人工智能教育伦理：参考框架

（2026年版）

Ethics of Artificial Intelligence in Education: A Reference Framework

（2026）

主体归人 · 协同共生 · 适境致善 · 分类施治

Human Agency · Synergistic Symbiosis · Context-Appropriate Beneficence · Differentiated Governance & Application

世界数字教育联盟专家咨询委员会

Expert Advisory Committee of World Digital Education Alliance

人工智能开放联盟伦理与治理专业委员会

Professional Committee on Ethics and Governance of Artificial Intelligence Open Alliance

中国 · 杭州

Hangzhou, China

2026

人工智能教育伦理：参考框架

（2026 年）

Ethics of Artificial Intelligence in Education: A Reference Framework

主体归人 · 协同共生 · 适境致善 · 分类施治

Human Agency · Synergistic Symbiosis · Context-Appropriate Beneficence · Differentiated Governance & Application

世界数字教育联盟专家咨询委员会

人工智能开放联盟伦理与治理专业委员会

中国 · 杭州

2026

目录

一、 前言 3

（一） 制定背景 3

（二） 定义与内涵 4

（三） 目的与愿景 5

（四） 核心理念 6

（五） 适用范围 7

二、 基本行为导向 9

（一） 强化人机协作边界，凸显育人主体价值 10

（二） 精准适配教育场景，完善分类治理机制 11

（三） 筑牢数据安全防线，严守隐私保护底线 12

（四） 推动算法公平透明，健全责任追溯机制 13

三、 人工智能教育应用的风险类型与责任认定 15

（一） 滥用风险与主观过错归责 15

（二） 故障风险与客观缺陷归责 16

（三） 系统性风险与生态治理归责 17

四、 教育者伦理行为规范 19

（一） 基础教育阶段 19

（二） 高等教育阶段 21

（三） 职业教育阶段 22

五、 学习者伦理行为规范 24

（一） 基础教育阶段 24

（二） 高等教育阶段 26

（三） 职业教育阶段 27

六、 教育机构伦理行为规范 29
（一） 基础教育阶段 29
（二） 高等教育阶段 31
（三） 职业教育阶段 33
七、 分类应用指引与动态治理 35
（一） 守护教育底线的禁止准入 35
（二） 引导人机协作的有限使用 36
（三） 激发创新潜能的鼓励使用 37
（四） 分类动态调整机制 38
八、 结语 40

一、前言

（一）制定背景

人工智能技术正深刻重塑全球教育生态。其应用范围涵盖基础教育、高等教育和职业教育，并贯穿教育教学与科学研究。以生成式人工智能为核心的教学智能体、自适应学习平台、人工智能辅助评价系统等正被逐步采纳并持续迭代。作为与蒸汽机、电力、计算机、半导体和互联网同具变革性的通用目的技术，人工智能不仅增强了人类改造世界的物理能力，更拓展了人类认知世界的智力边界，为教育带来了前所未有的发展机遇和变革挑战：赋能学习者个性成长、助推教育者专业发展、提升教育治理效能、催生科研范式创新、守护教育公平，实现教育的优质均衡和全纳包容。

然而，人工智能向教育领域的深度融入，不仅是单纯的技术应用，更是牵引着教育教学、科研范式及治理模式的系统性变革。在此过程中，也凸显出一系列严峻的伦理风险。1）认知主体退化风险。过度依赖算法易引发思维钝化，导致学习者在人机协同中面临主体性消解的风险，削弱自主创新能力。2）教育公平冲击风险。人工智能技术赋能效果的分化，易诱发强者愈强、弱者愈弱的马太效应。若无普惠性干预，将进一步加剧数字与认知鸿沟。3）数据滥用与隐私泄露风险。智能教育平台对教学全过程数据的采集与挖掘，易导致数据被越权访问或过度收集，进而侵犯师生个人隐私，挑战数据治理与隐私保护底线。

若对上述风险不加以防范，将可能严重背离“促进人的全面发展与社会全面进步”的教育根本宗旨。基于此背景，本参考框架旨在为基础教育、高等教育、职业教育等多元场景中的人工智能应用，提供系统性的伦理框架与行为导引。

本参考框架立足教育的内在规律与育人使命，确立**“主体归人、协同共生、适境致善、分类施治”**的核心理念：以人的全面发展为根本目的，以“师—生—机”良性互动为核心机制，以教育情境的适切性为价值尺度，以多学段多场景的分类应用与治理为实践路径，系统构建人工智能时代的教育伦理体系。

（二）定义与内涵

人工智能作为新一轮科技革命与产业变革的核心驱动力量，其影响已深度融入经济发展、社会进步、教育教学等领域，彰显出赋能社会的技术属性。人工智能正重塑社会形态——催生出人与机器之间人机共融的新型关系。在人工智能时代，伦理的探讨不再局限于人与人、人与自然既定事实之间的关系，更延伸到人类与自身创造的、具有某种智能属性的技术和发明的“自主或半自主”的产品之间的新型关联。人工智能伦理，是指在对人工智能系统进行设计、开发、部署与治理的全过程中，围绕人与机器、人与系统、人与人、人与社会之间所产生的新型关系，建立的一套价值准则、行为规范与治理体系和机制。

人工智能教育伦理，则是人工智能伦理在教育这一特定场域中的具体展开。它研究当人工智能系统进入教育过程，成为教学、学习、科研、评价、管理的重要参与者时，教育活动中原有的“师与生”“教与学”“社会教育目标与学习者自我成长”等核心关系发生了何种变化，以及应当依据怎样的准则来重构这些关系，以确保教育始终服务于人的全面发展。

（三）目的与愿景

本参考框架旨在为基础教育、高等教育和职业教育中教育者、学习者和教育机构建立伦理认知框架，引导各方在人工智能技术的选择和使用过程中坚守伦理底线，推动技术与教育教学的正向融合与共生发展。具体而言，本参考框架致力于：

- 对齐价值导向:明确人工智能技术在教育场景中应用的价值标准与伦理边界，确保技术服务于人的全面发展；
- 培养伦理素养：引导师生建立科学的人工智能认知，提升数字时代的伦理敏感度与人机协同素养,规范日常使用行为；
- 供给制度基准:为教育机构制定符合自身学情与教情的人工智能伦理规范提供实施参照框架；

- 创新治理模式：探索构建“人工智能+教育”的分类分级治理体系，将人工智能在教育场景中的应用划分为“禁止准入”“有限使用”与“鼓励使用”三类，推进精准施治；
- 凝聚全球共识：在世界数字教育的宏阔视野下，促进国际社会在人工智能教育伦理议题上的深度对话、经验共享与协同治理。

（四）核心理念

本参考框架以**“主体归人、协同共生、适境致善、分类施治”**为核心理念。这四项理念既汲取了“以人为本”“止于至善”和“各得其宜”等人类教育智慧，又融合了当代教育哲学与全球人工智能治理的共识性成果，力图为国际社会提供一个开放、包容的对话参考框架。

- **主体归人：人类决策、技术赋能**

教育的本质是人在自主探索中实现认知建构与人格完善。无论技术如何演进，教育的终极目的与主导权始终归于人类。教育者和学习者是教育活动的核心主体，人工智能应致力于激发、而非替代人的思考、判断和创造能力，确保技术始终服务于人的全面与可持续发展。

- **协同共生：师生机协同、智慧共融**

教育不仅是知识传递，更是师生之间、同伴之间意义共建与情感联结的过程。人工智能融入不同教育场景，应在“师

—生—机”之间构建平等、互惠、可持续的三元协同发展生态。这一生态应立足于广阔的社会情境，使技术丰富而非削弱人际联结，实现人类智慧与机器智能的优势互补，最终赋能人与社会的共融发展。

- **适境致善：契合情境，向善而行**

人工智能在教育中的应用，其最终目的是促进人的全面发展。为实现这一目标，技术的使用场景、方式与程度应适配教育规律、符合伦理规范、回应师生关切。同时，尊重各国依据自身国情与教育实际，做出最适切的技术选择，因地制宜地实现科技向善。

- **分类施治：因类制宜，各得其所**

依据教育对象的心智成熟度、育人目标及场景特殊性，将人工智能的教育应用科学划分为“禁止准入”“有限使用”与“鼓励使用”三类。通过分类分级的治理改革，在严守安全与伦理底线的同时，精准释放技术潜能，实现敏捷治理与动态平衡。

（五）适用范围

本参考框架面向全球各级各类教育机构与教育活动，涵盖但不限于基础教育、高等教育和职业教育等教育阶段与类型。适用对象包括教育者、学习者和教育机构等。不同国家

的教育机构可根据实际，依据本参考框架制定具体的实施细则。

二、基本行为导向

人工智能在教育中的应用应始终坚持以教育目的为中心，而非以技术手段为中心。技术应作为实现教育愿景的赋能工具，受控于人并服务于学习者的福祉与教育的公共利益。

为构建包容、公平且以人为本的人机教育生态，基于“主体归人、协同共生、适境致善、分类施治”的核心理念，本参考框架提出以下四项辩证统一的基本行为导向。

- 强化人机协作边界，凸显育人主体价值。明确教育者、学习者在教育活动中的核心主导与主动参与地位，规范人工智能在教学中的应用边界，推动“师—生—机”形成各尽其责、互补增效的协作模式。
- 精准适配教育场景，完善分类治理机制。结合不同学段、不同教学场景的需求，细化人工智能教育伦理规范，建立场景化的治理标准与动态调整机制，确保技术应用贴合教育实际需求。
- 筑牢数据安全防线，严守隐私保护底线。规范教育教学数据的采集、存储与使用权限，完善全生命周期的合规管理体系，防范数据滥用与越权访问，切实保障师生数字权益与信息安全。
- 推动算法公平透明，健全责任追溯机制。防范算法模型中的隐性偏见与歧视，增强智能系统决策的可解释性，明确

技术研发、管理与应用各方的权责边界，确保教育伦理风险可防范、可评估、可问责。

以上四项导向相互支撑、动态平衡，共同构成人工智能教育必须遵循的伦理基准。

（一）强化人机协作边界，凸显育人主体价值

本导向体现主体性与共生性的辩证统一。捍卫“主体性”，要求教育的根本目的与最终决策权必须由人类掌握，严防技术替代核心育人职能，切实凸显人类主体价值。落实“共生性”，则需明确人工智能是现代教育生态的有机组成部分，积极利用技术提升效能。坚持人类主导，在规范协同与动态平衡中实现教育价值最大化。

实施要点：

- **坚持主导中有协作。**核心教育环节的最终决策权归属人类。支持人工智能在探究提问和科研探索等任务中发挥作用，构建“人类发起—智能辅助—人类复核—人类决策”的行动闭环。
- **坚持协作中守边界。**利用人工智能分担计算密集型及重复性任务，保障教育者专注于价值引领与高阶思维培养。设定技术应用阈值，保障师生面对面深度交流的空间，防范技术弱化人际联结。

- **坚持共生中促发展。**推动构建师生、校企及国际社会参与的人工智能教育伦理共同体。在开放协作中完善技术伦理标准，维护教育数据主权与学术自主权，实现高水平安全与高质量发展。

（二）精准适配教育场景，完善分类治理机制

本导向体现价值统一性与实践多样性的辩证统一。坚守“价值统一性”，要求确立科技向善的整体治理准则，确保各类技术应用均不偏离育人初衷。统筹“实践多样性”，则需精准适配教育场景，尊重教育规律、认知阶段及学科属性等客观差异，进而完善分类治理机制。坚持统一准则与因地制宜相结合，确保治理举措贴合教育实际。

实施要点：

- **坚持价值统一性。**建立人工智能教育应用的价值评估制度。严格规范技术落地导向，对于可能诱发认知惰性、加剧教育焦虑或违背教育根本目的的应用模式与服务，应建立严格的准入限制与否决机制。
- **尊重实践多样性。**针对学段差异，基础教育侧重保护基础认知能力，防范生成式人工智能对独立思考的替代。高等教育侧重培养高阶创新能力，支持人工智能在探究式学习和科学研究中广泛应用。职业教育侧重强化技能实训，支持利用人工智能技术构建虚拟仿真等实训环境。针对学科差异，须依据不同学科特点，差异化设计人工智能的介入

深度与方式。针对区域差异，鼓励资源优势地区探索前沿应用，支持资源薄弱地区利用人工智能弥补师资和资源短板，防范数字鸿沟演变为智能鸿沟。

- **实施动态调适评估。**建立伦理影响动态评估机制，根据技术成熟度与教育实效性，定期调整治理标准与应用清单，确保向善原则在具体情境中有效落地。

（三）筑牢数据安全防线，严守隐私保护底线

本导向体现开放共享与安全可控的辩证统一。教育智能化升级依赖数据的合规共享与算法优化，但教育数据会涉及师生个人隐私和科研成果知识产权等，必须在保障数据安全的前提下，推进教育创新的合规开展。

实施要点：

- **实行分类分级管控。**建立教育数据分类分级管理体系。对教育领域核心数据与重要数据实施严格管控，落实本地化部署与数据安全审查。对师生个人信息与隐私数据采取加密传输及脱敏存储等技术防护措施。对公共教育资源等一般数据，鼓励在依法合规的前提下促进数据要素的开放共享。

- **强化数据全生命周期治理。**严格遵循"最小必要"原则开展数据采集，规范数据在存储、处理与应用环节的合规使用，严禁超出原定教育目的的二次开发。针对阶段性教学

行为数据，在达成既定教育目标后应及时予以销毁或匿名化处理，防范数据的过度采集、违规留存与非授权滥用。

- **健全知情同意与权益保障。** 完善数据收集知情同意制度，统筹兼顾学习者数据权益与教育机构的管理需求。未成年人数据采集必须获得监护人授权，切实保障学习者拒绝非必要数据收集的正当权利。

