ChatGPT 教育应用的伦理风险与规避进路

王佑镁　王　旦　梁炜怡　柳晨晨

（温州大学 大数据与智慧教育研究中心，浙江温州 325035）

［摘要］ 智能时代 ChatGPT 强势崛起，生成式人工智能惊艳大众，引领人工智能走向场景落地，为教育领域变革带来巨大机遇。ChatGPT 为教育创造有益价值的同时，也带来一系列伦理风险。本文阐述了 ChatGPT 的发展脉络和内涵特征，揭示 ChatGPT 教育应用存在的伦理风险，包括：数据隐私的泄露与滥用、机器算法的歧视与偏见、师生关系的弱化与破坏、学术公平的失信与失衡。本文从博弈论视角出发，从道德伦理角度剖析“教育-ChatGPT”之间的最优关系，提出 ChatGPT 教育应用伦理困境的规避建议：唤醒大众意识与保护数据隐私，警惕惯性认知与防范算法偏见，把握任务重心与调节师生关系，规训道德行为与重塑学术公平，以此增强“教育-ChatGPT”的应用价值利益，共建教育人工智能伦理规范，促进教育人工智能理性发展。

［关键词］ AIGC；ChatGPT；伦理风险

［中图分类号］G434　　［文献标识码］A　　［文章编号］1007-2179(2023)02-0026-10

一、问题提出

随着新一代人工智能技术的快速普及应用，教育工作者不得不认真思考如何更好地将人工智能与教育融合，以应对人工智能对传统教育的革命性冲击（董文娟等，2019）。近日，一款人工智能聊天机器人应用——ChatGPT 的面世，引发热议。作为一项沉淀多年的生成式人工智能研发成果，ChatGPT 利用海量数据对基于文本的输入生成类似人类的响应，在五天内积攒超 100 万用户，呈现了一场里程碑式的生成式人工智能多场景的应用效果。基于 ChatGPT 在编写与调试代码、改善课堂评估、辅助教学设计和调节互动交流等方面的潜在优势，教育人士对其充满期待（Euchner，2023）。然而，ChatGPT 可以成为支持教师教学和学生学习的有效工具，改善师生关系和给予学生情感寄托的纽带，但也可能成为影响学生注意力、学习创造性或教育公平性的干扰因素。ChatGPT 增强学习的同时，其潜在的伦理风险遭到质疑（Dowling et al.，2023），研究者探讨该如何在教育领域合乎道德地使用。基于此，本文深究 ChatGPT 应用于教育可能存在的伦理风险，通过博弈论分析“教育-ChatGPT”的关系的最优解，提出 ChatGPT 教育应用伦理风险的规避路径，以增强“教育-ChatGPT”的应用价值，共建教育人工智能伦理规范，促进教育人工智能理性发展。

二、变革 AIGC：ChatGPT 的本质属性

美国《科学》杂志刊文指出，人工智能对人类生

［收稿日期］2023-01-10　　［修回日期］2023-01-20　　［DOI 编码］10.13966/j.cnki.kfjyyj.2023.02.004

［基金项目］全国教育科学规划 2021 年度国家一般课题“教育领域人工智能应用的伦理风险与防范对策研究”（BCA210086）。

［作者简介］王佑镁，博士，教授，博士生导师，温州大学大数据与智慧教育研究中心主任，研究方向：智慧教育、人工智能教育、数字阅读 (wangyoumei@126.com)；王旦、梁炜怡，硕士研究生，温州大学教育学院教育技术系，研究方向：信息化教育；柳晨晨，博士，副教授，温州大学教育学院教育技术系，研究方向：人工智能教育、数字化学习。

［引用信息］王佑镁，王旦，梁炜怡，柳晨晨(2023). ChatGPT 教育应用的伦理风险与规避进路[J]. 开放教育研究,29(2)：26-35.

活的渗透最初是缓慢的，但 2022 年开启了一场“抢地战”。《科学》发布 2022 年度科学“十大”突破，人工智能自动生成内容（AI Generated Content，AIGC）赫然在列（Harold，2022）。人工智能生成内容指由人工智能作为内容创作主体，利用深度学习算法与场景决策模型等技术生成结果，例如人工智能写诗、人工智能绘画、人工智能视频等。人工智能生成内容正成为内容创作领域新的生产方式（王诺等，2023）。目前人工智能生成内容尽管存在争议，但毋庸置疑的是，人工智能生成内容正在开拓人类的新型创造力。2022 年 11 月 30 日，人工智能研究实验室（OpenAI）最新推出的全新聊天机器人模型 ChatGPT，一经面世便惊艳大众，引发人工智能生成内容领域新一轮变革。ChatGPT 强势崛起，引领人工智能生成内容走向场景落地，为人工智能变革日常领域带来巨大机遇。

ChatGPT 是使用对话式人工智能技术的大型语言对话模型，是一种新式人工智能技术驱动的自然语言处理工具。ChatGPT 建立于多次迭代升级的生成式预训练语言模型（Generative Pre-trained Transformer 3.5，GPT-3.5），拥有 1750 亿个模型参数，接受了 570GB 的文本训练，同时使用有监督和无监督的人工智能学习技术（OpenAI，2022）。ChatGPT 具有强大的学习和信息整合能力，能从庞杂的数字书籍、在线文章和其他媒体数据库获取所需知识，并利用这些知识与用户流畅对话和生成文本，甚至辅助用户制作新颖的图像和视频、编写和调试代码等（Rudolph et al.，2023）。

与其他搜索软件和聊天机器人不同的是，ChatGPT 采用从人类反馈中强化学习（Reinforcement Learning from Human Feedback，RLHF）的方式进行训练，并在训练过程中注重道德水平因素。这使它不仅能理解用户问题包含的人类意图，还能提供有用、真实且无害的答案，甚至对问题进行挑选并丢弃其中有危害或无意义的部分。正是这种独特的训练方式和大规模数据来源，ChatGPT 被描述为“病毒式”聊天机器人，几乎可以对任何类型的问题给出智能且类似人类的答案（King，2022）。教育工作者认为，ChatGPT 能帮助学生完成课程资源搜索、论文大纲生成和家庭作业，能协助老师完善课程设计、做好课前准备等，并认定未来 ChatGPT 一定可以为教育界带来意外之喜。

