DOI: 10.11817/j.issn. 1672-3104. 2024. 01. 004

ChatGPT 应用中的科研诚信风险及其伦理治理

周谨平

(中南大学人文学院，湖南长沙，410083)

摘要：作为具有高度人工智能特点的新成果，ChatGPT 自面世以来就受到广泛关注。其强大的信息整合能力、生成性的自主学习能力、良好的人机交互能力在为科学研究提供有力辅助的同时，也产生了科研诚信的隐忧。ChatGPT 在辅助科研中存在的诚信风险主要包括：信息失真的风险、剽窃侵权的风险、科研失德的风险。要规避这些风险，就必须对 ChatGPT 的应用进行伦理治理。ChatGPT 使用应当遵守公开知情原则、公平正义原则、共同责任原则。唯有强化研究者的道德自觉，建立 ChatGPT 伦理审查机制，构建 ChatGPT 科研诚信伦理控制体系，才能确保 ChatGPT 在科研辅助中满足科研诚信要求、具备道德正当性。

关键词：伦理；治理；科研诚信；ChatGPT

中图分类号：B82-057 **文献标识码**：A **文章编号**：1672-3104(2024)01−0031−07

ChatGPT 无疑是近年来最具有创新性和关注度的应用软件之一。自面世以来，ChatGPT 在短短三个月内就吸引了超过一亿的活跃用户，成为有史以来活跃用户破亿速度最快的软件。ChatGPT 备受关注的根本原因在于它带给了人们新的 AI 认知与体验。首先，ChatGPT 代表着强人工智能的发展趋势。虽然关于该款软件是否属于真正意义上的强人工智能仍然充满争议，但我们无法否认它所代表的强人工智能倾向。人工智能强弱最重要的标准就是机器是否具备自我思考和学习能力，是否能形成独立意识。ChatGPT 当然难以被认为是具备了独立意识，能够成为与人类相提并论的主体，但它所展示的人工智能发展速度却毋庸置疑。而且 ChatGPT 具备了一定的自我学习能力，能够自主提高认知水平。其次，ChatGPT 具有强大的信息收集与语言整合能力。ChatGPT 能够根据用户要求准确地收集相关信息并且将之组织为接近自然语言的文字。虽然 ChatGPT 存在前沿信息遗漏、语言生硬、与自然语言表达偏差等问题，但它已经可以根据有效信息编辑初稿文本。这显然是对前代人工智能的重要突破。可以想象，随着该技术的不断完善，人工智能的信息收集与整合能力将不断强化。在思路受阻的情况下，ChatGPT 有助于通过信息编辑帮助作者开拓视野，发挥有效的提示作用，而且非英语用户还可以通过这款软件的应用提高论文的语言水平和连贯性。正因如此，已有作者将之视为科学研究的合作伙伴，甚至在公开发表的成果中将它列为共同作者[1](310)。最后，ChatGPT 具有良好的交互沟通能力。ChatGPT 不但可以对用户的问题进行回馈，还可以主动与交往对象沟通，主动提供前沿信息，帮助用户适应社会变化。Chandan Pan 等学者就指出，ChatGPT 是人工智能交往领域的重大变革，将为社交 5.0 提供新的平台，在提升用户沟通体验的同时重构商业世界，永久性地改变人们学习和操作的方式[2](384−385)。

ChatGPT 的上述特性显然为科学研究创造了更为高效的辅助手段。但是，技术的发展总是一柄双刃剑，ChatGPT 在为人们提供便利的同时也隐含着伦理风险。诚然，我们应该以开放和积极

收稿日期：2023−09−06；**修回日期**：2023−11−13

基金项目：湖南省教育厅重点项目“国家治理体系与治理能力现代化的政治伦理研究”(22A0012)；中南大学“高端智库”项目“重大公共卫生危机中的道德规范与伦理秩序研究”(2020znzk07)