（四）推动算法公平透明，健全责任追溯机制

本导向体现实质正义、程序正义与责任机制的辩证统一。以公平追求教育机会与结果的实质正义，以透明保障决策过程的程序正义，以可问责确保权责一致与救济渠道畅通。通过构建分类分层的问责机制，提升人工智能教育应用的治理效能。

实施要点：

- **落实算法公平。** 定期对涉及招生录取、学业评价、资源配置等场景的人工智能应用系统开展公平性审计。防范并最大限度消除算法模型中的隐性偏见，切实保障各类受教育群体享有平等的教育机会与资源。

- **提升机制透明。** 针对涉及师生重大权益的关键教育决策场景，须明确告知受教育者智能算法的应用范围与基础决策逻辑。通过提供具有可读性的算法可解释性说明，有效规避技术黑箱带来的伦理风险，切实维护师生的知情权。

- **完善问责体系。**明确智能教育生态内各参与主体的责任边界，同步配套风险预警与申诉处置机制，确保相关权益受损时能够获得及时救济。

三、人工智能教育应用的风险类型与责任认定

伴随人工智能与教育的深度融合，其引发的风险日趋复杂且相互交织。在技术提供方、教育者、学习者与教育机构构成的动态复杂网络中，亟需打破责任重叠与模糊的困境。本参考框架确立“权责对等、按因归责”的裁决规则，即风险源于谁、责任归于谁。以下为三类核心风险及其责任界定：

（一）滥用风险与主观过错归责

滥用风险是技术两面性在教育场景下的直接体现，根源于教育参与主体将人工智能用于违背教育初衷的恶意目的。其典型表现为利用智能技术伪造学业成果、操控教育评价、实施过度监控、侵害师生隐私以及散布虚假信息等。此类行为严重异化了技术的赋能属性，挑战了教育评价的真实性与校园生态的道德底线。基于技术向善原则，必须将道德与规范深度融入人工智能算法与应用场景。

【滥用风险责任界定】

滥用风险多源于主观故意，实施滥用行为的个体须承担直接主体责任。针对教育者，若违规滥用技术操纵教育评价或侵犯师生权益，须追究其师德失范与行政纪律责任。针对学习者，若恶意利用人工智能实施代写或造假，须依据学段与情节严重程度承担学业违纪乃至学术不端等责任。针对教育机构，若存在明显的监管盲区与制度缺失，须承担相应的

失察失管与主体管理责任。针对技术提供方，若未尽到安全护栏设计的义务，须承担产品缺陷或连带赔偿责任。

（二）故障风险与客观缺陷归责

故障风险指因人工智能系统自身局限性导致功能失灵，进而对教育教学活动造成的现实误导与阻碍。其典型表现为生成式大模型产生认知幻觉并输出错误事实，以及智能评价与推荐系统因隐性偏见导致学情误判等。该风险根源在于当前人工智能系统高度依赖数据驱动的统计拟合与概率预测，缺乏人类所具备的常识推理与情感洞察能力，导致其难以真正理解教育现象背后的因果逻辑。在缺乏专业介入与人工复核的情况下，系统故障不仅会误导教育者的教学评估与干预决策，更易对学习者的认知建构及各类群体的教育公平造成伤害。

【故障风险责任界定】

故障风险主要源于技术客观局限与系统能力边界，其责任界定须严格遵循按过错程度分配的原则。针对教育者，若在应用过程中放弃必要的人工复核与专业评估而盲目采用智能生成结果，须承担关键教学环节的把关失察责任。针对教育机构，应重点考量其是否履行了合理审慎义务。若教育机构在引入系统时未落实基本的准入评估，或在发现系统输出明显错误及潜在风险后，未能及时采取有效的干预与阻断措施任由损害扩大，须承担相应的管理疏忽与应急处置迟缓

责任。针对技术提供方，须对算法底层缺陷与系统的基础可靠性承担保障与修复补救责任。

（三）系统性风险与生态治理归责

系统性风险指人工智能在教育领域大规模部署后引发的结构性次生危机。该风险并非源于单一系统缺陷，而是深度介入教育生态后引发的宏观异化。其主要表现为智能资源分配不均加剧区域间的智能鸿沟，过度依赖智能化工具导致学习者批判性反思与自主创新能力整体退化，以及海量多模态教学数据汇聚潜藏着隐私泄露风险。应对此类宏观危机应超越单点技术修补，统筹全域力量构建跨界协同的综合治理体系。

【系统性风险责任界定】

系统性风险深刻影响教育宏观生态，其责任界定须遵循“共责但有区别”与“长效预防”的治理原则，核心在于评估各参与主体是否履行了防范生态损害的社会义务。针对教育者，有义务在教学中监测技术对学习者认知发展的长期影响，若发现过度工具依赖却未能适时采取人工干预，须承担教学引导缺失责任。针对学习者，需主动提升数字素养，以抵御系统性认知操纵与信息茧房效应带来的影响。针对教育机构，作为智能教育的组织者需承担核心统筹与干预责任，若未能积极建立普惠接入机制弥合数字鸿沟，须承担宏观治

理缺位责任。针对技术提供方，须承担防范算法偏见规模化蔓延与保障个体隐私安全的底线责任。

四、教育者伦理行为规范

教育者是人工智能进入教育场景的关键责任主体。教育者应明确人工智能应用中的行为边界与责任要求，并坚持“人类决定”的基本立场，即关键教育决策与价值判断必须由人类决定，任何技术系统不得替代教育者的专业判断与责任承担。

无论基础教育、高等教育还是职业教育，三类教育都以人才培养为根本，都承担着促进学习者知识建构、能力发展、价值塑造与责任养成的共同任务。人工智能进入教学设计、课堂教学、实践训练、学习评价与科学研究等环节时，首先必须服务于这一共同育人目标，而不能让智能工具的直接输出替代学习者必要的认知探索与深度思考过程。

同时，不同教育阶段的人工智能伦理重点并不相同：基础教育强调对未成年人主体发展与价值启蒙的保护，高等教育强调学术训练、知识生产与创新活动的规范，职业教育强调课堂学习和实践教学的统一。因此，教育者应在把握共同育人要求的基础上，结合不同阶段特性审慎界定人工智能的功能定位与使用强度。

（一）基础教育阶段

基础教育阶段的重点是保护学习者的主体发展。该阶段学习者的认知能力与价值判断仍在形成，其核心素养的养成高度依赖于持续的亲身参与以及深度的反思内化。教育者需

科学规划人工智能的介入广度与深度，主动为学习者创设充裕的独立思考与自主探究空间，将人工智能定位为启发思维的辅助脚手架，保障学习者在知识建构过程中的核心能动性得以充分激发。

教育者应通过明确规范与持续调控，系统培育学习者面对智能工具时的独立思辨能力。基础教育阶段的人工智能应用，应以促进学习者兴趣、培养学习习惯和夯实基础素养为主要方向。

【行为建议】

课前准备：教育者应主动创设具有挑战性与多维开放性的探究任务，引导学习者将智能工具视为开启思路的辅助工具，使其投入核心问题的建构与逻辑推演过程。

课堂教学：教育者应引导学习者对人工智能生成内容进行比较、解释与再加工，将其作为分析对象而非直接答案使用。同时，指导学习者对相关信息的来源、合理性及潜在偏差进行基本辨识，防止对错误信息与不当价值导向的无意识接受。

成果提交：教育者应要求学习者说明人工智能的使用方式与参与程度，并通过口头表达、过程记录等方式验证其真实理解。

学习评价：教育者应聚焦学习者在学习过程中的投入程度与发展变化，避免以人工智能生成结果替代对其真实能力的判断，确保人工智能应用始终服务于学习者能动性的生成与发展。

（二）高等教育阶段

高等教育阶段的重点是维护知识生产的真实性与创新活动的规范性。教育者应通过学术训练，使学习者形成问题意识、证据意识、伦理判断能力以及研究责任意识。涉及学术观点形成、论证逻辑建构与研究结论生成等关键环节时，应充分发挥人的主体判断作用。

教育者应积极引导学习者辨析自动化内容生成与创造性知识创新的本质区别，强化学术活动中的证据意识与规范意识，确保研究过程与成果具有真实的学术价值与明确的责任归属。高等教育阶段的人工智能应用，应以保障学术诚信与促进高水平创新为核心，推动人机协同服务于知识创造。

【行为建议】

学术选题：教育者应指导学习者基于自身学术积累与社会观察开展选题，鼓励其将智能工具作为拓宽学科视野与激发科研创新的辅助手段，引导研究起点建立在学习者自身的学术理解、问题意识和研究兴趣上。

文献整理：教育者应指导学习者对人工智能提供的信息进行来源核查与真实性验证，避免引用不可靠或不可追溯的内容。

研究推进：教育者应明确要求学习者对核心观点推演、论证逻辑建构及最终成果表述承担独立责任。同时，应规范智能工具在研究推进过程中的使用范围与方法，防止其替代关键分析判断或介入研究结论生成。

成果提交：教育者应要求学习者对人工智能的使用情况进行必要说明，包括其使用环节、方式及对成果的影响。

学术评价：教育者应通过过程性任务、口头答辩等多元证据，鉴别学习者在研究过程中的真实学术贡献。在成果发表时，应遵循学术署名规范，根据期刊等要求如实说明人工智能技术所起的作用，确保研究过程与学术成果的真实性、规范性与责任归属。

（三）职业教育阶段

职业教育阶段的重点是推动专业理论内化与真实工程实践的深度融合，致力于塑造学习者具备高度专业性与社会责任感的综合职业胜任力。现代职业教育的人才培养体系不仅涵盖课堂内的专业知识传授与职业规范建构，更高度依赖于校内模拟实训、真实项目驱动以及企业岗位实战等实践过程。教育者应将人工智能定位为解析复杂生产流程、重构工作协作情境与支持工作决策、流程优化和质量控制的辅助工具，促使学习者在智能技术赋能下实现实践技能与职业操守的提升。

教育者应从职业能力形成的内在逻辑出发，积极构建人机协作完成实际工作任务的教学情境，确保学习者在智能技术辅助下依然能够精准主导核心操作流程并独立承担职业决策责任，推动人工智能应用全方位服务于高素质技能人才的培养。

【行为建议】

基础技能训练：在基础技术训练中教育者应引导学习者将手工实践与自主思考相结合，避免智能工具替代关键操作，保障基础技能与认知经验得到有效积累。

综合实训操作：在复杂工序的标准化实践中教育者应指导学习者合理调用智能工具优化工作流。需重点要求学习者建立实训日志并主动标注智能技术的辅助节点与参与深度，确保技术介入过程具备透明度与可追溯性。

项目实践任务：面对综合性工程项目教育者需重点培养学习者的风险管控与质量保障意识，在涉及生产安全与核心质量的关键节点必须由学习者本人独立完成风险评估与操作决策。

校企实习：教育者需协同企业导师引导学习者严格遵守行业的保密协议与数据合规制度。应指导学习者在企业伦理与规章制度边界内规范应用技术，树立岗位角色认知与明确的责任担当意识。

胜任能力考核：开展职业技能评价时教育者应客观评估学习者的真实专业素养。需通过考核智能辅助方案背后的工程原理与适用条件重点检验学习者对工作产出的专业解释能力与责任承担意识，确保其达到相关岗位的真实胜任要求。

五、学习者伦理行为规范

学习者是人工智能应用的直接使用者，也是自身学习成长与数字素养培育的践行者，需要清晰界定学习者在智能技术应用中的行为边界与责任准则，坚守学习活动“自主成长”的核心立场，即关键学习决策与认知判断必须由学习者自主完成，任何智能技术系统不得替代学习者的独立思考与知识建构。

无论基础教育、高等教育还是职业教育，学习者虽然处在不同发展阶段、承担不同学习任务，但都要在教育过程中实现知识积累、能力提升、品格塑造与责任养成。人工智能的使用因此不应仅以完成任务或提高效率为目标，还应服务于学习者真实理解、主动参与和自我发展。

与此同时，不同教育阶段的学习者在心智成熟度、学习任务与实践要求上存在差异，相应的人工智能伦理要求也有所不同：基础教育更强调在引导和监护中培育自主学习与正确价值判断，高等教育更强调学术诚信、创新意识与批判性思维，职业教育则更强调实践情境中的规范操作、职业诚信与责任意识。因此，学习者应在共性成长目标基础上，结合不同阶段要求审慎把握人工智能的使用边界。

（一）基础教育阶段

基础教育阶段是学习者认知结构建立、思维习惯养成、学习品格塑造的关键奠基期。该阶段学习者的心智发育尚未

完全成熟，需警惕应用人工智能工具时产生依赖性与盲目性。同时该阶段也是提升数字素养与建立社会化规则意识的关键窗口期。

学习者应在教育者引导与家长监护下规范使用人工智能，确保学习过程始终以自主探究与独立思考为核心。必须明确人机协同中的辅助关系，将人工智能定位为拓展知识来源、辅助理解难点与优化认知路径的工具。通过在真实学习场景中的规范应用，引导学习者在解决实际问题的过程中，实现从基础技术操作向数字素养提升的转变。

【行为建议】

课前预习：学习者应优先独立梳理知识框架并标注疑难问题，在教育者或家长指导下选用合规适龄的人工智能工具进行素材拓展。应将智能技术主要用于启发思路、提示疑难和拓展素材，通过自主思考夯实学科基础，避免形成认知依赖。

课堂学习：学习者应坚持先个体思考、再合作交流的学习路径，在深度参与课堂讨论的基础上，适度借助智能工具拆解重难点或梳理逻辑脉络。主动参与课堂互动交流，不断提升语言表达能力与批判性思维水平。

课后实践：学习者应以独立完成作业与动手实操为核心，结合生活实践拓展学习场景，仅将智能工具用于查漏补缺与辅助理解。在完成开放性任务或整理实践素材时，应主动标注智能技术辅助的环节与程度，确保学业成果的真实性。