三、风险预警：ChatGPT 教育应用的伦理困境

ChatGPT 自发布以来，研究者挖掘了其大量的教育教学功能，例如提升教学成果的灵活度与创意感，重构数字导师的角色与虚拟化，提高自适应学习的易用性与精准度，促进教学策略与方式的智慧化与创造性，支持教学反馈与评价的生成性和个性化。然而，这更像是一个警示，表明人工智能可以开始与教育工作者的某些能力相抗衡。面对 ChatGPT 教育应用带来的便利与机遇，随之而来的还有伦理风险。综合当前 ChatGPT 的应用场景，本文将其存在的伦理风险问题归纳为四类：风险意识薄弱造成数据和隐私的泄露与滥用、风险认知固化造成机器算法的歧视和偏见、风险重心偏移造成师生关系的弱化与破坏、风险鉴别偏误造成学术公平的失信与失衡（见图 1）。

（一）风险意识浅薄，导致数据隐私泄露与滥用

教育领域涉及 ChatGPT 的数据隐私主要有两类：滥用 ChatGPT 提供的数据导致侵犯隐私和使用 ChatGPT 时被其获取数据导致隐私泄露。首先，ChatGPT 大量训练数据来源于互联网。它能针对用户问题提供答案。学生获得解决方案后，需学会如何正确且恰当地应用于学习。目前，ChatGPT 建立的数据库有限，未有监管机制监测其数据的真实性，难以划清数据使用与滥用的界限。教育者使用 ChatGPT 时，难以评估数据使用是否存在侵犯隐私。ChatGPT 为学生带来了诸多好处，但学生判断经验不足，难以明确数据使用的边界和范围，可能容易导致隐私泄露。第二，ChatGPT 为教育工作者带来便利的同时，仍需承担 ChatGPT 数据存储不当造成隐私泄露的风险。例如，学者利用 ChatGPT 上传研究数据以获得答案，ChatGPT 捕捉数据后如何安全地存储却不透明。其中，探讨最多的莫过于 ChatGPT 辅助编码问题，它对教授学生编写代码或许有帮助，但历经反复更正，ChatGPT 从用户获取的代码的所有权归属尚未有定论，很有可能发生第三方机构泄露或借鉴代码的事件。此外，学生寻求答案时会忽视自身信息输入的风险，仅关注 ChatGPT 的使用效果，造成数据隐私泄露。

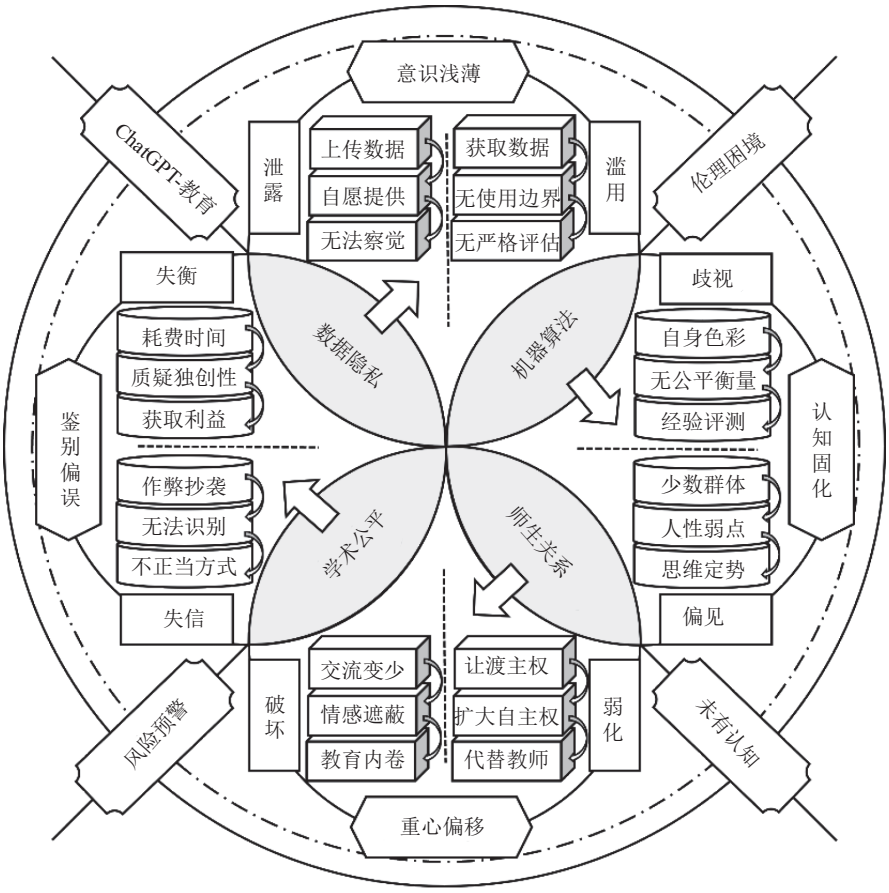

图 1　ChatGPT 教育应用的四类伦理问题

近日，国外研究者米维尔等（Mijwil et al.，2023）称，安全专家发现 ChatGPT 可能为犯罪分子提供网络犯罪的途径，使网络犯罪大众化；数据与隐私的泄露，还会诱发网络“钓鱼”事件。类似情况如果发生在教育领域，后果不堪设想，这与 ChatGPT 的初衷截然相反。