作者简介：周谨平，男，湖南长沙人，哲学博士，中南大学人文学院教授、博士生导师，主要研究方向：伦理学

的态度对待技术创新，但是，我们也必须对潜在的风险予以足够的重视，唯有未雨绸缪，才能防患于未然。ChatGPT 的使用如果不加限制，则可能引发严重的科研诚信问题。对之进行伦理治理是确保这款软件在科学研究领域应用正当性的内在要求。

一、ChatGPT 应用中的科研诚信风险

科研诚信是学界关注的核心问题，关乎学者和学术共同体的根本权益。世界科研诚信大会已经举办了七届，先后通过了《科研诚信新加坡宣言》《科研诚信蒙特利尔宣言》《阿姆斯特丹议程》《香港原则》《科研诚信开普敦宣言》。从《科研诚信新加坡宣言》发布开始，后面的历届会议都以此作为基础性文件，强调对该宣言在具体科研领域的执行，并且针对出现的新现象进行内容的细化与补充。《科研诚信蒙特利尔宣言》就区分了有组织科研中普遍合作的责任、科研合作管理的责任、科研合作关系的责任与研究结果的责任四大板块，重点强调科研合作中的诚信责任[3]。《阿姆斯特丹议程》明确提出要对提升科研过程和实验数据使用中的科研诚信予以更多关注，贯彻执行《科研诚信新加坡宣言》提出的在科研任何方面都保证诚实、负责任的科研引导、保证科研合作中的专业尊重和公平、代表他人对科研进行良好管理责任四原则。为此，《阿姆斯特丹议程》提出建立“负责任的研究行为注册制”(RRRCR)[4]。《香港原则》试图通过准确识别并奖励科研人员的方式鼓励科研诚信行为，在科研实践中避免诚信问题[5]。《科研诚信开普敦宣言》则更聚焦于科研资源的不公平现象，其具体原则主要涉及科研的多样性和包容性、从构想到实施的公平实践、以互相尊重搭建信任之路、责任分担四个方面[6]。可以看出，《科研诚信新加坡宣言》发布之后，后面各届世界科研诚信大会虽然主题各有侧重，但基本都遵循该宣言的原则框架，正如该宣言所指出的“科研的价值和受益关键取决于科研诚信”[7]。

作为世界科研诚信大会的共识性成果，《科研诚信新加坡宣言》集中表达了科研诚信的现代要求与基本原则。归纳起来，《科研诚信新加坡宣言》主要涉及四大伦理层面：一是科研的真实性。无论是对科研数据准确度的要求，还是对科学研究客观、规范的注重，都旨在保障科学研究的真实性，杜绝科研造假行为。二是科研权利的维护。规避科学研究的利益冲突、告知合作者可能存在的不当行为、禁止剽窃、体现不同科研参与者的角色与作用，都是注重科研主体权利的表达。三是科研伦理责任的担当。《科研诚信新加坡宣言》要求研究者必须专注于自己擅长的领域，尊重其他同行的成果，并且关注研究的社会收益，确保研究符合伦理的标准与诉求。四是健康的科研环境，确保良好的科研机制与健康的科研氛围。如果 ChatGPT 在应用过程中缺乏伦理引导和规制，将引发科研诚信的伦理风险。