数据管理：学习者应树立数据安全与个人隐私保护意识，在教育者或家长的指引下规范操作智能学习平台。应通过日常的合规应用，逐步掌握网络风险辨识方法，养成安全理性的数字技术使用习惯。

（二）高等教育阶段

高等教育阶段承担着深化专业理论、开展学术创新、培养高阶思维与拓展社会实践等多重使命。高等教育应围绕课程学习、课题探究与综合实践等多元成长目标，引导学习者成长为具备学术诚信、创新意识和实践能力的高素质人才。

学习者应遵循高等教育的学习规律与学术伦理，科学规范地应用人工智能工具。应在确保自主学习主体地位与原创研究核心价值的基础上，明晰人机协同的功能边界，将其作为提升专业深度、拓宽学术视野与优化知识生产效率的辅助手段。通过在复杂科研情境与社会实践中的理性应用，推动人工智能深度赋能专业成长与综合素养提升。

【行为建议】

专业知识建构：学习者在开展课程预习与专业资料研读时，应优先完成教材的独立阅读与核心逻辑的自主梳理。可利用智能工具辅助查阅前沿文献或构建知识图谱，但须确保对基础理论的深度理解源于个体的钻研，防止自主学习能力被弱化。

课堂互动研讨：学习者应将课堂视为学术对话与思想交流的场所，在专注听讲与积极参与小组讨论的基础上，适度借助智能工具

解析疑难术语或拓展研讨维度。应拒绝机械复制技术生成内容，坚持通过独立表达与观点辨析，将外部信息转化为个体的认知结构。

学术探究创新：学习者在参与课题研究、社会调研或实验开发时，必须确保选题逻辑、论证框架与研究结论均由本人独立建构。智能工具仅可用于辅助性的文献梳理、程序调试或基础数据统计，涉及研究核心价值的判断与创新点提炼须由学习者本人主导。

学业产出表达：学习者在提交课程作业、论文成果或项目报告时，应秉持真实的学术态度并客观展示个体的原创贡献，主动且如实地披露人工智能的使用范围与介入程度，杜绝以技术生成内容替代个体的真实学业付出。

数据管理：学习者在管理科研原始数据、处理课程素材或进行网络交互时，应严守信息安全规定与隐私保护红线。应具备辨识算法偏见与数据误导的敏感性，通过规范的技术操作规避数据滥用风险，确保智能技术的应用符合严谨的科研伦理与社会道德要求。

（三）职业教育阶段

职业教育阶段紧扣产教融合与校企协同的育人目标，致力于岗位核心技能锤炼、实操能力落地与职业规范养成。该阶段以“做中学”为核心特色，通过聚焦实训实操与岗位实景应用，将专业技术素养提升与社会化实践任务深度整合。教育引导应适配技能打磨与职场实践的成长规律，旨在培育符合现代产业需求的综合职业能力。

学习者应立足专业实训与职业发展需求，在岗位技能研习、项目实践及技能考核等核心流程中规范应用人工智能。应明确人工智能在职业场景中的辅助定位，坚持实操优先与自主历练的原则，在确保职业诚信的基础上发挥技术的实践赋能作用。通过在真实生产任务中的理性应用，推动人工智能助力技能进阶与职业素养养成。

【行为建议】

基础技能训练：学习者在进行标准化操作练习时，应确保完整经历核心工艺流程与技术要点。可利用智能工具开展虚拟仿真模拟或原理辅助理解，但必须保证关键动作的练习量与基础认知的自主建构，防范因技术代劳而削弱必要的经验积淀。

综合实训操作：在串接复杂工序与执行系统性任务过程中，学习者应有针对性地使用智能工具优化工艺路径或辅助方案设计。应主动记录技术在各个操作节点的应用详情，建立清晰的实训档案，确保操作过程符合行业规范并具备可追溯性。

项目实践任务：面对具有实际交付要求的项目任务，学习者必须在涉及生产安全、产品质量与技术决策的关键点保持独立判断。应将智能生成建议作为优化思路的参考，确保最终的操作执行与风险把控由本人主导，强化作为职场行动主体的责任意识。

校企实习：学习者在真实或模拟的职场环境中，应严格遵守所在行业的安全保密协议与数据合规准则。在享受技术便利的同时，牢固树立职业诚信并自觉维护公平竞争的行业生态。

六、教育机构伦理行为规范

教育机构作为人工智能进入教育场域的重要组织主体，决定着人工智能在教育活动中的实施方式。教育机构应从制度层面明确人工智能应用的要求。关键教育制度安排应由教育机构依法依规统筹落实。

无论基础教育、高等教育还是职业教育，教育机构都承担着教书育人、制度供给、资源配置、风险防范和环境营造的共同责任。三类教育虽然在培养侧重、组织方式和应用场景上存在差异，但都必须共同聚焦于如何在利用人工智能促进教育质量提升的同时，确保其不削弱育人本质并能够有效维护教育公平。

不同教育阶段的教育机构伦理责任侧重点有所不同。基础教育阶段以保护责任为主，着重保障未成年人权益并引导其价值观形成。高等教育阶段以规范治理责任为主，重点维护学术活动与知识生产的制度边界。职业教育阶段以实践保障责任为主，强调服务学习者的真实工作过程体验与职业能力的形成。

（一）基础教育阶段

基础教育阶段的重点在于保障学习者的主体发展，引导其构建正确的价值观。在基础教育阶段，教育机构所面对的是主体意识、认知能力与价值判断尚处于形成过程中的未成年学习者。人工智能的引入不仅会影响教学方式和学习过程，

还可能对学习者独立思考习惯、价值观建构、学业诚信意识与数字环境中的自我判断能力产生深层影响。

因此，基础教育阶段教育机构的伦理责任不仅在于拓展技术应用，更在于通过审慎规划，确保人工智能的应用不替代学习者核心认知的建构过程。教育机构应致力于维护学习者关键能力发展的自主性，通过界定技术介入的程度与边界，保障学习者养成良好的学习品格并能够独立进行思维判断。

【行为建议】

组织保障机制：教育机构应建立以学习者为本的组织保障机制。学校应构建覆盖校级、部门、年级与班级的人工智能治理架构，明确教学、德育、信息化、学习者管理、数据安全、家校沟通等相关主体的职责边界与协同流程，形成事前有规则、事中有监测、事后可复盘的治理体系。校级层面应制定人工智能应用规范与配套细则，明确不同场景下的允许、限制与禁止边界。对涉及学习者学业评价、纪律处理、资源分配等高利害事项，不得将人工智能输出作为最终依据，须由教育者或相关管理人员进行人工复核并承担最终责任。

沟通协商与人工支持责任：教育机构应承担充分的沟通协商与人工支持责任。对可能影响学习者权益、学习过程或成长环境的人工智能工具，学校应就其使用目的、基本流程、适用范围、潜在风险作出清晰说明，并通过教育者、学习者、家长等多方沟通机制吸纳意见，形成可追溯的决策记录。同时，学校应建立纠错渠道，对

疑似因人工智能导致的误判，及时提供人工复核，保障学习者在发生权益纠纷时能够有实质性的人工介入进行裁定。

隐私保护与公平保障责任：基础教育机构应履行更严格的隐私保护与公平保障责任。学校应建立教育数据分类分级管理制度，明确数据采集的“最小必要”原则、访问权限、保存期限、销毁机制等，并特别规范外部大模型的使用边界，防止师生隐私信息被泄漏。对于低龄学习者的使用场景，学校应实施更高标准的监护人知情同意机制。同时，学校应关注不同学习者的差异，通过个性化支持，防止人工智能应用加剧教育资源分配上可能的不平等。

（二）高等教育阶段

高等教育阶段的重点在于完善规范治理体系，统筹兼顾人才培养的育人使命与知识生产的学术边界。在这一阶段，高等教育机构是学术秩序的维护者与科技创新的引领者，也是高阶思维培养的主阵地。随着人工智能逐渐融入教学与科研各环节，若缺乏明晰的制度规范，智能技术不仅可能模糊研究责任归属、削弱学术训练的真实性并冲击知识生产的专业底线，更可能消解学习者的批判性思维与创新主体性，最终弱化高等教育在人类可持续发展中的作用与意义。

因此，高等教育机构的伦理责任，核心在于将育人作为技术赋能的价值原点。通过制度化治理，机构应在维护学术诚信与保障程序正当的同时，确保人工智能应用始终服从于知识创新的专业逻辑和人的全面发展的育人逻辑。

【行为建议】

组织保障机制：高等教育机构应建立系统化的组织保障机制。学校应设立人工智能教育伦理治理协调机构，统筹教学、科研、管理与服务等全场景应用，明确各相关业务职能部门的职责边界，建立跨部门会商、信息共享与风险处置协同机制。学院及科研机构应落实单位管理责任，结合学科特点细化本单位人工智能工具的适用范围，形成校院联动、权责分明的治理体系。学校应通过制度文件明确不同应用场景中的使用边界，使规范既具有伦理导向，也具有可操作性。

学术诚信与程序正当责任：高等教育机构应突出学术诚信与程序正当责任。学校应围绕生成式人工智能参与课程作业、学位论文、科研写作、代码生成与数据分析等情形，明确允许范围、披露义务、责任边界与署名规范，防止人工智能对学术观点形成、论证逻辑建构与研究结论生成产生不当替代。对涉嫌学术不端的行为认定，不得仅依据人工智能检测结果作出结论，而应坚持证据链审查、人工复核与申辩申诉保障，确保认定过程具备事实基础。同时，学校应加强知识产权、引用规范与成果责任教育，引导师生认识人工智能生成内容在授权、引用、侵权与责任归属方面的伦理风险。

数据治理与高风险事项管控责任：高等教育机构应承担严格的数据治理责任。对于科研数据、未发表成果、评审材料、个人信息等高敏感数据，学校应明确其使用边界，防止研究资料在外部模型平台中泄露。在涉及招生录取、奖助评审、职称评聘、学术评价、处分决定等高利害事项时，学校必须坚持人工复核制度，禁止将人

工智能输出直接作为最终依据。对于引入的第三方人工智能产品，学校应开展必要的合规审查。

（三）职业教育阶段

职业教育阶段的重点在于保障实操主体责任，服务于真实工作过程的完整经历与职业能力的扎实养成。在职业教育阶段，培养目标不仅在于学习者获得知识和技能，更在于其形成能够进入真实工作情境的职业胜任力。人工智能在职业教育中的应用，若缺乏明确边界，容易以“效率提升”之名替代关键操作、压缩实践过程、弱化经验积累，进而影响学习者对完整工作流程的理解以及对职业责任的形成。

因此，职业教育阶段教育机构的伦理责任，核心在于围绕课堂教学与真实工作过程审慎引入技术，使人工智能始终作为辅助工具而非实践主体，服务于职业能力培养而不削弱其实践性与责任性。

【行为建议】

场景化组织保障与制度规范：职业教育机构应围绕实践教学建立场景化的组织保障与制度规范。应明确人工智能工具在课堂教学、实训操作、岗位模拟、技能考核、实习管理与就业服务等场景中的应用边界，形成分类分级、可操作、可追责的制度安排。应根据行业领域特点、岗位风险水平和操作规范要求，细化本单位人工智能工具的可用范围与重点风险提示，防止在高责任、高风险或高度依赖人工经验判断的训练环节中，以技术替代关键操作、过程决

策与经验积累。对发现的违规使用、数据风险、评价争议等问题，应建立问题上报、台账管理、复盘整改的闭环机制，确保实践教学中的风险可识别、问题可追溯、整改可验收。

数据安全保护责任：职业教育机构应高度重视教育教学与生产实践中的数据安全管理。由于职业教育常涉及真实企业案例数据、生产实训记录、职业资格认证信息及实习过程材料，学校应针对校企协同平台建立严格的数据使用规范，通过技术手段与制度约束防止核心数据被不当泄露或扩散。

外部合作治理责任：职业教育机构应加强对第三方技术服务商及智能化设备的准入审查与过程监管。对于引入的各类人工智能实训平台、行业训练系统及智能化实训设备，学校应开展系统的合规性评估，并通过合同明确各方的安全义务、事故响应机制与责任承担要求。应警惕在技能认证、岗位评价与就业推荐等环节形成黑箱式技术依赖，确保算法逻辑的透明性与评价结果的公正性。

七、分类应用指引与动态治理

面向多层次的教育生态，人工智能的融入需确立“分类施治”的核心原则。这一原则包含了“实施使用”与“施加治理”的双重内涵：“治”明确指向严密的治理管控与风险防范；“分类”则充分体现了不同教育场景的学情特征与差异化育人要求。通过张弛有度的分类施治，方能在安全规范的前提下，精准释放人工智能赋能教育的创新价值。

（一）守护教育底线的禁止准入

“禁止准入”类场景主要涉及教育的核心评价机制、价值观建构及安全底线，须实施严格的物理或技术阻断。

在基础教育阶段，核心认知能力的奠基与价值观养成领域须严格限制人工智能的直接介入。在学业测试、期末考试等评价环节，教育机构应建立严密的防作弊与隔离机制。教育者需亲自引导师生间的情感互动与价值建构，避免由算法直接生成价值判断。学习者须独立完成基础认知结构的建构，防范因过早依赖技术导致的思维惰化。

在高等教育阶段，重点维护学术诚信底线与核心专业能力的真实认证。在考试、学位资格审查与专业资格认证等场景中，教育机构需部署有效的技术防范手段，保障学历学位认证的严肃性与公信力。学习者须独立完成学位论文的核心原创章节，以及涉及复杂伦理思辨的专业任务（如医学临床