（二）风险认知固化，致机器算法歧视与偏见

ChatGPT 应用还存在两项局限：训练数据截至 2021 年且传输信息不准确。鉴于此，ChatGPT 对许多未知问题回复不够准确，不能提供精准的解决措施，甚至给出有歧视与偏见的回复。首先，训练数据受限和数据偏差会影响输出模型。事实上，ChatGPT 在训练小众数据组时就已表现出偏见：对少数群体带有歧视，不能公平对待。许多研究者测试了 ChatGPT 的算法偏见。例如，姆巴奎等（Mbakwe et al.，2023）通过观察 ChatGPT 参加美国医学执照考试的表现，发现 ChatGPT 应用医学教育的缺陷：由于发达国家的医学研究和描述无法代表全球患者，ChatGPT 训练时无法识别并将其直接应用，从而影响考试的准确性和完整性。第二，OpenAI 承认 ChatGPT 有时会给出看似合理但不正确或荒谬的答案。ChatGPT 的会话流畅性掩盖了它无法区分事实和虚构的缺陷，甚至从 ChatGPT 生成的文本能提取到充斥着人性弱点的信息。有研究提出 ChatGPT 可能生成有害指令，并提出由于 ChatGPT 的快速迭代，这些错误可能会被放大（Williamson et al.，2023）。在教育领域，ChatGPT 通过算法可能加深教育管理者的偏见与歧视，从而影响教学管理，为教师提供错误指引。学生无形中受认知偏见的影响，进而形成认知固化，造成恶性循环，降低 ChatGPT 教育应用的可持续性。此外，ChatGPT 很难区分真理与谬误，如使用者将旗鱼归为哺乳动物时，ChatGPT 不能立即指出其错误；当使用者在下一个问题指出旗鱼不是哺乳动物时，ChatGPT 又会立刻承认自己的错误，由此造成学生的认知误差和偏见。

邓建阳等(Deng et al., 2022)综述了 ChatGPT 的特点、优势与挑战，指出 ChatGPT 不仅无法对用户问题提供准确或最新的信息，也可能无法回答复杂和非常规的问题，并产生偏见或攻击性反应。如果 ChatGPT 未及时修正所包含的偏见与歧视，学生会认为机器呈现的一定是“客观”的，因而得到有歧视或偏见的知识。

(三)风险重心偏移，致师生关系弱化与破坏

师生关系是重要的人际关系之一，它直接影响学生对知识的吸收、掌握和建构。然而，ChatGPT 的出现可能消解师生的主体地位。教师让渡教学主权与学生扩大学习自主权，有可能使师生情感关系发生异化，师生交流变少与学生情感遮蔽。首先，ChatGPT 能辅助学生写诗、续写故事、学术写作与编写代码等，学生可以借助 ChatGPT完成作业与测验，学习和巩固知识，降低对教师的依赖，拓展自身学习主体性，致使出现教学主体角色混乱、学习惰性增强等问题。有学者指出，ChatGPT 能为学生提供教师未掌握的知识，学生以 ChatGPT 代替真实教师，用 ChatGPT 的观点与教师争辩，这会加剧师生关系破坏的风险。第二，ChatGPT 对患有社交恐惧症、自卑等特殊群体的学生很有帮助。学生有一对一的情感寄托，但在相对受益的情况下可能造成破坏师生情感的伦理风险。一旦学生对 ChatGPT 产生依赖，减少师生沟通，学生不愿与教师分享自己的想法，那么 ChatGPT 就不再是帮助学生最恰当的工具，而是师生关系弱化的成因。教师需要权衡 ChatGPT 的应用，趋利避害，扬长避短。当下“教育内卷”严重，学生学习动机本不是自发的，家长通常借助外力对孩子施压，教师负担重。就业率、竞赛率、升学率以及社会对学生的高期望，都可能让管理者、教学者和学习者放大 ChatGPT 的优势，从而破坏教育主体间的关系平衡。

有研究者开发了一种利用 AR、语音机器人和 ChatGPT 技术的语言学习软件，培养儿童学习外语的兴趣。研究结果表明，此类外语教学软件能改善低龄儿童学习外语的抗拒感，增进师生关系。因此，教师应在 ChatGPT 的辅助下另辟蹊径帮助学生了解所学知识的价值，尝试用它增添教学的趣味性，并考虑如何投入更多情感于人文交流中(Topsakal, 2023)。只有如此，师生关系才可能避免被破坏或恶化。显然，教育需将学生培养成人而非机器，拯救学生的灵魂而非堆积知识和认识(冯永刚等，2021)。

(四)风险鉴别偏误，致学术公平失信与失衡

“教师担心学生作弊”“教授警告 ChatGPT 帮助作弊”“ChatGPT 改变作弊者的游戏规则”等在 ChatGPT 发布一月后成为了热点讨论话题，教育研究者纷纷质疑 ChatGPT 是否会加剧学术不端和导致教育不公平(Moosmosis, 2022)。出现此类问题的原因在于：学生使用 ChatGPT 作弊和从 ChatGPT 获取内容进行改写或代写的所有权归属不明。首先，并非每位学生都能使用 ChatGPT，使用权并不平等。同时，到目前为止，抄袭检测器并未做到全面监控 ChatGPT 的输出结果，这让评估人员难以公平地辨别学生是否利用 ChatGPT 作弊。此外，剽窃问题是识别学生作业真实性的棘手难题，布置学习任务的初衷是帮助学生掌握知识，从而使教师逐步更新教学计划以符合学生学情。然而，面对 ChatGPT 的诱惑，学生能否合乎道德地使用而不以不正当的方式获取额外利益呢？学生的选择是他人无法保证的，只有学生自己才能对作弊行为承担责任。第二，学生有足够能力让 ChatGPT 先生成初步答案，然后按个人风格精炼与修改后使用，这样抄袭的监测难度更高。针对此类情况，学术不公现象会只增不减，人们或许可以质疑学术成果的独创性，但由此开展调查所耗费的时间与精力也巨大。已有期刊、出版机构发表声明，严令禁止将 ChatGPT 列为论文合著者。此外，ChatGPT 也受到多地教育部门“封杀”。有学者搜集并分析了 ChatGPT 发布一个月的 233914 篇英文推文发现，ChatGPT 能为学生的创意写作、论文写作和回答问题等带来积极影响，也指出 ChatGPT 能辅助考试并可能引发作弊行为，甚至影响教育公平(Taecharungroj, 2023)。教师虽然可以将可疑文本输入 ChatGPT 查找检验，但是 ChatGPT 每次生成的答复不同，因而很难证明论文的出处来自于 ChatGPT。同时，论文所有权的伦理法监管需要全面的证据，这导致难以追究剽窃的责任归属。