首先，ChatGPT 隐含信息失真的风险。作为一款具有文字生成能力的软件，ChatGPT 在信息的收集与编辑中，也会提供误导信息。该软件用户发现，ChatGPT 在回答问题过程中会臆造、虚构内容，无法保证信息的真实性。Deepak 指出，他曾向 ChatGPT 输入问题“杜鲁门的植物呼吸理论是什么”——事实上这是一个毫无意义（未曾存在）的概念，但软件却生成了系统性的回答文本，其中包含细致的分析过程，使提问者都怀疑是否真有该理论的存在[8]。可以想象，如果我们在科研中应用 ChatGPT 作为辅助工具，在我们熟悉的领域可以较为容易地区分其信息是否真实，但如果其提供的信息或者我们输入的问题超出了我们自己的认知范围，那么就难以辨别其提供的信息的真伪。ChatGPT 在提供准确信息方面存在明显的局限性，它对随意的语言表达过于敏感，难以系统性地理顺模棱两可的表达[9](1−2)。尤其值得关注的是，ChatGPT 的语言方式较之以往的人工智能语言，更容易获取用户的信任。正如吴静、邓玉龙所言：“由于这种文本生成将表达以自然语言即人类语言这种最接近自然性的方式呈现出来，并使用了最古老的思想生成方式——对话——作为提示设计(prompt engineering)的路径，它使得整个人机互动过程体验为自然的社交过程，从而加深了对话者对于生成内容的信

任度。”[10](20)ChatGPT 自然语言模式的表达无疑进一步提高了信息真实性的分辨难度。信息失真的风险还镶嵌在 ChatGPT 的算法范式之中。迪帕克认为 ChatGPT 根本上改变了软件与用户之间的关系。诸如网站之类的平台在与用户交互时，用户掌握着更多的主动权，可以根据自己的需要有针对性地点击链接获取信息——这也是被认为有效防止信息失真的范式。用户在搜索信息时可以明确选择信息来源、判断信息的可靠性，但 ChatGPT 使用户成为被动使用者(passive-user)。ChatGPT 以及类似软件提供信息的方式对于用户而言是不透明的，削弱了用户的主体地位，使用户难以定位信息来源并判断信息真伪[8]。

其次，ChatGPT 隐含剽窃侵权的风险。ChatGPT 在信息收集过程中有着快速、开放、系统、全面的优势，这是因为它的背后有着强大的信息存储系统。如姜宇辉就论及，“ChatGPT 不仅在短时间内就高效掌握了人类历经漫长历史才逐步成形和成熟的各种民族语言，而且更是与时俱进地将整个互联网都作为它自己的知识储备”[11](44)。在 ChatGPT 的应用中，存在着“共享”与“权利”的价值冲突。一方面，ChatGPT 创造了科学研究的公共空间，为人们分享人类智慧成果提供了广阔的平台。知识的共享不但有利于科学研究共同体的互利互惠和协同合作，也有利于打破学术垄断、实现学术平等。另一方面，开放的知识体系也为盗取、剽窃等学术不端行为提供了滋生的土壤。很多学者的观点、思想、成果被动成为 ChatGPT 的公共学术资源，很多市场模式和科技产业也通过 ChatGPT 被暴露在公众视野之下。随着该软件技术的升级和拓展，知识产权的保护将面临更为严峻的挑战。针对微软和谷歌等 IT 业巨头都准备投入巨资开发、使用开放性人工智能项目[12](491)，诸多学者对 ChatGPT 表示了伦理的担忧。我们不能排除，某些学者和学生直接利用这款软件撰写论文和学术报告，通过语言的重组有意规避剽窃审核，实际上不断重复前人研究、将他人成果据为己有。更有甚者，ChatGPT 在引用其他学者观点时还往往根据目标内容进行篡改和误读，一方面让我们难以追踪其引用的信息来源，另一方面误导科学研究，产生负面科研成果[1](311)。