诊断决策、法学法理辨析等），严禁使用人工智能进行实质性内容的直接生成或思维替代。

在职业教育阶段，严格划定特种行业与高危职业技能鉴定的技术边界。在涉及生命财产安全的职业培训与考核中，必须由具备资质的专业教育者进行核心决策与安全评判。学习者须在真实或等效真实的考核环境中完成职业资格的理论与实操考核，核心操作与关键评判不得由算法替代，以确保行业准入标准的社会公信力不容妥协。

（二）引导人机协作的有限使用

“有限使用”类场景强调在使用人工智能时必须明确主体责任、限定使用深度，并强化过程监督。

在基础教育阶段，突出教育者引导与认知内化。教育机构应建立人工智能辅助教学的规范声明制度。教育者需发挥把关人作用，指导学习者进行信息甄别，并把控表达优化、实验数据核对等辅导环节。学习者在知识拓展或技能训练中使用人工智能获取的信息，必须经过自身的认知加工与逻辑重组，方可转化为最终的学习成果。

在高等教育阶段，应重点强化对学科内涵与能力边界的认知。教育机构应出台具有学科针对性的人工智能使用指南，明晰合理技术辅助与学术不端的界限。教育者需依据学科属性制定差异化规则。学习者可利用智能技术辅助文献梳理或

代码查错，但研究假设的提出、批判性分析与核心算法优化等关键智力劳动须保持自主性。

在职业教育阶段，侧重理实交融与辅助赋能。教育机构应合规配置数字孪生与仿真教学资源，将人工智能作为辅助训练工具。在复杂工作情境模拟中，教育者需结合系统生成的数据报告，对学习者的临场应变与职业素养进行针对性指导。学习者可依托技术开展虚拟工艺流程优化，但最终的精密加工操作与实物验证须由人工独立完成。

（三）激发创新潜能的鼓励使用

“鼓励使用”类场景旨在发挥人工智能的技术杠杆作用，推动教学模式创新与科研范式变革。

在基础教育阶段，聚焦好奇心激发与综合素养培育。教育机构应建立适度的创新容错机制，保护学习者在人机协同探索中的创新热情。鼓励教育者利用人工智能驱动的虚拟仿真与增强现实等技术开展项目式及探究式学习。支持学习者在跨学科项目、编程教育或艺术创作中，将人工智能作为创意催化工具，通过不断试错与迭代生成多元化的解决方案。

在高等教育阶段，聚焦创新人才培养与前沿学术探索。鼓励教育机构提供高性能算力基础设施，支持高风险、探索性的人工智能交叉学科研究。鼓励教育者探索“人机协同科研”新模式，通过算法完成大规模数据清洗与预处理任务，

将人类智力聚焦于理论创新。鼓励学习者利用技术挖掘跨领域的知识图谱关联，在复杂系统设计等领域开展人机共创。

在职业教育阶段，聚焦产教深度融合与产业技术创新。鼓励教育机构联合企业建设“人工智能+”职业场景的智慧实训基地，培养适应智能时代需求的高素质技术技能人才。鼓励教育者对接最新产业动态，带领学习者在高度仿真的生产环境中，利用数字孪生技术解决真实工程难题。支持学习者使用人工智能工具进行产品原型设计、市场分析与商业推演，熟练掌握新型生产与服务模式。

（四）分类动态调整机制

三类应用场景的划分应具备灵活性，需建立常态化的评估与动态调整机制，建议每一到两年开展一次系统性评估。评估过程应由教育主管部门牵头，技术专家、教育学专家、一线师生及公众代表共同参与。评估维度应涵盖技术成熟度（可靠性与安全性）、教育适配度（对育人目标的实际达成效果）及社会接受度等。

场景分类的变更须遵循审慎原则。拟将应用场景由“禁止准入”调整为“有限使用”的场景，须开展严格的区域性或校本化试点，经论证确认安全可控后方可推广。拟将应用场景由“有限使用”放宽至“鼓励使用”的场景，须同步出台配套的素养培育方案。

当出现重大底层技术突破（如通用人工智能 AGI 的迭代）或面临突发性严重伦理风险时，应立即启动应急评估机制，及时校准分类标准与管控措施，让分类施治始终与教育实践同频共振，形成持续进化、自我修正的治理闭环，确保人工智能在教育领域的应用始终沿着“适境致善”“分类施治”的方向健康发展。

八、结语

本参考框架立足人工智能融入教育的新形势，以“主体归人、协同共生、适境致善、分类施治”为核心理念，系统构建了人工智能教育应用的伦理框架。参考框架明确提出“强化人机协作边界，凸显育人主体价值；精准适配教育场景，完善分类治理机制；筑牢数据安全防线，严守隐私保护底线；推动算法公平透明，健全责任追溯机制”等行为导向，并进一步对人工智能教育应用中的风险类型、责任认定和治理边界作出回应。其根本关切，不在于简单回答“人工智能能否进入教育”，而在于进一步明确“人工智能应当以何种方式、在何种边界内、遵循何种价值秩序进入教育”，力求在技术发展与育人使命之间建立清晰的规范基础。

在此基础上，参考框架进一步面向教育实践展开，对教育者、学习者和教育机构在基础教育、高等教育和职业教育等不同场景中的责任与行为边界进行了具体规范，并提出分类施治的应用思路，强调在不同学段、不同任务和不同教育情境中审慎界定人工智能的功能定位与使用强度。由此，参考框架并非单纯鼓励技术进入教育，而是试图通过价值引领、行为规范、风险防范和制度安排的协同推进，确保人工智能始终服务于教育。

值得高度关注的是，由于区域与群体间存在技术应用不平衡的结构性差异，教育公平是伦理原则落实过程中的核心

议题。推进人工智能教育应用须始终把教育公平置于与技术创新同等重要的位置，确保教育智能化转型不应仅仅局限于技术能力的更迭，而应成为扩大优质教育机会、改善学习支持条件与优化教育公共服务的重要契机。人工智能虽具备改善教育资源配置、提升个性化支持水平、增强弱势群体学习可及性和缓解部分结构性短缺问题的潜力，但它同样可能因基础设施差异、数据偏向、平台门槛、使用成本、教育者能力差异及制度支持不均等因素，进一步放大区域、城乡、校际、群体之间的发展落差。也就是说，人工智能是中性的，其并不天然导向公平，它既可能成为弥合差距的工具，也可能演变为制造新鸿沟、加剧原有不平等的负面力量。决定其最终效应的不是技术本身，而是技术被置于何种制度环境之中。

因此，面向未来须推动人工智能发展成果更加公平地惠及全体学习者，尤其应重点关注农村社区、城市低收入群体、残障人士和女性等弱势群体，确保其在数字基础设施、教育资源供给、技术服务支持、教育者专业指导、伦理保护与制度保障等方面获得持续稳定的支持。与此同时，还应把公平要求贯穿于人工智能教育应用的设计、研发、部署、使用、评价与退出全过程，既关注教育机会是否公平可及，也关注教育过程是否公平有效，进一步关注教育结果是否有助于促进人的全面发展和社会整体进步。人工智能进入教育不应仅

作为提升效率的工具选择，更应成为推动教育公正、构建高质量教育体系的建设性力量。

致谢

来自浙江大学、北京师范大学、香港中文大学的学者撰写了《人工智能教育伦理：参考框架》，世界数字教育联盟专家咨询委员会和人工智能开放联盟伦理与治理专业委员会组织国内外专家对《人工智能教育伦理：参考框架》进行了评审，在此一并感谢，名单如下：

浙江大学:吴飞 陈静远 李艳 黄昌勤 孙凌云 陈立萌 黄正行 廖备水 朱强 杨洋 况琨 金方舟

北京师范大学：黄荣怀 李艳燕 杨俊锋 王欢欢 曾海军 王珺怡

香港中文大学：金国庆 蒙美玲

评审专家（按评审时间排序）：杨宗凯 龚克 郑庆华 李永智 薛澜 刘桐汐 唐虔 苗逢春 秦昌威 熊璋 胡祥恩 梁正 王国豫 白惠仁 侯万军 沈明喆 叶民 贡森 陶锋 林建武 秦岩丁 Asha S. Kanwar 陈丽 Mohamed Jemni 展涛 刘德建 刘庆峰 何莲珍 徐雪英 Torunn Gjelsvik Habibah Abdul Rahim Maxim Jean-Louis

Ethics of Artificial Intelligence in Education: A Reference Framework (2026)

Human Agency · Synergistic Symbiosis · Context-Appropriate Beneficence · Differentiated Governance & Application

Expert Advisory Committee of World Digital Education Alliance
Professional Committee on Ethics and Governance of Artificial Intelligence Open Alliance
Hangzhou, China
2026

Table of Contents

I. Introduction 1

(I) Background 1

(II) Definition and Connotation 2

(III) Purpose and Vision 3

(IV) Core Philosophy 4

(V) Scope of Application 6

II. Fundamental Behavioral Guidelines 7

(I) Strengthen the Boundaries of Human-Machine Collaboration and Foreground the Value of Human Agency 8

(II) Precisely Align with Educational Contexts and Perfect Differentiated Governance Mechanisms 9

(III) Fortify Data Security Defenses and Strictly Safeguard Privacy Baselines 10

(IV) Promote Algorithmic Fairness and Transparency, and Establish Robust Accountability Mechanisms 11

III. Risk Types and Responsibility Definition in AI Educational Applications 13

(I) Misuse Risks and Accountability for Subjective Fault 13

(II) Malfunction Risks and Accountability for Objective Defects 14

(III) Systemic Risks and Accountability for Ecological Governance 15

IV. Ethical Behavioral Norms for Educators 17

(I) Basic Education Stage 17

(II) Higher Education Stage 19

(III) Vocational Education Stage 20

V. Ethical Behavioral Norms for Learners 23

(I) Basic Education Stage 23

(II) Higher Education Stage 25

(III) Vocational Education Stage 27

VI. Ethical Behavioral Norms for Educational Institutions 29

(I) Basic Education Stage 29

(II) Higher Education Stage 31

(III) Vocational Education Stage 33

VII. Directives for Differentiated Application and Dynamic Governance 36

(I) Prohibited Access: Safeguarding Educational Baselines 36

(II) Restricted Use: Guiding Human-Machine Collaboration 37

(III) Encouraged Use: Stimulating Innovative Potential 38

(IV) Dynamic Adjustment Mechanism for Categorization 39

VIII. Conclusion 41

I. Introduction

(I) Background

Artificial intelligence (AI) technologies are profoundly reshaping the global educational ecosystem. Their scope of application spans basic education, higher education, and vocational education, permeating both pedagogical practices and scientific research. Pedagogical agents powered by Generative AI (GenAI), adaptive learning platforms, and AI-assisted evaluation systems are being progressively adopted and continuously iterated. As a general-purpose technology on par with the transformative impacts of the steam engine, electricity, computers, semiconductors, and the Internet, AI not only augments humanity's physical capacity to shape the world but also expands our intellectual boundaries in comprehending it. Consequently, AI presents unprecedented developmental opportunities and transformative challenges for education: it harbors the potential to empower learners' personalized growth, facilitate educators' professional development, enhance educational governance efficacy, catalyze innovations in research paradigms, and safeguard educational equity, thereby realizing quality, balanced, and fully inclusive education.

However, the deep integration of AI into the educational sphere goes beyond mere technological application; it drives a systemic transformation in pedagogy, research paradigms, and governance models. During this transition, a series of critical ethical risks have been brought to the fore: 1) Risk of Cognitive Degradation. Over-reliance on algorithms can induce intellectual complacency, exposing learners to the risk of losing their human agency in human-machine collaboration, and thereby undermining their capacity for independent innovation. 2) Threats to Educational Equity. Disparities in the empowering effects of AI technologies may trigger a Matthew effect, where the advantaged benefit disproportionately. Without universally accessible and inclusive interventions, this threatens to further exacerbate digital and cognitive divides. 3) Risks of Data Abuse and Privacy Infringement. The pervasive collection and mining of data throughout the educational process by intelligent platforms can lead to unauthorized access or excessive data harvesting. This infringes upon the personal privacy of educators and learners, posing profound challenges to the foundational baselines of data governance and privacy protection.

Left unmitigated, these risks could cause education to depart from its fundamental purpose: "promoting the holistic development of the individual and the comprehensive progress of society." Against this backdrop, this Reference Framework aims to provide a systematic ethical architecture and behavioral guidelines for the application of AI across diverse scenarios, including basic, higher, and vocational education.

Grounded in the intrinsic laws of education and its mission of nurturing individuals, this Reference Framework establishes the core philosophies of "**Human Agency, Synergistic Symbiosis, Context-Appropriate Beneficence, and Differentiated Governance & Application**." It structurally constructs an educational ethics system for the AI era, positing the holistic development of human beings as the ultimate goal, a constructive "teacher-student-machine" interaction as the core mechanism, contextual appropriateness as the value yardstick, and categorized application and governance across various educational stages and scenarios as the practical pathway.

(II) Definition and Connotation

As a core driving force of the latest scientific and technological revolution and industrial transformation, AI has deeply penetrated economic development, social progress, and educational practices, underscoring its role as a socially empowering technology. AI is reshaping societal structures, giving rise to a novel paradigm of human-machine integration. In the AI era, ethical discourse transcends the traditional boundaries of relationships between humans, and between humanity and nature; it now encompasses the unprecedented dynamics between humanity and its own creations — "autonomous or semi-autonomous" technological artifacts possessing certain intelligent attributes. AI Ethics thus refers to a comprehensive set of value principles, behavioral norms, and governance systems and mechanisms established to navigate the novel relationships between humans and machines, humans and systems, individuals, and society, throughout the entire lifecycle of designing, developing, deploying, and governing AI systems.

AI Ethics in Education (AIED Ethics) represents the specific operationalization of AI ethics within the distinct domain of education. It examines how foundational educational relationships—such as "teacher and student," "teaching and learning," and "societal educational objectives versus learner self-actualization"—are transformed when AI systems enter the educational process to become active participants in teaching, learning, research, assessment, and administration. Furthermore, it delineates the normative principles required to reconstruct these relationships, ensuring that education invariably serves the holistic development of the individual.