四、纳什均衡解：“教育—ChatGPT”博弈分析模型

人工智能技术的革命是一场利害关系的博弈。

ChatGPT 能为教育增添“光明”，也会带来“阴暗”。这不仅与教育紧密相关，与社会公众紧密相关，也与 ChatGPT 背后的技术管理人员相关。在明晰 ChatGPT 的内涵后，面对 ChatGPT教育应用的诸多伦理困境，本文借鉴博弈论的分析模型构建互惠互利的教育与 ChatGPT 之间的关系，寻求“教育—ChatGPT”的纳什均衡解。

（一）博弈论与纳什均衡解

博弈论源自《孙子兵法》，最初主要用于象棋、桥牌比赛，后延伸至不同学科，在生物学、经济学、管理学等应用广泛。它指研究多个个体或团队之间在特定条件制衡下的对局中，利用相关方的策略谁受影响实施对应策略的学科（卑力添，2019）。完整的博弈过程主要受五个方面影响：博弈的参加者、博奕信息、博弈次序、博弈方可选择的全部行为或策略的集合和博弈方的收益。基本模型有“囚徒困境”“智猪模型”与“斗鸡模型”等。博弈论的教育应用也很多。例如，李广海等（2022）从博弈论视角审视“双减”政策下政策执行的本质、主体与目标；俞慧刚（2020）从博弈论视角探讨校企合作的演化过程，揭示校企合作的内在演变规律。

纳什均衡解指在一组策略的组合中，所有参与者面临他方不改变策略时，我方此时的策略是最好的。在纳什均衡点上，理性的参与者不会单独改变本方策略。它为博弈论提供了重要的分析手段，使研究者在博弈结构中找寻有意义的结果。关于“纳什均衡”，教育界一直持续关注，田夏彪（2016）针对课堂教学的纳什均衡危机，提出了相应的解决方案；李宝斌（2021）基于博弈论思考与强化教育惩戒博弈的立德树人纳什均衡支点，以更好地落实教育惩戒规则。

（二）“教育—ChatGPT”博弈模型

进入人工智能生成内容时代，ChatGPT 的出现改变了技术革命的格局，为教育带来新的希望，也对教育与 ChatGPT 之间的关系提出考验：研究者对拥抱还是禁止 ChatGPT 提出了不同看法。本文基于博弈论分析如下：假如教育和 ChatGPT 都是理性方，即理性方的行为目的是获得最大效用值。在道德伦理中，教育和 ChatGPT 在作出自身判断前不知道彼此的博弈行为，但对以往的博弈结果十分清楚。假设按照以下模型进行静态的博弈分析（易开刚等，2019），通过 10 以内的整数量化博弈双方的受益情况，构建“教育—ChatGPT”受益矩阵模型。它就可以分为四种方式：如果教育做到“优质使用，积极接纳”，ChatGPT 做到“优质开发，积极监管”，则 ChatGPT 教育应用的伦理困境能得到及时的规避或防范，各方受益均为 5 个单位；ChatGPT 如果仅做到“一般开发，消极监管”，那么教育方因 ChatGPT 出现算法偏见、输出信息不准等问题受益减少 2 个单位，ChatGPT 方因教育带来的不良影响受益减少 1 个单位；如果教育实行“一般使用，消极接纳”，ChatGPT 做到“优质开发，积极监管”，那么 ChatGPT 方因师生关系恶化、学术公平质疑等问题受益减少 2 个单位，教育方因 ChatGPT带来的不良影响受益减少 1 个单位；ChatGPT 如果仅做到“一般开发，消极监管”，那么 ChatGPT 教育应用的伦理困境会层出不穷并得不到治理，而导致两方受益都减少 3 个单位（见图 2）。

在此矩阵模型中，本文将代表受益最高的坐标（5，5）描述为教育与 ChatGPT 之间的“合作共善”关系，将代表受益一般的坐标（3，4）与（4，3）表述为教育与 ChatGPT 之间的“相互适应”关系，将代表受益不佳的坐标（2，2）表述为教育与 ChatGPT 之间的“双轨对立”关系。总结来说，只有当教育与 ChatGPT 都选择最优策略，即教育做到“优质使用和积极接纳”，ChatGPT 做到“优质开发和积极监管”，才能得到二者之间的纳什均衡解，双方受益最大化。因此，合作共善应成为 ChatGPT 建设的“零号”道德伦理准则，更好地构建 ChatGPT 教育应用的新伦理契约乃是当下的重点讨论方向（见图 3）。维贝克的道德物化理论提出，人工智能教育应用中，人们可以充分发挥技术的调节作用，形成双向促进和协同发展的“人—技术”关系（孙田琳子，2021）。因此，教育工作者应各司其职，充分发扬 ChatGPT 的优势，规避 ChatGPT 可能带来的风险与挑战；ChatGPT 开发人员应精益求精，积极监管 ChatGPT 使用效果，形成从博弈到共善的转变。

