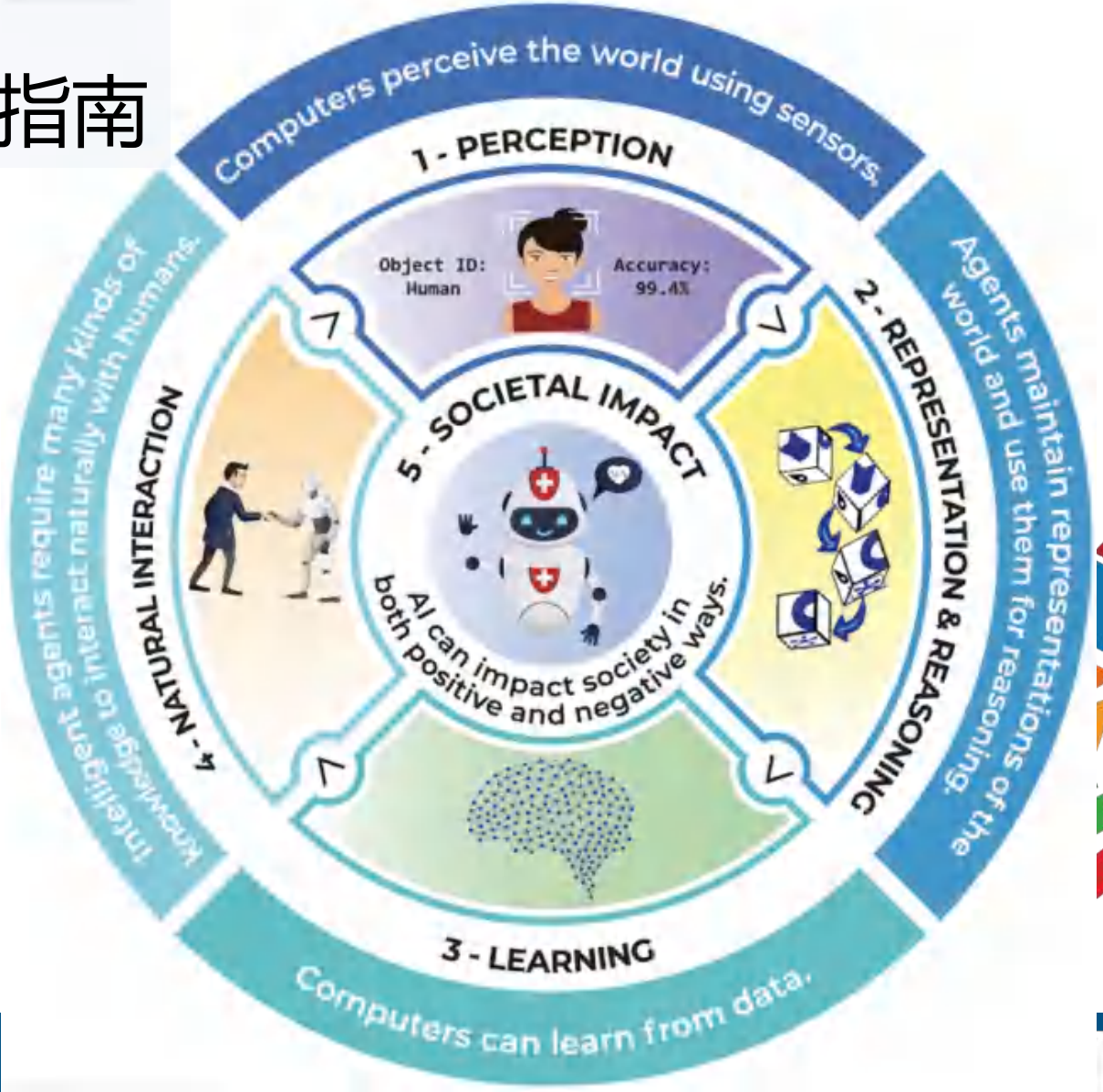
[1]倪琴,刘志,郝煜佳等.智能教育场景下的算法歧视：潜在风险、成因剖析与治理策略[J].中国电化教育,2022(12):93-100.