(III) Purpose and Vision

This Reference Framework aims to establish an ethical cognitive paradigm for educators, learners, and educational institutions across basic, higher, and vocational education. It guides all stakeholders to uphold ethical baselines during the selection and application of AI technologies, fostering the positive integration and symbiotic evolution of technology and pedagogy. Specifically, the Framework is dedicated to the following imperatives:

Aligning Value Orientations: To articulate clear value standards and ethical boundaries for AI applications in educational settings, guaranteeing that technology serves the holistic development of human beings.

Cultivating Ethical Literacy: To guide educators and learners in forming a scientific understanding of AI, enhancing their ethical sensitivity and human-machine collaborative competencies in the digital age, thereby normalizing responsible daily usage.

Providing Institutional Benchmarks: To offer an implementation reference framework that enables educational institutions to formulate tailored AI ethical guidelines reflecting their specific academic and pedagogical realities.

Innovating Governance Models: To pioneer a differentiated and tiered governance system for "AI + Education." By classifying AI applications in educational settings into three categories—"Prohibited," "Restricted," and "Encouraged"—the Framework advances precise and targeted governance.

Forging Global Consensus: Situated within the broader vision of global digital education, to facilitate profound dialogue, experience sharing, and collaborative governance within the international community regarding the ethics of AI in education.

(IV) Core Philosophy

This Reference Framework is anchored in four core philosophies: **Human Agency, Synergistic Symbiosis, Context-Appropriate Beneficence, and Differentiated Governance & Application**. These principles draw upon enduring human educational wisdom—such as "human-centeredness," "striving for the highest good," and "fittingness for all"—while integrating contemporary educational philosophy with the consensus achievements of global AI governance. They endeavor to provide an open and inclusive reference framework for international dialogue.

Human Agency: Human Decision-Making and Technological Empowerment

The essence of education lies in the cognitive construction and character development achieved through human-driven, autonomous exploration. Irrespective of technological evolution, the ultimate purpose and agency of education must indelibly remain with humanity. Educators and learners are the central subjects of educational activities. AI should be dedicated to stimulating—rather than supplanting—human capacities for critical thought, judgment, and creativity, ensuring that technology invariably serves the holistic and sustainable development of people.

Synergistic Symbiosis: Teacher-Student-Machine Collaboration and Blended Intelligence

Education transcends the mere transmission of knowledge; it is a profound process of shared meaning-making and emotional connection between educators and learners, as well as among peers. The integration of AI across diverse educational settings should forge an egalitarian, reciprocal, and sustainable tripartite ecological symbiosis among "teachers, students, and machines." Rooted in broad social contexts, this ecosystem must utilize technology to enrich, rather than diminish, interpersonal connections, achieving a complementary fusion of human wisdom and machine intelligence to ultimately empower the harmonious development of individuals and society.

Context-Appropriate Beneficence: Contextual Alignment and Technology for Good

The ultimate objective of applying AI in education is to foster the holistic development of human beings. To actualize this, the contexts, methodologies, and extent of technological utilization must be aligned with the intrinsic laws of education, comply with ethical norms, and be responsive to the needs and concerns of educators and learners. Furthermore, it respects the autonomy of diverse nations to make the most appropriate technological choices based on their own socio-cultural realities and educational contexts, ensuring that "technology for good" is realized in a locally adapted manner.

Differentiated Governance & Application: Targeted Measures for Optimal Utility

Based on the cognitive maturity of the learners, pedagogical objectives, and the specificities of the educational context, the educational applications of AI are scientifically categorized into three tiers: "Prohibited," "Restricted," and "Encouraged." Through this differentiated and tiered governance reform, the Framework aims to strictly safeguard safety and ethical bottom lines while precisely unlocking the potential of technology, achieving agile governance and dynamic equilibrium.

(V) Scope of Application

This Reference Framework is directed towards educational institutions and pedagogical activities at all levels and of all types globally, encompassing, but not limited to, basic education, higher education, and vocational education. Its intended audience comprises educators, learners, and educational institutions. Educational authorities and institutions across different nations are encouraged to formulate context-specific implementation guidelines grounded in this Reference Framework, tailored to their distinct operational realities.

II. Fundamental Behavioral Guidelines

The application of AI in education must consistently remain centered on educational purposes rather than technological means. Technology should serve as an enabling tool to realize educational visions, remaining under human control and dedicated to the well-being of learners and the public interest of education.

To construct an inclusive, equitable, and human-centered human-machine educational ecosystem, and grounded in the core philosophies of "Human Agency, Synergistic Symbiosis, Context-Appropriate Beneficence, and Differentiated Governance & Application," this Reference Framework proposes the following four dialectically unified fundamental behavioral guidelines:

Strengthen the boundaries of human-machine collaboration and foreground the value of human agency in education. Clarify the core leading and active participation roles of educators and learners in educational activities, regulate the boundaries of AI application in pedagogical practices, and promote a collaborative paradigm among "teachers, students, and machines" where each fulfills its responsibilities and complements the others to enhance overall efficacy.

Precisely align with educational contexts and perfect differentiated governance mechanisms. Tailored to the distinct needs of various educational stages and pedagogical scenarios, refine the ethical norms for AI in education, and establish contextualized governance standards alongside dynamic adjustment mechanisms to ensure that technological applications closely address actual educational demands.

Fortify data security defenses and strictly safeguard the bottom line of privacy protection. Regulate the permissions for collecting, storing, and utilizing educational and pedagogical data, improve the full-lifecycle compliance management system, prevent data abuse and unauthorized access, and effectively protect the digital rights and information security of educators and learners.

Promote algorithmic fairness and transparency, and establish robust accountability and traceability mechanisms. Prevent implicit bias and discrimination within algorithmic models, enhance the interpretability of intelligent system decisions, clarify the boundaries of rights and responsibilities among technology developers, managers, and users, and ensure that educational ethical risks are preventable, assessable, and accountable.

The above four guidelines are mutually supportive and dynamically balanced, collectively constituting the ethical baselines that must be adhered to in the realm of AI education.

(I) Strengthen the Boundaries of Human-Machine Collaboration and Foreground the Value of Human Agency

This guideline embodies the dialectical unity of human agency and symbiosis. Safeguarding "human agency" dictates that the fundamental purpose and ultimate decision-making authority of education must firmly reside in human hands, strictly preventing technology from supplanting core educational functions, thereby unequivocally highlighting the value of the human subject. Implementing "symbiosis" necessitates recognizing AI as an organic component of the modern educational ecosystem and actively leveraging technology to enhance efficiency. By upholding human leadership, the maximization of educational value is achieved through regulated collaboration and dynamic equilibrium.

Implementation Imperatives:

Maintain human primacy within collaboration: The ultimate decision-making authority in core educational processes belongs to humans. AI is supported in facilitating tasks such as inquiry-based questioning and scientific exploration, constructing a closed-loop action model of "human initiation—intelligent assistance—human review—human decision."

Respect boundaries within collaboration: Utilize AI to offload compute-intensive and repetitive tasks, ensuring educators can concentrate on value guidance and the cultivation of higher-order thinking. Establish thresholds for technology application to safeguard the necessary space for deep, face-to-face interactions between educators and learners, preventing technology from attenuating interpersonal connections.

Promote development through symbiosis: Drive the construction of an AI educational ethics community involving educators, learners, school-enterprise partnerships, and the international community. Perfect technological ethical standards through open collaboration, safeguard educational data sovereignty and academic autonomy, and achieve high-level security alongside high-quality development.

(II) Precisely Align with Educational Contexts and Perfect Differentiated Governance Mechanisms

This guideline embodies the dialectical unity of unified values and diverse practices. Upholding "unified values" requires establishing an overarching governance principle of "technology for good," ensuring all technological applications do not deviate from the original intent of education. Accommodating "diverse practices" necessitates precisely aligning with educational contexts, respecting objective differences such as intrinsic educational laws, cognitive stages, and disciplinary attributes, thereby perfecting the differentiated governance mechanism. Combining unified principles with locally adapted measures ensures governance actions are closely tailored to educational realities.

<u>Implementation Imperatives:</u>

Uphold unified values: Establish a value assessment system for AI applications in education. Strictly regulate the trajectory of technology deployment; stringent access restrictions and veto mechanisms must be established for application models and services that may induce cognitive inertia, exacerbate educational anxiety, or contradict the fundamental purposes of education.

Respect practical diversity: Tailor approaches to varying educational stages. Basic education should prioritize protecting foundational cognitive abilities, preventing Generative AI from supplanting independent thinking. Higher education should focus on cultivating higher-order innovative capacities, supporting the extensive application of AI in inquiry-based learning and scientific research. Vocational education should emphasize strengthening practical skills training, supporting the use of AI to construct virtual simulation training environments. Regarding disciplinary differences, the depth and manner of AI intervention must be differentiated based on specific subject characteristics. Regarding regional disparities, regions with abundant resources are encouraged to explore frontier applications, while under-resourced regions are supported in using AI to bridge gaps in teaching staff and resources, preventing the digital divide from evolving into an intelligence divide.

Implement dynamic adjustment and assessment: Establish a dynamic assessment mechanism for ethical impacts. Regularly adjust governance standards and application inventories based on technological maturity and pedagogical efficacy to ensure the principle of beneficence is effectively actualized in specific contexts.

(III) Fortify Data Security Defenses and Strictly Safeguard Privacy Baselines

This guideline embodies the dialectical unity of open sharing and secure control. The intelligent upgrading of education relies on the compliant sharing of data and the optimization of algorithms; however, educational data inherently involves the personal privacy of educators and learners, as well as the intellectual property of research outcomes. The compliant advancement of educational innovation must therefore be pursued under the strict prerequisite of guaranteeing data security.

Implementation Imperatives:

Implement tiered and classified control: Establish a classified and tiered management system for educational data. Implement strict controls over core and important data within the educational sector, enforcing localized deployment and data security reviews. Adopt technical protective measures, such as encrypted transmission and desensitized storage, for the personal information and privacy data of educators and learners. For general data, such as public educational resources, the open sharing of data elements is encouraged under the premise of strict legal compliance.

Strengthen full-lifecycle data governance: Strictly adhere to the principle of "data minimization" during data collection. Regulate the compliant use of data across storage, processing, and application phases, strictly prohibiting secondary development beyond the original educational purpose. Formative pedagogical behavioral data should be promptly destroyed or anonymized upon achieving the designated educational objectives, preventing excessive data collection, illicit retention, and unauthorized misuse.

Perfect informed consent and rights protection: Improve the informed consent system for data collection, balancing the data rights of learners with the management needs of educational institutions. Data collection involving minors must obtain explicit authorization from guardians, effectively safeguarding learners' legitimate right to refuse non-essential data collection.

(IV) Promote Algorithmic Fairness and Transparency, and Establish Robust Accountability Mechanisms

This guideline embodies the dialectical unity of substantive justice, procedural justice, and accountability mechanisms. It pursues the substantive justice of educational opportunities and outcomes through fairness, guarantees the procedural justice of decision-making processes through transparency, and ensures the alignment of rights and responsibilities and unimpeded remedial channels through accountability. By constructing a differentiated and tiered accountability mechanism, the governance efficacy of AI applications in education is significantly elevated.

Implementation Imperatives:

Ensure algorithmic fairness: Conduct regular fairness audits on AI application systems involved in high-stakes scenarios such as admissions, academic assessment, and resource allocation. Prevent and maximally eliminate implicit biases within algorithmic models, earnestly ensuring that all learner demographics enjoy equal educational opportunities and resources.

Enhance mechanism transparency: In critical educational decision-making scenarios affecting the vital interests of educators and learners, stakeholders must be explicitly informed of the scope of application and the underlying decision-making logic of intelligent algorithms. By providing readable and interpretable explanations of algorithms, the ethical risks associated with technological black boxes can be effectively mitigated, and the right to informed consent of educators and learners safeguarded.

Perfect the accountability system: Clarify the responsibility boundaries of all participating entities within the intelligent education ecosystem. Concurrently, implement risk early-warning and appeal-handling mechanisms to ensure that timely remedies are available when relevant rights are infringed upon.

III. Risk Types and Responsibility Definition in AI Educational Applications

As artificial intelligence becomes deeply integrated with education, the emergent risks grow increasingly complex and inextricably intertwined. Within the dynamic and intricate network comprising technology providers, educators, learners, and educational institutions, it is imperative to resolve the dilemmas of overlapping and ambiguous responsibilities. This Reference Framework establishes the adjudicative principle of "Equivalence of Rights and Responsibilities, and Causality-Based Accountability," asserting that accountability must lie where the risk originates. The following delineates three core categories of risk and their respective attribution of responsibility:

(I) Misuse Risks and Accountability for Subjective Fault

The risk of misuse is a direct manifestation of technology's dual-use nature in educational scenarios, stemming from stakeholders utilizing AI for malicious purposes that contradict the original intent of education. Typical manifestations include leveraging intelligent technologies to fabricate academic outcomes, manipulate educational evaluations, impose excessive surveillance, infringe upon the privacy of educators and learners, and disseminate disinformation. Such behaviors severely alienate the empowering attributes of the technology, challenging the authenticity of educational assessments and the moral baseline of the campus ecosystem. Based on the principle of "technology for good," ethical norms must be deeply embedded into AI algorithms and application scenarios.

[Attribution of Accountability for Misuse Risks]

Risks of misuse predominantly originate from subjective intent; therefore, individuals committing such acts must bear direct primary accountability.

For educators, illicitly abusing technology to manipulate assessments or infringe upon the rights of teachers and students entails accountability for professional misconduct and administrative disciplinary actions.