五、合作共善：ChatGPT 教育应用伦理风险规避路径

一直以来，教育人工智能的伦理问题难以得到有效约束和监管，一些国家建立实验室或成立机构发布治理指南，但主要靠教育主体自我伦理素养的

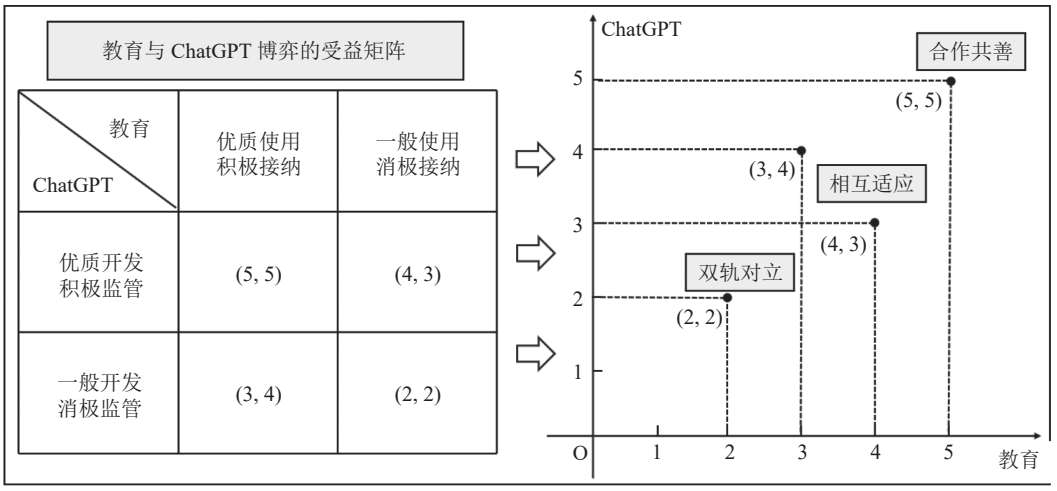

图 2　教育—ChatGPT 博弈分析模型

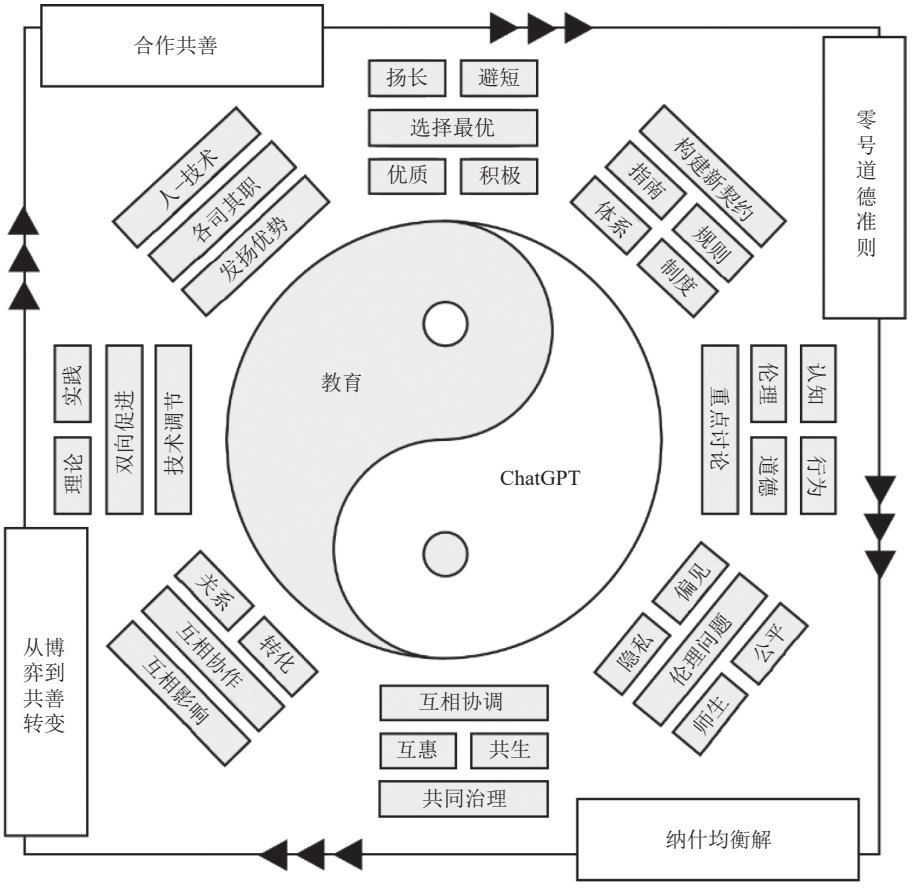

图 3　教育—ChatGPT 之间最优策略解

抉择。本文基于博弈论找寻教育与 ChatGPT 二者关系的最优解，针对 ChatGPT 教育应用的四类伦理问题，从相关主体的视角阐述“合作共善”视域下伦理问题的消解路径(见图 4)。相关主体包括政府管理者、ChatGPT 技术工作者、高级伦理研究人员、社会大众、学校管理者、教师与学生，各自承担不同角色：决策者、劳动者、监督者、导向者、把关者、守卫者与实施者致力于将“教育—ChatGPT”的协同融合环境中由被动、无约束、无边界与延时机制转变至主动、有约束、有边界与及时机制，塑

造良好的 ChatGPT 教育应用关系。

（一）“教育—ChatGPT”由被动变主动：唤醒大众意识，保护数据隐私

在数据隐私方面，政府管理者应设立实时监控与调查部门，从各方面考察 ChatGPT，发现问题及时制止，并致力于颁布权威的道德使用指南监管 ChatGPT 获取数据，并根据有关隐私法律法规加以处理（Lund et al.，2023）。ChatGPT 背后的技术工作者是最需要监管的人员之一，他们能获取教师或学生的第一手数据，应不断提升道德伦理素养，明确网络犯罪的严重性。高级伦理研究人员和社会大众需积极推动事态发展，在出现 ChatGPT 侵犯师生隐私或泄露师生研究数据时，做好记录备案和正面舆论引导，减少类似事件发生。学校管理者作为教育与 ChatGPT 之间合作共善的把关者，既受规制又有一定的权利，应重视学生数据隐私侵犯问题，提升伦理意识，主动组织有意义的讲座、座谈提醒师生，并在出现问题后公平且迅速地裁决。教师特别是中小学教师将 ChatGPT 应用于教学时，不仅要提醒学生不能轻易上传数据和隐私，还应保证学校电子产品安全软件达到最新状态、使用强密码和安装防病毒软件等（van Dis et al.，2023）。高校教师掌握人工智能工具后，需提前预判，提醒学生勿随意输入或生成实验数据、调研数据等。对于学生而言，具备安全隐私意识是最重要的，应多组织学生学习人工智能侵犯隐私的案例，增强学生自我保护的责任感。总之，教育界和 ChatGPT 需共同具备保护数据和防范被侵犯的意识，制定严格的监管体系，主动控制信息，打造绿色健康的 ChatGPT 教育应用环境。