3 案例

5 Big Ideas in AI—a guideline to support AI education in schools

人工智能的五大理念
——支持学校人工智能教育的指南

1. **感知：**计算机通过传感器感知世界
2. **表示与推理：**智能体保持对世界的表征，并用它们进行推理
3. **学习：**计算机可以从数据中学习
4. .**自然交互：**智能代理需要多种知识才能与人类自然互动
5. **社会影响：**人工智能可以积极和消极的方式影响社会



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8342268/

建议4

Micro-certificate/Credential Course on AI Ethics as recommended

《AI伦理课程》微认证
(建议供参考)

一、课程目标

1.学生掌握人工智能的基本原理和机制以及隐私、安全和透明度的相关问题。

2.学生了解人工智能研发、治理、数据来源等的背后主体关系和人工智能对社会的重要性。

3.学生能认识到人工智能对世界的影响，有哪些挑战、风险和机遇。

4.学生能深入了解人工智能发展的伦理、法律和政策的相关问题。

5.学生熟知使用人工智能的原则和正确方式，并能正确解决一些人工智能伦理的实际问题（以案例研究的形式分析和应用人工智能伦理知识）。

二、课程内容纲要

第一章 人工智能应用概述

第一节 人工智能技术的发展历程

第二节 人工智能关键技术和特点

第三节 人工智能应用现状和未来展望（教育领域）

第二章 人工智能与伦理风险

第一节 人工智能伦理道德内涵与原则

第二节 伦理风险的类型及成因

第三节 人工智能伦理现状及反思

第三章 人工智能伦理与法制

第一节 人工智能伦理政策梳理

第二节 人工智能道德主体与法律责任

第三节 人工智能时代的法制建设

第四章 人工智能伦理与应对策略

第一节 人工智能伦理对传统伦理的新挑战

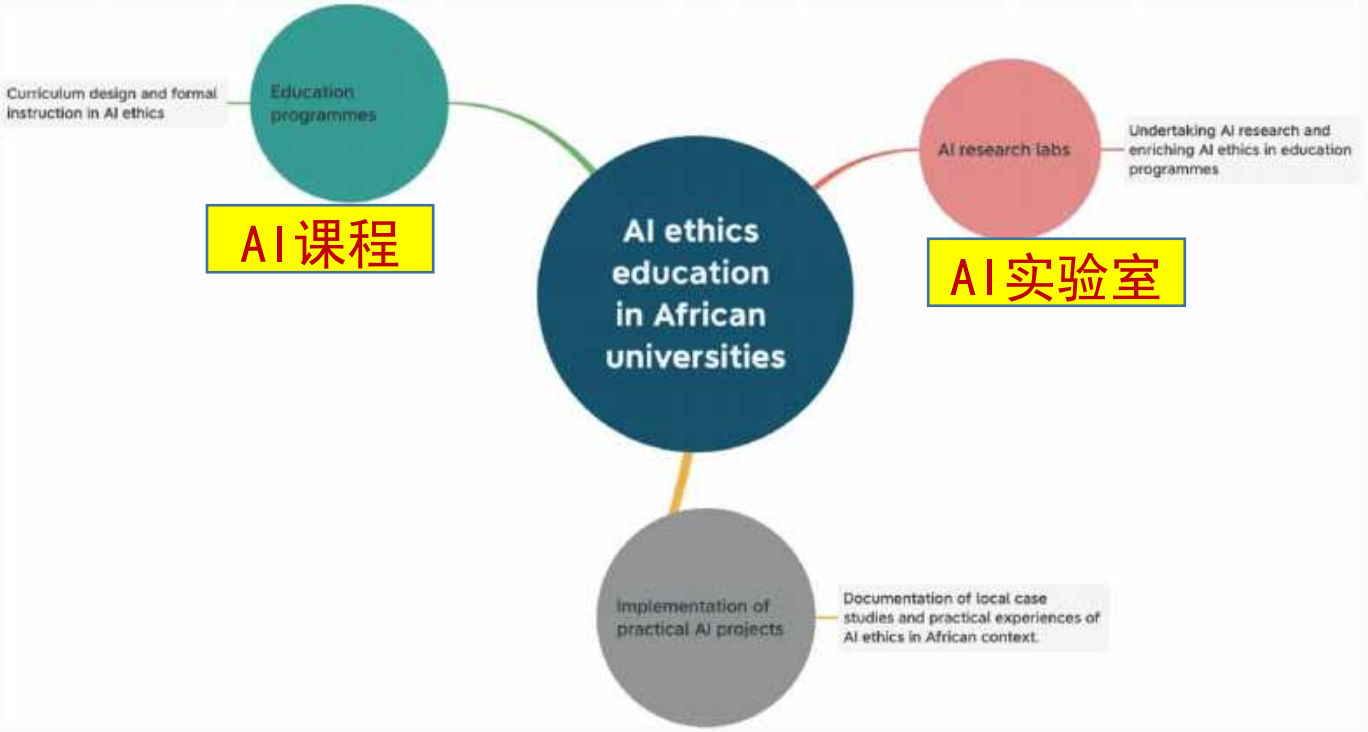
第二节 人工智能伦理问题解决路径

第三节 人工智能伦理体系建构

4 案例

AI Ethics in Higher Education:
Research Experiences from Practical Development and Deployment of AI Systems

非洲高教中人工伦理教育实践的研究经验



非洲大学人工智能伦理教育的关键要素

Theme	Lessons and emerging issues
AI ethics education in African Universities	**L1**: AI ethics is embedded in traditional research methods, although specific courses are emerging