For learners, maliciously utilizing AI for ghostwriting or fabrication incurs penalties ranging from academic disciplinary action to severe academic misconduct, commensurate with their educational stage and the gravity of the offense.

For educational institutions, obvious regulatory blind spots and systemic deficiencies result in accountability for negligent oversight and institutional mismanagement.

For technology providers, failure to fulfill the obligation of designing adequate safety guardrails necessitates bearing product liability or joint and several compensation responsibilities.

(II) Malfunction Risks and Accountability for Objective Defects

The risk of malfunction refers to functional failures arising from the intrinsic limitations of AI systems, which consequently cause practical misguidance and disruption to pedagogical activities. Typical manifestations include Generative AI models producing cognitive hallucinations and outputting factual errors, as well as intelligent evaluation and recommendation systems misjudging learning profiles due to implicit biases. The root cause of this risk lies in current AI systems' heavy reliance on data-driven statistical fitting and probabilistic prediction, lacking human capacities for commonsense reasoning and emotional insight, rendering them unable to truly comprehend the causal logic behind educational phenomena. Absent professional intervention and human review, system malfunctions not only mislead educators' assessment and intervention decisions but also easily harm learners' cognitive construction and the educational equity of diverse groups.

[Attribution of Accountability for Malfunction Risks]

Risks of malfunction primarily stem from objective technological limitations and system capacity boundaries; thus, responsibility must be strictly apportioned based on the degree of fault.

For educators, blindly adopting intelligently generated results while abandoning necessary human review and professional assessment incurs accountability for gatekeeping failures in critical teaching processes.

For educational institutions, the primary consideration is whether a reasonable duty of care was fulfilled. If institutions fail to conduct fundamental access evaluations upon introducing systems, or fail to promptly implement effective intervention and mitigation measures after discovering obvious system errors and potential risks—thereby allowing harm to escalate—they must bear accountability for management negligence and delayed emergency response.

For technology providers, they must bear the responsibility for guaranteeing the foundational reliability of the system, fixing underlying algorithmic defects, and providing remedial solutions.

(III) Systemic Risks and Accountability for Ecological Governance

Systemic risk refers to the structural, secondary crises triggered by the large-scale deployment of AI in the educational domain. This risk does not originate from a single system defect but rather emerges as a macro-level alienation following deep intervention in the educational ecosystem. Primary manifestations include the uneven distribution of intelligent resources exacerbating the regional intelligence divide; over-reliance on intelligent tools leading to an overall degradation of learners' critical reflection and autonomous innovation capacities; and the massive aggregation of multimodal pedagogical data harboring latent risks of privacy breaches. Addressing such macro-crises requires moving beyond isolated technological fixes to mobilize sector-wide efforts in constructing a cross-boundary, synergized, and comprehensive governance system.

[Attribution of Accountability for Systemic Risks]

Systemic risks profoundly impact the macro-educational ecosystem; thus, accountability attribution must adhere to the governance principles of "common but differentiated responsibilities" and long-term prevention. The core lies in assessing whether participating stakeholders have fulfilled their social obligations to prevent ecological harm.

For educators, there is an obligation to monitor the long-term impact of technology on learners' cognitive development during teaching; failure to implement timely human interventions upon discovering excessive tool reliance incurs accountability for a lack of pedagogical guidance.

For learners, proactive enhancement of digital literacy is required to resist the impacts of systemic cognitive manipulation and filter bubble effects.

For educational institutions, as the organizers of intelligent education, they bear core coordination and intervention responsibilities; failure to actively establish inclusive access mechanisms to bridge the digital divide results in accountability for macro-governance absence.

For technology providers, they must bear the baseline responsibility of preventing the large-scale proliferation of algorithmic biases and safeguarding individual privacy security.

IV. Ethical Behavioral Norms for Educators

Educators are the primary responsible subjects for introducing artificial intelligence into educational scenarios. Educators must clearly delineate the behavioral boundaries and responsibility requirements in AI applications, firmly upholding the fundamental stance of "human decision-making." This dictates that critical educational decisions and value judgments must be made by humans; no technological system shall supplant an educator's professional judgment and accountability.

Whether in basic, higher, or vocational education, the fundamental purpose remains the cultivation of talent. All three tiers share the common mission of facilitating learners' knowledge construction, capacity development, value shaping, and responsibility cultivation. When AI is integrated into instructional design, classroom teaching, practical training, learning assessment, and scientific research, it must foremost serve this unified educational objective. The direct outputs of intelligent tools must never be allowed to short-circuit the essential processes of cognitive exploration and deep reflection required of learners.

Concurrently, the ethical focal points of AI application vary across different educational stages: basic education emphasizes safeguarding the development of minors' learner agency and values formation; higher education prioritizes the normativity of academic training, knowledge production, and innovative endeavors; and vocational education focuses on the seamless integration of classroom learning with practical instruction. Therefore, while adhering to shared pedagogical requirements, educators must prudently define the functional positioning and intensity of AI utilization based on the distinct characteristics of each educational stage.

(I) Basic Education Stage

The primary focus of the basic education stage is to protect the development of learner agency. During this stage, learners' cognitive capacities and value judgments are still in their formative periods, and the cultivation of their core competencies relies heavily on sustained personal engagement and profound reflective internalization. Educators must scientifically plan the breadth and depth of AI intervention, proactively creating ample space for learners to engage in independent thinking and autonomous inquiry. AI should be positioned as an auxiliary cognitive scaffolding designed to inspire thought, ensuring that learners' core agency in the process of knowledge construction is fully galvanized.

Through explicit normative guidance and continuous regulation, educators should systematically cultivate learners' capacities for independent critical thinking when interacting with intelligent tools. In basic education, AI applications should be primarily directed toward stimulating learners' interests, fostering positive learning habits, and consolidating foundational literacies.

[Behavioral Suggestions]

Pre-class Preparation: Educators should proactively design challenging and multi-dimensional inquiry tasks, guiding learners to view intelligent tools as heuristic aids for opening up new lines of thought, thereby engaging them deeply in the construction and logical deduction of core problems.

Classroom Teaching: Educators should instruct learners to compare, interpret, and reprocess AI-generated content, treating it as an object of analysis rather than a definitive answer. Simultaneously, educators must teach learners to fundamentally discern the sources, rationality, and potential biases of relevant information, preventing the unconscious assimilation of erroneous data or inappropriate value orientations.

Assignment Submission: Educators should require learners to explicitly state the manner and extent of their AI usage, verifying their authentic understanding through oral articulation, process documentation, and other formative methods.

Learning Assessment: Educators should focus on the degree of learners' engagement and their developmental progression throughout the learning process. AI-generated outputs must not be used as a substitute for judging a learner's genuine capabilities, ensuring that AI applications consistently serve the generation and development of learner agency.

(II) Higher Education Stage

The primary focus of the higher education stage is to maintain the authenticity of knowledge production and the normativity of innovative activities. Through rigorous academic training, educators should cultivate learners' problem awareness, evidence-based reasoning, ethical judgment, and sense of research responsibility. In critical phases such as the formulation of academic perspectives, the construction of logical arguments, and the generation of research conclusions, the primacy of human subjective judgment must be fully exercised.

Educators should actively guide learners to distinguish the fundamental differences between automated content generation and creative knowledge innovation. They must reinforce evidence awareness and normative compliance in academic activities, ensuring that research processes and outcomes possess genuine academic value and clear attribution of responsibility. In higher education, AI applications should be centered on safeguarding academic integrity and promoting high-level innovation, leveraging human-machine collaboration to serve knowledge creation.

[Behavioral Suggestions]

Academic Topic Selection: Educators should guide learners to select topics based on their own academic accumulation and social observations. AI tools should be encouraged as supplementary means to broaden disciplinary horizons and catalyze research innovation, but the inception of the research must be firmly rooted in the learner's own academic comprehension, problem awareness, and research interests.

Literature Review: Educators should instruct learners to rigorously verify the sources and authenticity of information provided by AI, strictly avoiding the citation of unreliable or untraceable content.

Research Execution: Educators must explicitly require learners to assume independent responsibility for the deduction of core arguments, the construction of logical reasoning, and the articulation of final outcomes. The scope and methodology of utilizing intelligent tools during the research process must be regulated to prevent them from supplanting critical analytical judgments or intervening in the generation of research conclusions.

Outcome Submission: Educators should mandate that learners provide necessary disclosures regarding their use of AI, detailing the specific phases of use, methodologies applied, and the resulting impact on the final work.

Academic Evaluation: Educators should utilize diverse forms of evidence, such as process-oriented tasks and oral defenses, to authenticate the learner's genuine academic contributions during the research process. Upon publication, academic authorship norms must be strictly followed, and the specific role played by AI technologies must be truthfully declared in accordance with journal requirements, ensuring the authenticity, normativity, and accountability of the research process and academic outcomes.

(III) Vocational Education Stage

The primary focus of the vocational education stage is to drive the deep integration of professional theory internalization with authentic engineering practice, striving to mold learners with comprehensive vocational competency characterized by high professionalism and social responsibility. The talent cultivation system in modern vocational education encompasses not only the transmission of professional knowledge and the establishment of occupational norms within the classroom but also relies heavily on practical processes such as on-campus simulated training, real project-driven tasks, and enterprise-based practical operations. Educators should position AI as an auxiliary tool for analyzing complex production workflows, reconstructing collaborative work environments, and supporting decision-making, process optimization, and quality control. This facilitates the enhancement of learners' practical skills and professional ethics empowered by intelligent technologies.

Grounded in the intrinsic logic of how occupational capabilities are formed, educators should actively construct pedagogical scenarios where humans and machines collaborate to complete actual occupational tasks. It is imperative to ensure that learners, while assisted by intelligent technologies, maintain precise leadership over core operational processes and bear independent responsibility for occupational decisions, thereby directing AI applications to comprehensively serve the cultivation of highly skilled professionals.

[Behavioral Suggestions]

Foundational Skills Training: During foundational technical training, educators should guide learners to integrate manual practice with independent cognitive processing, preventing intelligent tools from replacing critical manual operations, thus ensuring the effective accumulation of foundational skills and cognitive experience.

Comprehensive Practical Operations: In the standardized practice of complex procedures, educators should instruct learners on rationally deploying intelligent tools to optimize workflows. Crucially, learners must be required to maintain training logs that proactively annotate the nodes and depth of intelligent technological assistance, ensuring transparency and traceability in the technological intervention process.

Project-Based Practical Tasks: When tackling comprehensive engineering projects, educators must prioritize the cultivation of learners' risk management and quality assurance awareness. At critical nodes involving production safety and core quality control, the learner must independently execute risk assessments and operational decisions.

School-Enterprise Internships: Educators, in collaboration with enterprise mentors, must guide learners to strictly adhere to industry non-disclosure agreements and data compliance regulations. Learners should be instructed to apply technologies normatively within the boundaries of corporate ethics and institutional rules, establishing a clear understanding of their occupational roles and a distinct sense of accountability.

Competency Assessment: When conducting vocational skills evaluations, educators should objectively assess the learner's genuine professional literacy. By examining their grasp of the engineering principles and applicable conditions underlying AI-assisted solutions, the assessment should emphatically test the learner's capacity for professional explication and responsibility assumption regarding their work output, ensuring they truly meet the competency requirements of their respective roles.

V. Ethical Behavioral Norms for Learners

Learners are both the direct users of AI applications and the active agents in their own learning, growth, and digital literacy development. It is imperative to clearly delineate the behavioral boundaries and responsibility guidelines for learners when engaging with intelligent technologies, firmly upholding the core stance of "autonomous growth." This dictates that critical learning decisions and cognitive judgments must be independently executed by the learner; no intelligent technological system shall supplant a learner's independent thinking and knowledge construction.

Irrespective of basic, higher, or vocational education, learners—though at varying developmental stages and undertaking diverse learning tasks—share the common educational objectives of knowledge accumulation, capacity enhancement, character shaping, and responsibility cultivation. Consequently, the utilization of AI should not be driven solely by the goals of task completion or efficiency enhancement; rather, it must serve the learner's authentic understanding, active engagement, and self-actualization.

Concurrently, learners across different educational stages exhibit variations in cognitive maturity, learning tasks, and practical requirements, necessitating tailored AI ethical guidelines. Basic education emphasizes the cultivation of autonomous learning and correct value judgments under guidance and supervision; higher education prioritizes academic integrity, innovative consciousness, and critical thinking; and vocational education focuses on normative operations, professional integrity, and a sense of responsibility within practical contexts. Therefore, while adhering to shared developmental goals, learners must prudently navigate the boundaries of AI usage in accordance with the specific requirements of their educational stage.

(I) Basic Education Stage

The basic education stage is the critical foundational period for constructing cognitive structures, forming thinking habits, and shaping learning character. As learners in this stage have not yet reached full cognitive maturity, there must be vigilance against the dependency and blind usage that can arise when applying AI tools. Simultaneously, this stage represents a pivotal window for enhancing digital literacy and establishing an awareness of societal rules.

Learners should utilize AI normatively under the guidance of educators and the supervision of guardians, ensuring that the learning process remains fundamentally centered on autonomous inquiry and independent reflection. The auxiliary nature of human-machine collaboration must be explicitly defined, positioning AI as a tool for expanding knowledge sources, assisting in the comprehension of difficult concepts, and optimizing cognitive pathways. Through regulated application in authentic learning scenarios, learners should be guided to transition from basic technical operations to the holistic enhancement of digital literacy while solving practical problems.

[Behavioral Suggestions]

Pre-class Preparation: Learners should prioritize independently structuring knowledge frameworks and identifying complex issues. Under the guidance of educators or guardians, they may select age-appropriate and compliant AI tools for material expansion. Intelligent technologies should be used primarily to inspire thought, highlight difficulties, and broaden resources, thereby consolidating foundational subject knowledge through independent reflection and avoiding cognitive dependency.