（二）“教育—ChatGPT”由无约变有约：警惕惯性认知，防范算法偏见

ChatGPT 接受了大量互联网代码和信息数据的操纵训练，能给予人性化的回复。针对 ChatGPT 自身训练存在的算法歧视与偏见，政府领导者、高级伦理研究人员与社会大众是偏见和歧视的来源，不能用固有的认知看待 ChatGPT，只有找到偏见的源头才能解决它的偏见问题。伦理研究人员和社会大众应给予正确的导向，引领 ChatGPT 技术人员平等训练模型。学校管理者虽无法改变 ChatGPT 自带的算法偏见，但可以与教师商讨如何恰当使用 ChatGPT，使学生避免接触偏见性知识，尤其是

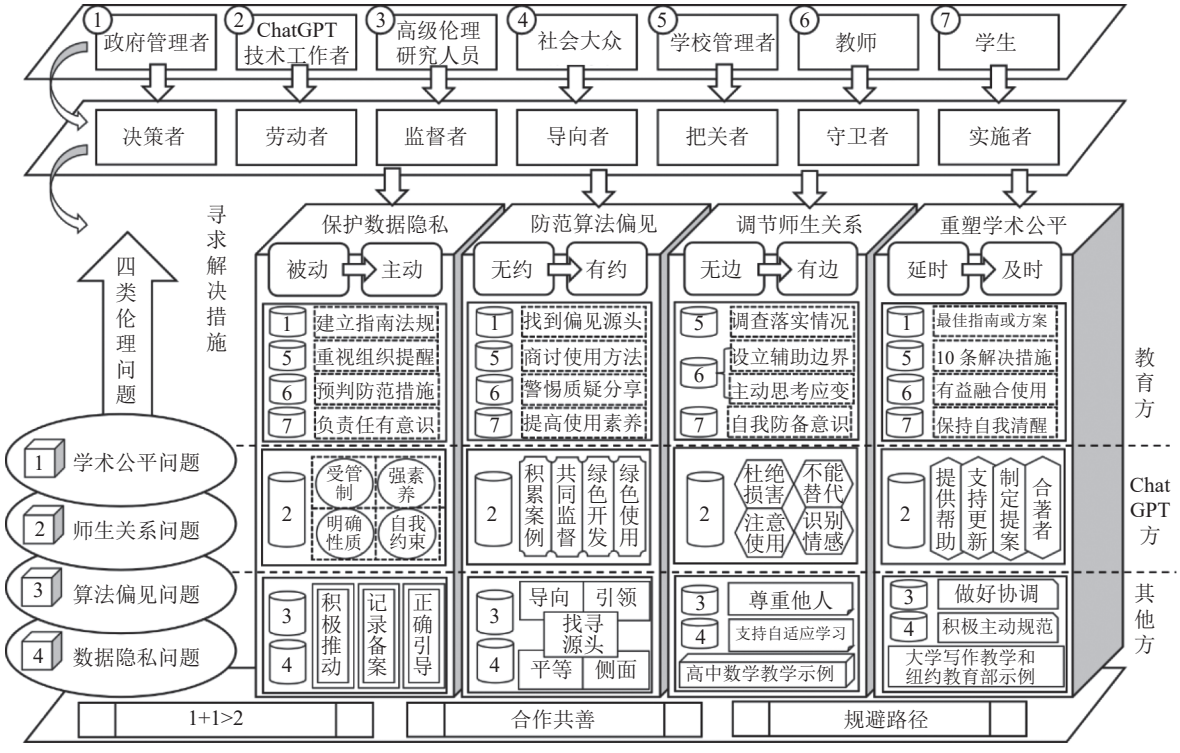

图 4　ChatGPT 教育应用伦理风险消解路径

政治、历史等学科教师，更需要在大规模使用前反复测验，确保课堂教学健康持续。学生使用 ChatGPT 时，教师应提醒学生脱离惯性思维与认知，勇敢提出疑问，进而再与社会大众分享，扩大 ChatGPT 教育应用的可持续性。算法歧视和偏见对个人影响或许很小，但长远来看往往会加剧教育不公平，扩大数字鸿沟（贺译葶，2022）。算法本身并不完美，继承了人类设计者对特定事物的偏见，更何况 ChatGPT 是由大量互联网信息训练而成的。只有教育者做到源头切割，把握事情真相，才能避免误导学生。ChatGPT应积累相关案例和偏见性的对话交流，从没有约束到有约束，共同监督 ChatGPT 的开发与使用，这对创建绿色的 ChatGPT 教育应用环境具有重要意义。

（三）“教育—ChatGPT”由无界变有界：把握任务重心，调节师生关系

师生关系伦理问题的解决主要在于学校管理者、教师和学生三者间的任务分配。首先，对 ChatGPT 在学业上的强大帮助，学校管理者要调查教师将 ChatGPT 用于职责范围内的教学落实情况，保证教师教学安排合理且有效，抽查学生使用 ChatGPT 对教师的评价。教师应设立技术与教学的辅助边界，做到不偏不倚，发挥 ChatGPT 在写作、编码和反馈等方面的长处，鼓励学生多思考，批评性接受 ChatGPT 提供的答案。例如，英语写作中，ChatGPT 能为学生提供清晰、连贯和有洞察力的提示，使教师猝不及防。这就需要教师重新评估教学内容，思考如何有效地为学生提供真实学习体验。因此，教师在 ChatGPT 的助力下，不能一成不变和坐享其成，需主动思考教学革新，与学生沟通，增加课堂趣味性，调节师生关系。学生应具备自我防范意识，不能过多依赖 ChatGPT，明确教师和人工智能的区别，厘清角色定位。其次，针对特殊群体存在师生情感异化的现象，从政府管理者至教师，乃至同级学生群体等各方人士需关爱与配合。ChatGPT 能实现个性化的自适应学习对此类学生帮助极大。在 ChatGPT 的辅助下，教师应利用更多时间与学生交流，使学生全面发展。教育相关主体要明晰 ChatGPT 与教师间的角色分配，划清人与机器的边界，使学生把握不同学习任务的重心，杜绝 ChatGPT 有损教师形象和身份等问题。教师应掌握 ChatGPT 使用注意事项，更好地把握学生的真实情感，走入学生内心，为构建绿色的 ChatGPT 教育应用环境奠定坚实基础。