	L2: AI ethics is offered across undergraduate and postgraduate levels at the University
	L3: There is the use of global AI ethics frameworks with some glocalisation
	L4: Institutional AI ethics local capacity is still developing
Role of AI research labs and practical projects in AI ethics education at African Universities	**L5**: African Universities are establishing AI labs
	L6: Minimal AI ethics-specific research themes
	L7: AI labs are providing relevant content for curricula and serve as a vehicle for experiential learning for AI ethics
	L8: AI labs are playing a critical role in promoting AI ethics research and training

非洲高等教育中人工智能伦理教学的经验

https://link.springer.com/chapter/10.1007/978-3-031-23035-6_4

4 案例

美国佛罗里达大学文理学院设《AI伦理》微认证课程

https://reg.pwd.aa.ufl.edu/search/publicCourseSearchDetails.do?method=load&courseId=1015790&selectedProgramAreaId=1015758&selectedProgramStreamId=1016506

课程描述

人工智能1小时课程简介与目标：此课是异步的，这意味着你可以在一年中的任何时候报名。人工智能伦理简介为识别和处理人工智能应用中出现的一些伦理问题提供了一个框架。课程结束时，您将对人工智能应用程序的道德后果有一个基本的了解。本课程结束时，学生将能够：确定伦理的性质和一些核心伦理概念。简要讨论这些概念如何应用于各个领域的人工智能应用，以及这些系统引起的一些伦理问题。

人工智能伦理导论4小时课程描述与目标：此课程是异步的，这意味着你可以在一年中的任何时候报名。完成后，授予0.4个CEU。人工智能伦理导论为识别和解决人工智能应用中出现的一些伦理问题提供了一个框架。课程结束时，您将对相关的伦理概念有更深入的理解，能够更好地评估当前和未来人工智能应用的伦理影响。本课程结束时，学生将能够：解释一些核心伦理概念认识到人工智能应用在各个领域引起的一些普遍道德问题。从道德角度对人工智能应用进行批判性评估。