Classroom Learning: Learners should adhere to the pedagogical pathway of "individual reflection prior to collaborative exchange." Building upon deep participation in classroom discussions, they may judiciously leverage intelligent tools to deconstruct key challenges or map logical trajectories. Learners must actively engage in interactive classroom dialogues to continuously elevate their linguistic articulation and critical thinking capacities.

Post-class Practice: Learners must center their efforts on the independent completion of assignments and hands-on practices, extending learning scenarios through real-life applications. Intelligent tools should only be employed for identifying knowledge gaps and supplementing understanding. When completing open-ended tasks or organizing practical materials, learners should proactively annotate the specific phases and extent of AI assistance to guarantee the authenticity of their academic outcomes.

Data Management: Learners should cultivate an awareness of data security and personal privacy protection, navigating intelligent learning platforms normatively under the instruction of educators or guardians. Through routine compliant usage, they should progressively master methods for identifying cyber risks, developing safe and rational digital technology habits.

(II) Higher Education Stage

The higher education stage bears the multifaceted missions of deepening professional theory, advancing academic innovation, cultivating higher-order thinking, and expanding social practice. Centered around diverse developmental goals—including coursework, research inquiry, and comprehensive practice—higher education must guide learners to evolve into highly qualified professionals characterized by academic integrity, innovative consciousness, and practical competence.

Learners should adhere to the intrinsic laws of higher education and academic ethics, applying AI tools scientifically and normatively. While safeguarding the primacy of autonomous learning and the core value of original research, learners must clearly define the functional boundaries of human-machine collaboration, utilizing AI as an auxiliary means to deepen professional insight, broaden academic horizons, and optimize the efficiency of knowledge production. Through rational application within complex research contexts and social practices, learners should leverage AI to profoundly empower their professional growth and holistic capacity enhancement.

[Behavioral Suggestions]

Professional Knowledge Construction: When undertaking course preparation and the study of professional materials, learners should first independently read textbooks and map out core logic. AI tools may be utilized to assist in reviewing frontier literature or constructing knowledge graphs; however, learners must ensure that their deep comprehension of foundational theories stems from personal intellectual effort, preventing the erosion of their capacity for autonomous learning.

Classroom Interactive Seminars: Learners should treat the classroom as a forum for academic dialogue and the exchange of ideas. Upon a foundation of attentive listening and active participation in group discussions, learners may appropriately use intelligent tools to decipher complex terminology or expand the dimensions of the seminar. They must reject the mechanical replication of technology-generated content, insisting instead on transforming external information into personal cognitive structures through independent expression and the critical analysis of diverse viewpoints.

Academic Inquiry and Innovation: When participating in research projects, social surveys, or experimental development, learners must guarantee that the logic of topic selection, argumentative frameworks, and research conclusions are independently constructed. Intelligent tools may only be utilized for auxiliary tasks such as literature organization, code debugging, or baseline data statistics; judgments concerning the core value of the research and the distillation of innovative points must be spearheaded by the learner.

Academic Output and Expression: Upon submitting coursework, thesis outcomes, or project reports, learners must uphold a genuine academic attitude and objectively present their original contributions. They must proactively and truthfully disclose the scope and degree of AI intervention, strictly eliminating the substitution of authentic academic effort with technology-generated content.

Data Management: When managing raw research data, processing course materials, or engaging in online interactions, learners must strictly observe information security regulations and the red lines of privacy protection. They should possess the sensitivity to identify algorithmic biases and misleading data, circumventing the risks of data abuse through normative technical operations to ensure that AI application aligns with rigorous research ethics and societal moral imperatives.

(III) Vocational Education Stage

The vocational education stage is intricately aligned with the objectives of integrating industry and education, and fostering school-enterprise collaboration. It is dedicated to honing core occupational skills, actualizing hands-on capabilities, and cultivating professional norms. Characterized by the core ethos of "learning by doing," this stage deeply integrates the elevation of professional technical literacy with socialized practical tasks through a focus on real-world training and authentic workplace applications. Educational guidance must adapt to the developmental logic of skill refinement and workplace practice, aiming to nurture comprehensive vocational competencies that meet the demands of modern industry.

Learners should base their actions on the requirements of professional training and career development, normatively applying AI across core processes such as occupational skill study, project execution, and competency assessment. They must recognize the auxiliary positioning of AI in vocational scenarios, adhering to the principles of prioritizing hands-on practice and autonomous experiential learning. By actualizing the practical empowerment of technology while ensuring professional integrity, learners should drive the use of AI to facilitate skill advancement and the cultivation of professional ethics through rational application in authentic production tasks.

[Behavioral Suggestions]

Foundational Skills Training: During standardized operational exercises, learners must ensure they fully experience the core procedural workflows and technical imperatives. Intelligent tools may be utilized for virtual simulations or to assist in grasping underlying principles; however, learners must guarantee a sufficient volume of practice for critical physical actions and the autonomous construction of foundational cognition, guarding against the dilution of necessary experiential accumulation caused by technological substitution.

Comprehensive Practical Operations: When integrating complex procedures and executing systemic tasks, learners should purposefully utilize intelligent tools to optimize process pathways or assist in solution design. They must proactively record the details of technology application at each operational node, establishing transparent training archives to ensure the operational process complies with industry standards and maintains traceability.

Project-Based Practical Tasks: When confronted with project tasks requiring actual deliverables, learners must maintain independent judgment at critical junctures involving production safety, product quality, and technical decision-making. AI-generated recommendations should be treated as references for optimizing approaches, ensuring that final operational execution and risk control are directed by the learner, thereby reinforcing their sense of accountability as the primary actor in the workplace.

School-Enterprise Internships: Within authentic or simulated workplace environments, learners must strictly adhere to the safety, non-disclosure agreements, and data compliance protocols of their respective industries. While benefiting from technological conveniences, learners must firmly establish professional integrity and consciously uphold a fair and competitive industry ecosystem.

VI. Ethical Behavioral Norms for Educational Institutions

As the pivotal organizational entities facilitating the entry of artificial intelligence into the educational domain, educational institutions determine the modalities of AI implementation in pedagogical activities. Institutions must explicitly define the requirements for AI application at the institutional level. Critical educational policy arrangements must be comprehensively implemented by these institutions in strict accordance with relevant laws and regulations.

Regardless of whether they are engaged in basic, higher, or vocational education, educational institutions share common responsibilities: educating and nurturing learners, providing institutional frameworks, allocating resources, mitigating risks, and cultivating the educational environment. Although the three tiers differ in their pedagogical focus, organizational modalities, and application scenarios, they must collectively concentrate on a singular imperative: how to leverage AI to enhance educational quality while ensuring it neither undermines the essence of education nor compromises educational equity.

The focal points of ethical responsibilities for educational institutions vary across different educational stages. The basic education phase prioritizes protective responsibilities, focusing on safeguarding the rights of minors and guiding the formation of their values. The higher education phase emphasizes normative governance responsibilities, focusing on maintaining the institutional boundaries of academic activities and knowledge production. The vocational education phase centers on practical assurance responsibilities, emphasizing service to learners' authentic workflow experiences and the formation of occupational competencies.

(I) Basic Education Stage

The core focus of the basic education stage lies in safeguarding the development of learner agency and guiding the construction of sound values. In this stage, educational institutions cater to minor learners whose sense of agency, cognitive capacities, and value judgments are still nascent. The introduction of AI not only alters pedagogical methods and learning processes but may also exerts profound impacts on learners' habits of independent thinking, value construction, awareness of academic integrity, and self-determination capacities within digital environments.

Consequently, the ethical responsibility of basic educational institutions extends beyond expanding technological applications; it necessitates prudent planning to ensure that AI usage does not short-circuit the learner's core cognitive construction processes. Institutions must commit to preserving the autonomy of learners' critical capability development. By delineating the degree and boundaries of technological intervention, they ensure learners cultivate strong learning character and the capacity for independent critical judgment.

[Behavioral Suggestions]

Establish Learner-Centered Organizational Safeguard Mechanisms: Schools should construct an AI governance architecture spanning the school, department, grade, and class levels. The responsibility boundaries and collaborative workflows of stakeholders involved in teaching, moral education, information technology, learner management, data security, and home-school communication must be explicitly defined. This establishes a governance system characterized by ex-ante regulation, in-progress monitoring, and ex-post review. At the institutional level, normative guidelines and supporting protocols must be formulated to clarify the boundaries of permitted, restricted, and prohibited use across diverse scenarios. For high-stakes matters involving learner academic evaluation, disciplinary action, and resource allocation, AI outputs must never serve as the final determinant; human review by educators or administrators is mandatory, and they must bear ultimate accountability.

Fulfill Responsibilities for Communication, Consultation, and Human Support: Institutions must undertake rigorous communication and human support duties. For AI tools that may impact learners' rights, learning processes, or developmental environments, schools must transparently articulate their purpose, operational flow, scope of application, and potential risks. Opinions should be assimilated through multi-stakeholder communication mechanisms involving educators, learners, and parents, resulting in traceable decision records. Concurrently, schools must establish grievance channels for remediation. In instances of suspected AI-induced misjudgments, timely human review must be provided, guaranteeing substantive human intervention and arbitration when rights disputes arise.

Uphold Strict Privacy Protection and Equity Assurance Responsibilities: Basic education institutions must enforce elevated standards of privacy protection and equity assurance. Schools should implement a tiered and classified data management system, explicitly defining the "data minimization" principle in collection, access permissions, retention periods, and destruction mechanisms. The boundaries for utilizing external Large Language Models (LLMs) must be strictly regulated to prevent the leakage of privacy information of educators and learners. For scenarios involving younger learners, schools must enforce higher standards for guardian informed consent. Simultaneously, schools must remain attentive to learner diversity, preventing AI applications from exacerbating potential inequalities in educational resource allocation through the provision of personalized support.

(II) Higher Education Stage

The higher education stage focuses on perfecting the normative governance system, balancing the pedagogical mission of talent cultivation with the academic boundaries of knowledge production. In this stage, higher education institutions are the custodians of academic order, the vanguard of scientific and technological innovation, and the primary bastions for cultivating higher-order thinking. As AI progressively permeates all facets of teaching and research, the absence of clear institutional norms risks blurring the attribution of research responsibilities, diluting the authenticity of academic training, and challenging the professional baselines of knowledge production. Furthermore, it threatens to erode learners' critical thinking and innovative agency, ultimately weakening the role and significance of higher education in sustainable human development.

Therefore, the core ethical responsibility of higher education institutions lies in anchoring "nurturing individuals" as the foundational value of technological empowerment. Through institutionalized governance, institutions must ensure that AI applications consistently remain subordinate to the professional logic of knowledge innovation and the pedagogical logic of holistic human development, while simultaneously upholding academic integrity and ensuring procedural justice.

[Behavioral Suggestions]

Establish Systematic Organizational Safeguard Mechanisms: Higher education institutions must establish an AI ethics governance coordinating body to oversee applications across all scenarios—teaching, research, administration, and services. The responsibility boundaries of relevant functional departments must be clarified, establishing cross-departmental mechanisms for consultation, information sharing, and collaborative risk mitigation. Colleges and research institutes must fulfill their managerial responsibilities, refining the applicable scope of AI tools within their units based on disciplinary characteristics, thereby forging a governance system with clear institutional-departmental linkage and distinct demarcations of rights and responsibilities. Schools should issue policy documents defining usage boundaries across varying scenarios, ensuring that norms are both ethically oriented and operationally actionable.

Foreground Responsibilities for Academic Integrity and Procedural Justice: Schools must clarify the permitted scope, disclosure obligations, liability boundaries, and authorship norms concerning the involvement of Generative AI in coursework, degree theses, academic writing, code generation, and data analysis. This prevents AI from improperly supplanting the formulation of academic perspectives, logical argumentation, and the generation of research conclusions. Determinations of alleged academic misconduct must not be based solely on AI detection results; institutions must insist on chain of evidence reviews, human verification, and the guarantee of appeal and defense rights, ensuring the investigative process rests on a solid factual basis. Furthermore, schools should intensify education on intellectual property, citation norms, and accountability for outcomes, guiding the academic community to recognize the ethical risks associated with AI-generated content regarding authorization, citation, infringement, and liability attribution.

Execute Responsibilities for Data Governance and High-Stakes Control: Higher education institutions must undertake rigorous data governance. For highly sensitive data—such as research data, unpublished findings, peer-review materials, and personal information—schools must strictly demarcate usage boundaries to prevent the leakage of research materials into external model platforms. In high-stakes matters involving admissions, scholarship reviews, professional title promotions, academic evaluations, and disciplinary decisions, schools must mandate human-in-the-loop review systems, explicitly prohibiting AI outputs from serving directly as the final basis for decisions. Necessary compliance reviews must be conducted for all onboarded third-party AI products.

(III) Vocational Education Stage

The vocational education stage prioritizes safeguarding the agency of learners in practical operations, serving the complete experience of authentic workflows and the solid formation of occupational competencies. In vocational education, the developmental objective is not merely the acquisition of knowledge and skills, but the cultivation of comprehensive vocational competency enabling learners to enter authentic workplace scenarios. If AI applications in this sector lack clear boundaries, they can easily usurp critical operations under the guise of "efficiency enhancement," compressing practical processes and diluting experiential accumulation. This subsequently impairs learners' comprehension of complete workflows and the formation of their occupational accountability.

Consequently, the core ethical responsibility of vocational education institutions lies in prudently introducing technology around classroom instruction and authentic workflows. AI must remain an auxiliary tool rather than the subject of practice, serving the cultivation of occupational competencies without undermining their practical and accountable nature.