（四）“教育—ChatGPT”由延时变及时：规训道德行为，重塑学术公平

一直以来学术公平是教育界追求的最终目标，然而学生是否诚信、教师是否一视同仁取决于自我道德认知和行为操作。目前，研究者针对 ChatGPT 可能会造成的抄袭问题，提出了相应的治理方法和建议（Rudolph et al., 2023）。政府管理者应制定 ChatGPT 教育应用的指南或方案。伦理研究人员和社会大众做好各教育区域的协调工作，主动规范 ChatGPT 的使用，最大限度地保留优势，减少负面影响。ChatGPT 应用在于发挥学生潜力而不是作弊，应为学生带来更好的学习体验和推动心智能力得到拓展，帮助学生发展能力而不是在思维中取代他们。学校管理者和教师防止学生抄袭的方法很多（见图 5）：1）监控 ChatGPT 在学术环境的使用，一旦发现作弊或其他不道德目的，立即采取行动；2）向学生普及抄袭或剽窃的危害性，并阐述案例；3）布置作业和制定考试内容时，可根据 ChatGPT 的内容只更新于 2021 年而设定；4）作业题与课堂材料和笔记相连；5）尝试布置复杂和含有特定上下文的作业，并增加及时性考察；6）避免布置过于笼统或易解决的任务；7）避免布置已被广泛讨论或分析

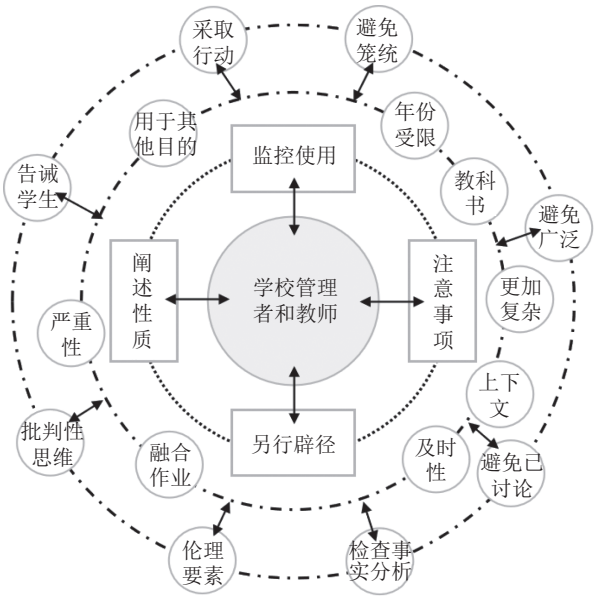

图 5　学术公平问题的治理策略

却仍有争议的主题、趋势等作业题目；8)检查学生对客观事实的分析；9)将批判性思维和伦理要素纳入课程，帮助学生养成负责任地使用 ChatGPT 的技能和习惯；10)如教师无法规避 ChatGPT的影响，可专门布置需借助 ChatGPT 的作业，如论文写作和编写代码等。

在学校管理者和教师的监督下，学生要时刻保持清醒，不能被眼前利益蒙蔽，明确获得真实知识才是关键。技术工作者在学术诚信方面也可以为剽窃检测机构提供协助，支持及时更新，使 ChatGPT 良性发展。同时，对 ChatGPT 的内容所有权和合著者问题，应制定保护提案，如发现有人用于获取自身利益，就追究其责任或给予处分等。目前，GPT-2 输出检测器虽可以检测文本是否用于 GPT 技术生成并确认真假，但不能指向互联网中的任何来源。教育是国家发展的重要根基，因此，未来发布类似新产品时，应建立教育领域的治理部门，制定突发情况的解决措施(IBL News, 2022)。

六、结论与讨论

事实上，由于 ChatGPT 拥有超强的自然语言处理能力，许多师生动了“歪心思”，有学术机构及管理部门甚至为此公开发布使用 ChatGPT 的禁令，但人工智能技术的发展与介入势不可挡(Thorp, 2023)。教育工作者有必要积极拥抱生成式人工智能技术，尽可能使用人工智能推进教学创新，同时精准制定解决措施，规避其不良影响和伦理风险。本文归纳总结 ChatGPT 教育应用的四类伦理困境，基于博弈论的分析模型探讨“教育—ChatGPT”之间的纳什均衡解，明确以合作共善的方式提升师生的伦理意识与规范行为；从合作共善的视角出发提出相应的解决路径：注重保护 ChatGPT 教育应用的数据隐私，防范其可能带来的算法歧视与偏见，正确利用并鼓励师生多情感交流，规训学生道德行为创造公平的学术环境，实现“教育—ChatGPT”从博弈到共善的转化，推动教育智能化和公平化发展，为未来教育提供使用 ChatGPT 的伦理指南。随着越来越多人工智能生成内容产品的研发落地，教育领域需要保持倾听利益相关者的想法和观点的习惯，以合乎伦理的方式推动人工智能生成内容的教育应用，增强教学与学习的创新性和价值感，探索生成式人工智能教育应用的场景与伦理规范。

［参考文献］

[1] 卑力添，蒋柯，李先春，熊哲宏(2019). 博弈论视角下的超扫描多人互动任务新模型 [J]. 心理科学进展，(7)：1284-1296.

[2] Deng, J., & Lin, Y.(2022). The benefits and challenges of chatGPT: An overview[J]. Frontiers in Computing and Intelligent Systems, 2(2): 81-83.