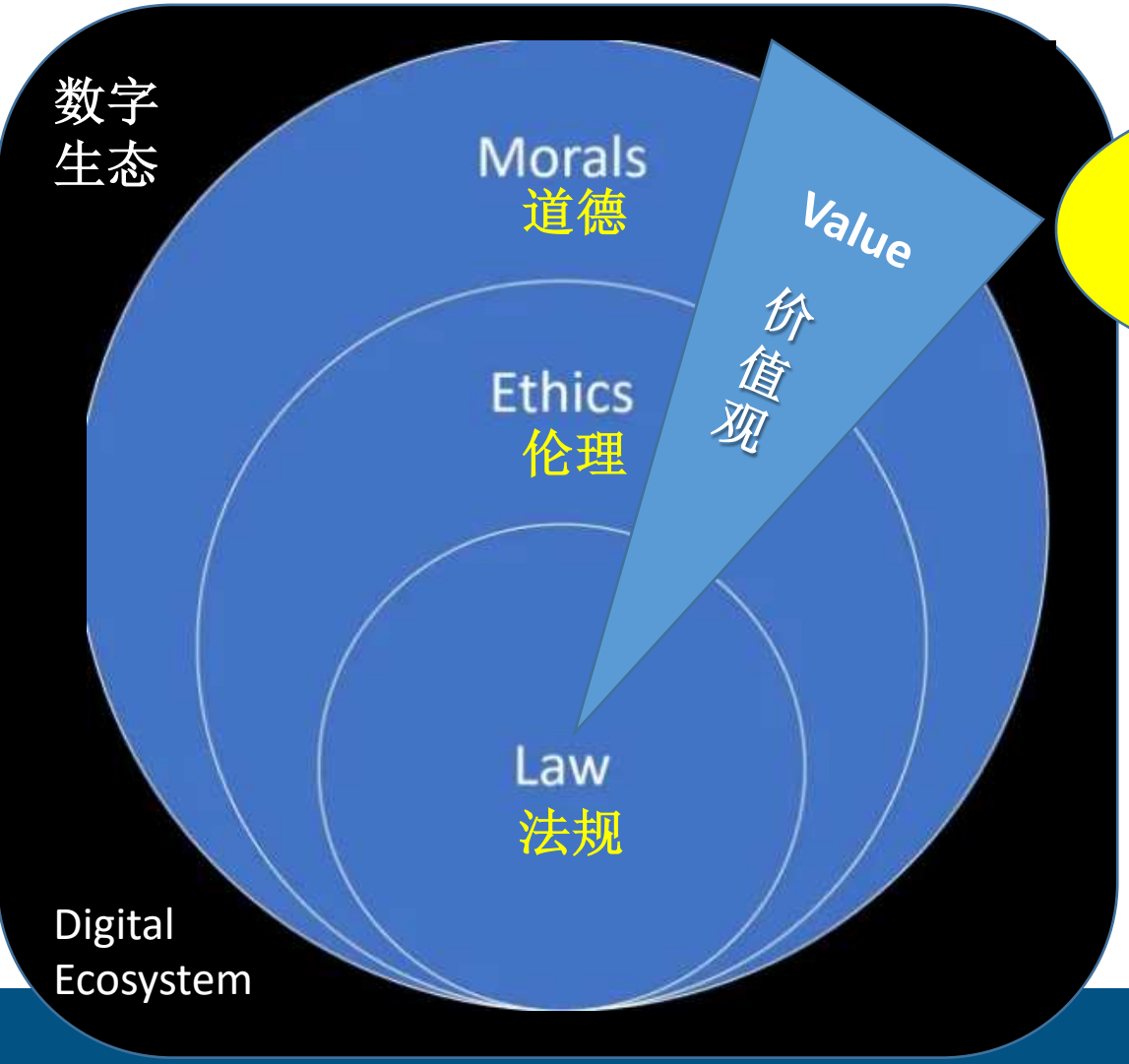
建议5

高教应为建设人技共善的AI伦理体系做贡献

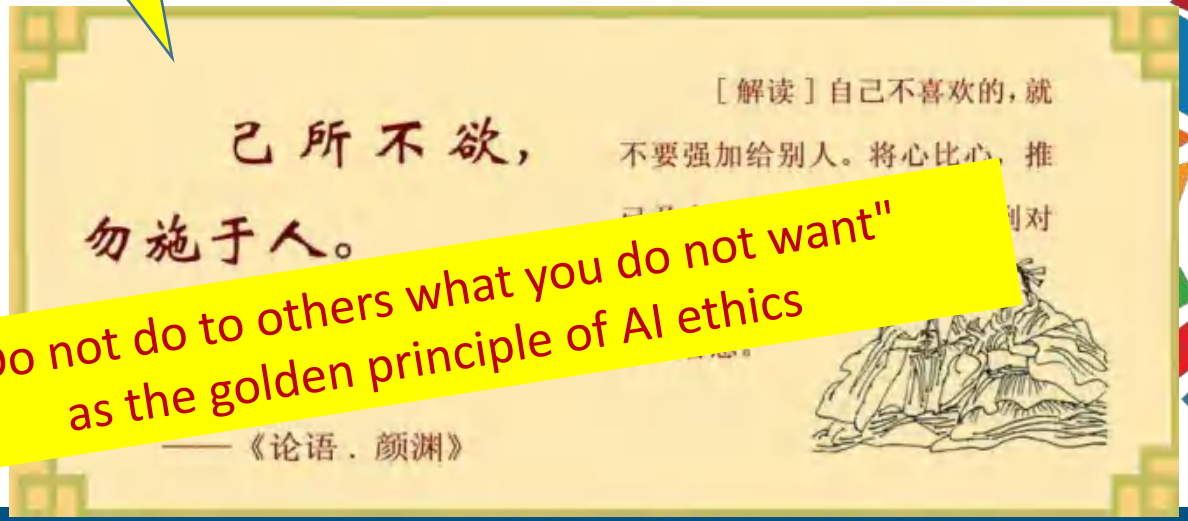
Higher education should contribute to the building of ethics system for human-technology co-goodness