[Behavioral Suggestions]

Implement Scenario-Based Organizational Safeguards and Institutional Norms: Vocational institutions must establish scenario-based safeguards centered around practical teaching. The boundaries of AI tool application must be explicitly defined in scenarios encompassing classroom teaching, practical training operations, job simulations, skills assessments, internship management, and employment services, forming tiered, actionable, and accountable institutional arrangements. Based on industry characteristics, job risk levels, and operational norms, institutions should refine the permissible scope of AI tools and highlight key risk advisories. This prevents technology from replacing critical operations, process decision-making, and experiential accumulation in training phases characterized by high responsibility, high risk, or a heavy reliance on human empirical judgment. A closed-loop mechanism—comprising incident reporting, ledger management, and post-event rectification—must be established for identified non-compliant usage, data risks, and evaluation disputes, ensuring that risks in practical teaching are identifiable, problems traceable, and rectifications verifiable.

Fulfill Responsibilities for Data Security Protection: Vocational education institutions must prioritize data security management in educational practices and production activities. As vocational education frequently involves authentic enterprise case data, production training logs, occupational qualification certification information, and internship process materials, schools must establish stringent data usage protocols tailored to school-enterprise collaborative platforms. Core data must be safeguarded against improper leakage or diffusion through both technological safeguards and institutional constraints.

Strengthen Responsibilities for External Collaboration Governance: Institutions must enhance the access review and process supervision of third-party technology providers and intelligent equipment. For various onboarded AI practical training platforms, industry training systems, and intelligent equipment, schools must conduct systematic compliance assessments. Contracts must explicitly stipulate the safety obligations, incident response mechanisms, and liability requirements of all parties. Institutions must remain highly vigilant against the formation of black-box technological dependencies in processes such as skills certification, job evaluation, and employment recommendations, guaranteeing the transparency of algorithmic logic and the fairness of evaluation outcomes.

VII. Directives for Differentiated Application and Dynamic Governance

Facing the multi-layered educational ecosystem, the integration of artificial intelligence must establish the core principle of "Differentiated Governance & Application." This principle encompasses the dual connotations of "application execution" and "governance administration": "Governance" explicitly targets rigorous control and risk prevention; "Differentiated" fully reflects the diverse learner profiles and varied pedagogical requirements across distinct educational contexts. Only through a well-calibrated approach of differentiated governance can the innovative value of AI-empowered education be precisely unlocked under the premise of safety and normativity.

(I) Prohibited Access: Safeguarding Educational Baselines

"Prohibited access" scenarios primarily involve core educational evaluation mechanisms, value construction, and safety baselines, necessitating strict physical or technological blockades.

In Basic Education: The foundational domains of core cognitive capabilities and value formation must strictly limit direct AI intervention. In evaluation phases such as academic tests and final examinations, educational institutions must establish rigorous anti-cheating and isolation mechanisms. Educators must personally guide emotional interactions and value construction among learners, preventing algorithms from directly generating value judgments. Learners must independently complete the construction of foundational cognitive structures to prevent cognitive inertia caused by premature technological dependency.

In Higher Education: The focus is on maintaining the baselines of academic integrity and the authentic certification of core professional competencies. In scenarios such as examinations, degree qualification reviews, and professional certifications, educational institutions must deploy effective technological countermeasures to safeguard the solemnity and credibility of academic credentialing. Learners must independently complete the core original chapters of their degree theses, as well as professional tasks involving complex ethical reasoning (e.g., clinical diagnostic decisions in medicine, jurisprudential analysis in law). The use of AI for the direct generation of substantive content or as a substitute for human thought in these areas is strictly prohibited.

In Vocational Education: Strict technological boundaries must be drawn for specialized industries and high-risk occupational skills assessments. In vocational training and evaluations involving the safety of life and property, core decision-making and safety judgments must be conducted by qualified professional educators. Learners must complete theoretical and practical assessments for occupational qualifications in authentic or equivalently authentic testing environments. Core operations and critical evaluations must not be supplanted by algorithms, ensuring the societal credibility of industry entry standards remains uncompromised.

(II) Restricted Use: Guiding Human-Machine Collaboration

"Restricted use" scenarios emphasize that when utilizing AI, primary accountability must be clearly defined, the depth of use must be delineated, and process supervision must be reinforced.

In Basic Education: Emphasizes educator guidance and cognitive internalization. Educational institutions should establish a normative declaration system for AI-assisted teaching. Educators must act as gatekeepers, guiding learners in information verification and overseeing auxiliary processes such as expression optimization and experimental data checking. Information acquired by learners via AI during knowledge expansion or skills training must undergo their own cognitive processing and logical restructuring before being translated into final learning outcomes.

In Higher Education: The focus is on reinforcing the understanding of disciplinary connotations and capability boundaries. Educational institutions should issue discipline-specific AI usage guidelines, clarifying the demarcation between legitimate technological assistance and academic misconduct. Educators must formulate differentiated rules based on disciplinary attributes. Learners may utilize intelligent technologies to assist with literature reviews or code debugging; however, critical intellectual labor — such as proposing research hypotheses, conducting critical analyses, and optimizing core algorithms—must maintain strict human autonomy.

In Vocational Education: Focuses on the integration of theory and practice alongside auxiliary empowerment. Educational institutions should compliantly deploy digital twins and simulated teaching resources, utilizing AI as an auxiliary training tool. In simulations of complex work scenarios, educators must provide targeted guidance on learners' on-the-spot adaptability and professional ethos based on system-generated data reports. Learners may rely on technology to optimize virtual process workflows, but the ultimate precision machining operations and physical verifications must be completed independently by human hands.

(III) Encouraged Use: Stimulating Innovative Potential

"Encouraged use" scenarios aim to leverage the technological capacities of AI to drive innovations in pedagogical models and transformations in research paradigms.

In Basic Education: Focuses on stimulating curiosity and cultivating comprehensive literacies. Educational institutions should establish appropriate safe-to-fail mechanisms for innovation, protecting learners' enthusiasm for exploration during human-machine collaboration. Educators are encouraged to utilize AI-driven virtual simulations and augmented reality technologies to conduct project-based and inquiry-based learning. Learners are supported in using AI as a catalyst for creativity in interdisciplinary projects, programming education, or artistic creation, generating diverse solutions through continuous trial, error, and iteration.

In Higher Education: Focuses on cultivating innovative talent and frontier academic exploration. Higher education institutions are encouraged to provide high-performance computing infrastructure to support high-risk, exploratory, and interdisciplinary AI research. Educators are encouraged to explore new paradigms of "human-machine collaborative scientific research," utilizing algorithms to accomplish large-scale data cleaning and preprocessing tasks, thereby focusing human intellect on theoretical innovation. Learners are encouraged to utilize technology to mine cross-disciplinary knowledge graph associations and engage in human-machine co-creation in domains such as complex system design.

In Vocational Education: Focuses on the deep integration of industry and education, and industrial technological innovation. Educational institutions are encouraged to collaborate with enterprises to build smart training bases for "AI+" vocational scenarios, cultivating highly qualified technical and skilled personnel adapted to the demands of the intelligent era. Educators are encouraged to align with the latest industrial dynamics, leading learners to solve authentic engineering challenges using digital twin technologies within highly simulated production environments. Learners are supported in using AI tools for product prototyping, market analysis, and business simulation, thereby achieving proficiency in novel production and service models.

(IV) Dynamic Adjustment Mechanism for Categorization

The classification of the three application scenarios must possess flexibility, necessitating the establishment of a normalized evaluation and dynamic adjustment mechanism. It is recommended to conduct a systematic evaluation every one to two years. The evaluation process should be spearheaded by educational authorities and involve the collaborative participation of technology experts, pedagogy experts, frontline educators and learners, and public representatives. Evaluation dimensions should comprehensively encompass technological maturity (reliability and safety), educational suitability (actual efficacy in achieving pedagogical goals), and social acceptance.

Modifications to scenario categorizations must adhere to the principle of prudence. Scenarios proposed for transitioning from "Prohibited Access" to "Restricted Use" must undergo rigorous regional or school-based pilot testing, and may only be expanded after demonstration and confirmation of safety and controllability. Scenarios proposed to be relaxed from "Restricted Use" to "Encouraged Use" must be accompanied by the simultaneous rollout of supporting literacy cultivation programs.

In the event of major foundational technological breakthroughs (such as the iterative advancement of Artificial General Intelligence, AGI) or the emergence of sudden, severe ethical risks, an emergency assessment mechanism must be immediately activated. This allows for the prompt recalibration of classification standards and control measures, ensuring that differentiated governance evolves in tandem with pedagogical practices. By forging a continuously evolving and self-correcting closed-loop governance system, it guarantees that the application of AI in the educational domain invariably progresses in a healthy manner along the established trajectory of "Context-Appropriate Beneficence" and "Differentiated Governance & Application."

VIII. Conclusion

Grounded in the new realities of artificial intelligence integrating into education, this Reference Framework systematically constructs an ethical architecture for AI educational applications, anchored by the core philosophies of "Human Agency, Synergistic Symbiosis, Context-Appropriate Beneficence, and Differentiated Governance & Application." The Framework explicitly advances fundamental behavioral guidelines — namely, to strengthen the boundaries of human-machine collaboration and foreground the value of human agency; precisely align with educational contexts and perfect differentiated governance mechanisms; fortify data security defenses and strictly safeguard the bottom line of privacy protection; and promote algorithmic fairness and transparency while establishing robust accountability and traceability mechanisms. Furthermore, it decisively addresses the typologies of risk, the attribution of accountability, and the boundaries of governance in AI educational applications. Its fundamental concern is not merely to answer whether AI can enter education, but rather to explicitly articulate in what manner, within what boundaries, and adhering to what value order AI should be integrated, striving to forge a clear normative foundation between technological advancement and the mission of nurturing human beings.

Building upon this foundation, the Reference Framework extends deeply into pedagogical practice. It concretely regulates the responsibilities and behavioral boundaries of educators, learners, and educational institutions across the diverse scenarios of basic, higher, and vocational education. By proposing the operational paradigm of "Differentiated Governance & Application," it emphasizes the prudent definition of AI's functional positioning and intensity of use across varying educational stages, tasks, and contexts. Thus, the Reference Framework is not a blanket endorsement for injecting technology into education; rather, it represents a concerted effort to ensure that AI invariably serves education through the synergistic advancement of value guidance, behavioral norms, risk prevention, and institutional arrangements.

It merits profound attention that, due to the structural disparities in technological application across different regions and demographics, educational equity remains the paramount issue in the operationalization of ethical principles. The advancement of AI in education must consistently place educational equity on par with technological innovation. We must ensure that the intelligent transformation of education is not confined merely to the iteration of technical capabilities, but serves as a vital catalyst for expanding quality educational opportunities, improving learning support conditions, and optimizing educational public services. While AI possesses the undeniable potential to improve resource allocation, elevate personalized support, enhance learning accessibility for disadvantaged groups, and mitigate certain structural shortages, it simultaneously harbors the risk of exacerbating developmental divides—across regions, between urban and rural areas, among schools, and between demographic groups. These risks are compounded by disparities in digital infrastructure, data biases, platform barriers, usage costs, variations in educator capacities, and uneven institutional support.

In essence, AI is inherently neutral; it does not organically gravitate toward equity. It can serve as an instrument to bridge divides, or it can mutate into a negative force that engenders new chasms and aggravates pre-existing inequalities. The ultimate impact of the technology is determined not by the technology itself, but by the institutional environment in which it is embedded.

Looking to the future, we must therefore ensure that the developmental dividends of AI equitably benefit all learners. Special focus must be directed toward marginalized and vulnerable populations—including rural communities, urban low-income groups, persons with disabilities, and women—ensuring they receive sustained and stable support in digital infrastructure, educational resource provisioning, technical service assistance, professional pedagogical guidance, ethical protection, and institutional safeguards. Concurrently, the imperatives of equity must be seamlessly woven into the entire lifecycle of AI educational applications: from design, research and development, and deployment, to usage, evaluation, and eventual phase-out. We must scrutinize not only whether educational opportunities are equitably accessible, but also whether the educational process is equitably effective, and ultimately, whether the educational outcomes genuinely contribute to the holistic development of the individual and the comprehensive progress of society.

The entry of artificial intelligence into education must transcend its role as a mere tool for efficiency enhancement; it must rise to become a profoundly constructive force for advancing educational justice and building a high-quality educational system for all.

Acknowledgement

Scholars from Zhejiang University, Beijing Normal University, and The Chinese University of Hong Kong authored *Ethics of Artificial Intelligence in Education: A Reference Framework*. The Expert Advisory Committee of the World Digital Education Alliance and the Professional Committee on Ethics and Governance of the Artificial Intelligence Open Alliance organized domestic and international experts to review this document. We hereby express our sincere gratitude to them. The list is as follows:

Zhejiang University: Wu Fei, Chen Jingyuan, Li Yan, Huang Changqin, Sun Lingyun, Chen Limeng, Huang Zhengxing, Liao Beishui, Zhu Qiang, Yang Yang, Kuang Kun, Jin Fangzhou

Beijing Normal University: Huang Ronghuai, Li Yanyan, Yang Junfeng,Wang Huanhuan, Zeng Haijun, Wang Junyi

Chinese University of Hong Kong: Jin Guoqing, Meng Meiling

Participating Review Experts (in chronological order of review): Yang Zongkai, Gong Ke, Zheng Qinghua, Li Yongzhi, Xue Lan, Liu Tongxi, Tang Qian, Miao Fengchun, Qin Changwei, Xiong Zhang, Hu Xiang'en, Liang Zheng, Wang Guoyu, Bai Huiren, Hou Wanjun, Shen Mingzhe, Ye Min, Gong Sen, Tao Feng, Lin Jianwu, Qin Yanding, Asha S. Kanwar, Chen Li, Mohamed Jemni, Zhan Tao, Liu Dejian, Liu Qingfeng, He Lianzhen, Xu Xueying, Torunn Gjelsvik, Habibah Abdul Rahim, Maxim Jean-Louis