[3] 董文娟，黄尧(2019). 人工智能背景下职业教育变革及模式建构 [J]. 中国电化教育，(7)：1-7+45.

[4] Dowling, M., & Lucey, B.(2023). ChatGPT for (finance) research: The bananarama conjecture[J]. Finance Research Letters, 2023: 103662.

[5] Euchner, J.(2023). Almost human[J]. Research-Technology Management, 66(2): 10-11.

[6] 冯永刚，陈颖(2021). 智慧教育时代教师角色的“变”与“不变”[J]. 中国电化教育，(4)：8-15.

[7] Harold, M. (2022). The 10 biggest scientific breakthroughs of 2022[EB/OL]. [2022-12-22]. https://theweek.com/in-depth/1019386/the-10-biggest-scientific-breakthroughs-of-2022.

[8] 贺译葶(2022). 人工智能在信用监管中应用的法律风险及其应对 [J]. 甘肃社会科学，(4)：142-150.

[9] IBL News. (2022). Cheating on essays in higher education through chatGPT alarms academia[EB/OL]. [2022-12-28]. https://iblnews.org/cheating-on-essays-in-higher-education-through-chatgpt-alarms-academia/.

[10] 李宝斌，陈慧(2021). 教育惩戒规则落地的利益相关者博弈心理研究 [J]. 中国人民大学教育学刊，(4)：151-163.

[11] King, M. R.(2022). The future of AI in medicine: a perspective from a chatbot[J]. Annals of Biomedical Engineering, : 1-5.

[12] 李广海，李海龙(2022). 博弈论视角下“双减”政策执行的阻滞与疏解 [J]. 现代教育管理，(6)：10-19.

[13] Lund, B. D., & Wang, T.(2023). Chatting about chatGPT: How may AI and GPT impact academia and libraries?[J]. Library Hi Tech News, 2023: 0741-9058.

[14] Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., & Dagan, A.(2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education[J]. PLOS Digital Health, 2(2): e0000205.

[15] Mijwil, M., & Aljanabi, M.(2023). Towards artificial intelligence-based cybersecurity: The practices and chatGPT generated ways to combat cybercrime[J]. Iraqi Journal For Computer Science and Mathematics, 4(1): 65-70.

[16] Moosmosis. (2022). Ethics of using chatGPT openAI in writing essays for students[EB/OL]. [2022-12-18]. https://moosmosis.org/2022/12/20/ethics-of-using-chatgpt-openai-in-writing-essays-for-students/.

[17] OpenAI. (2022). ChatGPT: Optimizing language models for dialogue[EB/OL]. [2022-11-30]. https://openai.com/blog/chatgpt/.

[18] Rudolph, J., Tan, S., & Tan, S.(2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?[J].

Journal of Applied Learning and Teaching, 2023,6（1）：9.

[19] 孙田琳子（2021）. 人工智能教育中“人—技术”关系博弈与建构——从反向驯化到技术调解 [J]. 开放教育研究，27（6）：37-43.

[20] Taecharungroj, V.（2023）. “What can chatGPT do?” analyzing early reactions to the innovative AI chatbot on twitter[J]. Big Data and Cognitive Computing, 7（1）：35.

[21] Thorp, H. H.（2023）. ChatGPT is fun, but not an author[J]. Science, 379（6630）：313-313.

[22] Topsakal, O., & Topsakal, E.（2023）. Framework for a foreign language teaching software for children utilizing AR, voicebots and chatGPT（Large Language Models）[J]. The Journal of Cognitive Systems, 7（2）：33-38.

[23] 田夏彪（2016）. 走出课堂教学消极“纳什均衡”危机 [J]. 教学与管理，（15）：78-80.

[24] van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L.（2023）. ChatGPT：Five priorities for research[J]. Nature, 614（7947）：224-226.

[25] 王诺，毕学成，许鑫（2023）. 先利其器：元宇宙场景下的 AIGC 及其 GLAM 应用机遇 [J/OL]. 图书馆论坛：1-9 [2023-02-15]. http://kns.cnki.net/kcms/detail/44.1306.G2.20220916.1710.005.html.

[26] Williamson, B., Macgilchrist, F., & Potter, J.（2023）. Re-examining AI, automation and datafication in education[J]. Learning, Media and Technology, 48（1）：1-5.

[27] 易开刚，张琦（2019）. 平台经济视域下的商家舞弊治理：博弈模型与政策建议 [J]. 浙江大学学报（人文社会科学版），（5）：127-142.

[28] 俞慧刚（2020）. 从合作博弈到利益均衡：高校学生社团与企业合作的动态演化过程 [J]. 高教探索，（2）：77-82.

（编辑：赵晓丽）

Ethical Risks and Avoidance Approaches of ChatGPT in Educational Application

WANG Youmei, WANG Dan, LIANG WeiYi & LIU Chenchen

（*Research Center for Big Data and Smart Education*, *Wenzhou University; Wenzhou* 325035, *China*）

Abstract: *ChatGPT has emerged strongly in the era of intelligence, and the generative Artificial Intelligence (A.I.) has amazed the public, leading A.I. further toward extensive implementation with breakthrough opportunities for changes in education. While ChatGPT brings in beneficial value for education, it also introduces a series of ethical risk issues. This article describes characteristics of ChatGPT, and its development to reveal the ethical risks of ChatGPT applications in education including leakage and abuse of data privacy, discrimination and bias of machine algorithms, negative impact on teacher-student relationship, and loss of trust and imbalance of academic fairness. Based on the game theory, the article analyzes the "education-ChatGPT" relationship from a moral and ethical perspective and proposes to solve the ethical issues by enhancing public awareness, protecting data privacy, alerting inertia perception and preventing algorithmic bias, fostering and regulating teacher-student relationship. To enhance the value of "education-ChatGPT", the article suggests to establish an ethical code of educational A.I. and to promote the rational development of educational A.I. for education.*

Key words: *AIGC*; *ChatGPT*; *ethical risks*