数字德育引导人性向善与技术向善

Digital morality education promoting both human goodness and technology goodness



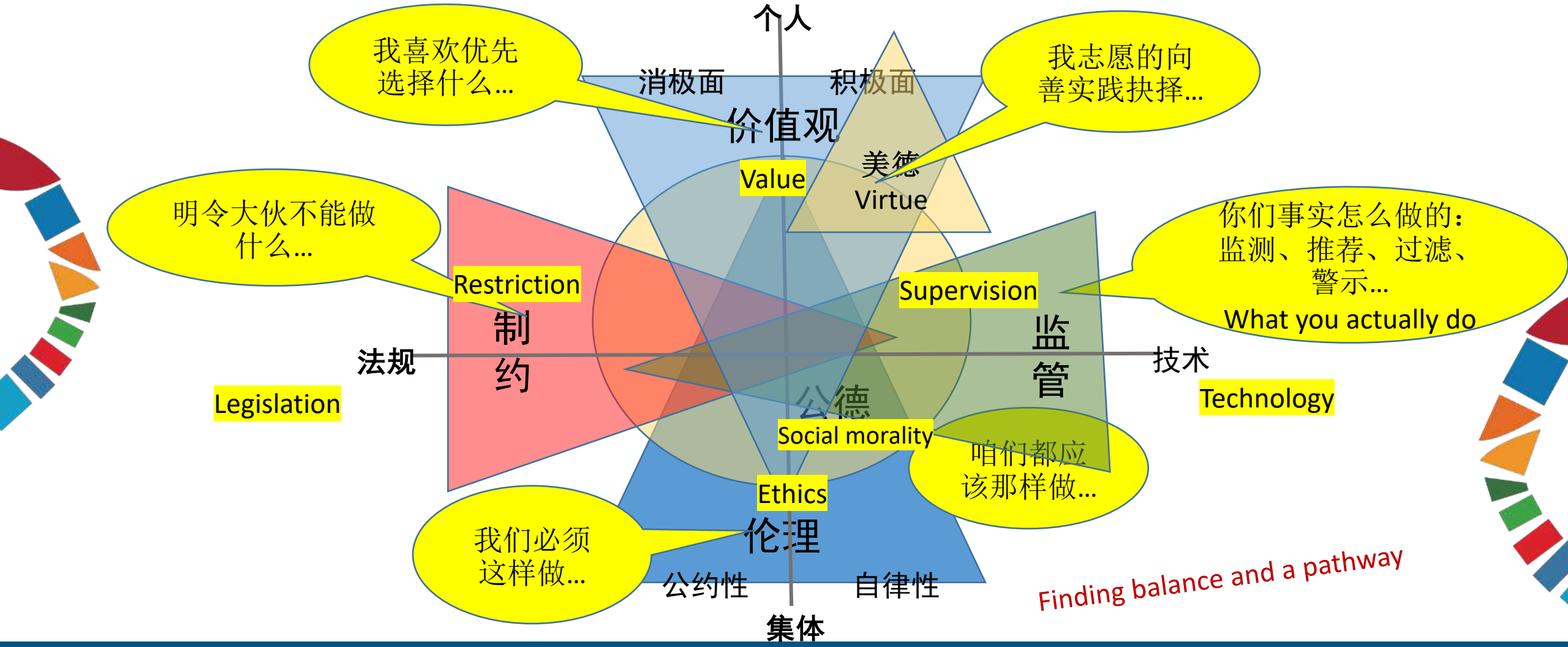
价值观教育的
“黄金法则”



"Do not do to others what you do not want"
as the golden principle of AI ethics

数智社会AI伦理综合治理框架

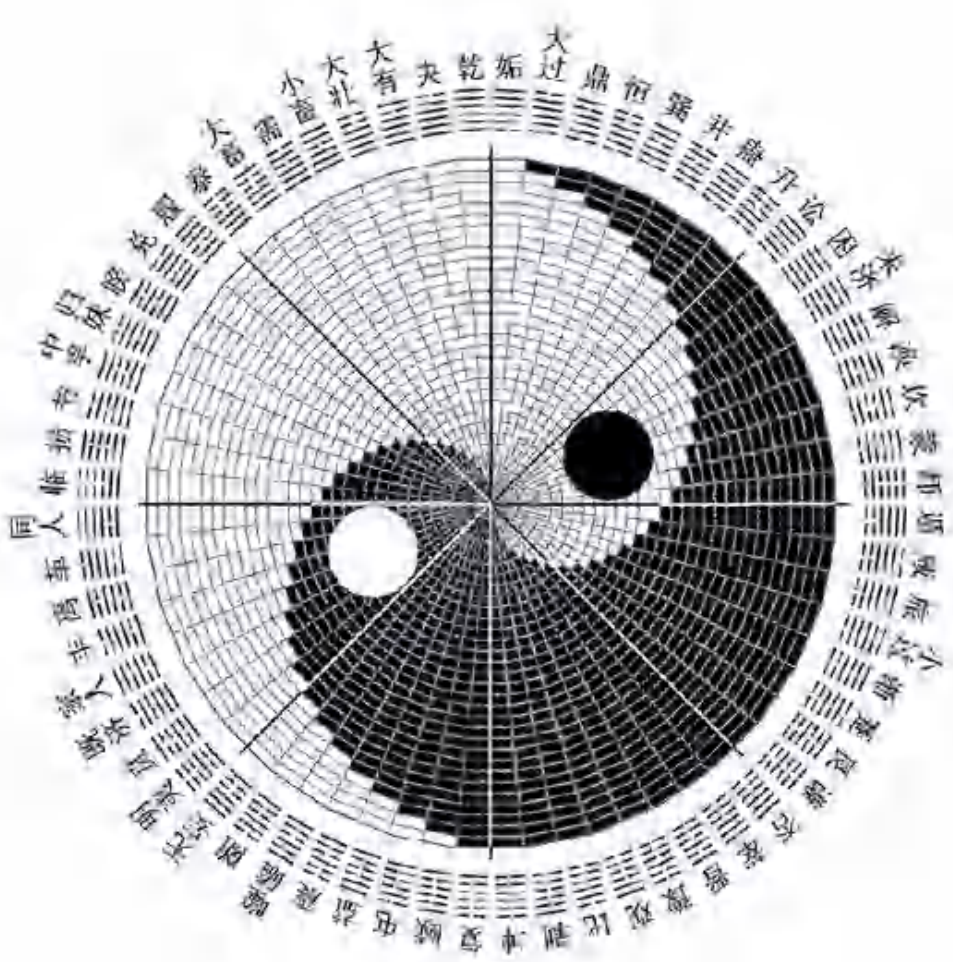
A comprehensive governance framework for AI ethics in the age of digital intelligence



Finding balance and a pathway

（来源：祝智庭，2022-10）

“世上唯一不变的是变化”



“The only thing that is constant is change.” – Heraclitus（赫拉克利特，古希腊哲学家）

“The measure of intelligence is the ability to change.” – Albert Einstein（爱因斯坦）

祝智庭

华东师大终身教授，教育技术学博导
享受国务院特殊津贴专家
教育部教育信息化技术标准委员会首席顾问
教育部人工智能助推教师队伍建设行动试点工作指导专家组副组长
全国信息技术标准化委员会专家委员
国家级教师培训管理者发展中心学术委员会委员
联合国教科文组织高等教育创新研究中心特聘顾问

上海市中山北路3663号华东师大开放教育学院 ztzhu1949@163.com; 15821155868

شكر 谢谢 Thanks Merci Спасибо Gracias

感谢华东师大教育学部戴岭对本报告所做的贡献!
Special Thanks to DAI Ling, Faculty of Education, ECNU, for his contribution to this presentation