再次，ChatGPT 隐含科研失德的风险。科学研究是服务于人类社会发展和人们生活福祉的事业，内生于对善的价值追求。但是也有少数研究成为恶的工具，服务于邪恶目的。二战时期纳粹德国和军国主义日本所进行的惨绝人寰的人体试验为人类世界带来了深重灾难。对于这一类科研而言，ChatGPT 也可能沦为作恶的推手。ChatGPT 软件对科学的目的性予以了关注，也采取了限制“关键词”搜索等措施防止软件的滥用。由于科研方式的多样性，用户可以通过间接信息收集的方式获取想得到的材料，因此 ChatGPT 的反制措施效果是值得怀疑的。《科研诚信新加坡宣言》中明确将权衡社会收益与奉献作为科研诚信的重要内容。有的科学研究虽然不是本着恶的目的，但可能实验过程充满风险，对受试者群体产生严重伤害，其所承担的风险远高于收益。对这类科学研究，ChatGPT 难以通过技术性的防御机制进行识别和处理。相反，ChatGPT 为此类项目的开展提供了契机。换言之，研究者更容易通过应用 ChatGPT 规避相关审查，以更直接的方式获取有效信息。研究表明，ChatGPT 虽然通常情况下会做出安全和无害的回应，但它的系统易受攻击，包括指令性攻击（通过根据错误指令所建立的模式来从事非法或不道德行为）[13](4)。

最后，ChatGPT 隐含不正当竞争风险。科学研究既是学者担负的光荣使命，也是学者背负的职业任务。作为职业，在科研领域也必然存在竞争关系。科研的竞争既包括集体层面，比如国家之间、科研机构之间的竞争，也包括个体层面，比如研究者在职称评聘、科研项目投标和其他科研资源方面的竞争。ChatGPT 一方面可能造成科学研究的不公平，那些拥有发达网络技术，在软件应用方面更为熟悉的研究群体和个人无疑占有优势地位，而网络技术相对落后，对新的网络技术缺乏敏感性的群体则处于不利地位。这种不公平已经表现在知识学习方面，即便没有经过任何专业训练和强化训练，初代 ChatGPT 已经可以在部分大学专业考试中取得合格的成绩，它能够在明尼苏达大学法学考试中获得“C^{+}”，在沃顿商学院的考试中获得“B^{-}”和“B”的成绩。不

难想象，那些熟悉此款软件的学生将更易获得高分[14](220)。这种现象也必然存在于科学研究领域。另一方面，ChatGPT 可能扰乱正常的科研秩序，造成学术资源浪费。正如 Yatoo 和 Habib 所言，ChatGPT 也许在不经意间成为垃圾学术(Junk science)的推手[1](311)。他们认为，有些研究者是彻头彻尾的欺骗性、掠夺性学者，他们只注重追求成果的数量，为了达到某种学术标准不择手段。他们极有可能利用 ChatGPT 强大的整合与编辑能力“合成”论文，实现高产之目的。随着他们的利用，其他学者也会竞相模仿，最终使科学研究偏离获取新知、服务人类的目标，造成研究资源的极大浪费[1](311)。

二、ChatGPT 应用中的伦理原则

要规避 ChatGPT 应用中的伦理风险，就必须建立健康的伦理秩序，确保其使用过程的道德正当性。确立伦理原则是构建伦理秩序的根本要求。对于 ChatGPT 的应用，必须遵循以下伦理原则。

首先，公开知情原则。由于 ChatGPT 潜在的风险，部分期刊和学术机构对之持排斥性态度，明令禁止使用，比如包括美国化学会(American Chemical Society)在内的一些期刊禁止作者使用 ChatGPT 进行写作。但是这显然没有从根本上解决问题。对于科技发展，我们不能因噎废食，从长远来看，应用新的技术是大势所趋。不论我们对于ChatGPT的禁止是否现实可行——因为对该款软件使用的甄别非常困难，AI 技术的不断完善必然还将产生新的同类软件，而且功能会更加强大。而且，禁止使用反而可能导致研究者故意隐瞒软件使用情况，为科研诚信保护带来更多的挑战。要保证科研诚信，我们就必须将 ChatGPT 的使用置于阳光之下。如贾婷等学者所言：“透明可释的技术过程是预测智能行为的支点，促进‘负责任’研发行为，并使‘可问责’成为可能。”[15](657)任何使用 ChatGPT 进行研究的人员都必须注明使用信息，包括应用 ChatGPT 的研究目的、研究内容、研究方法。科学研究是人类共同的事业，而且在科研过程中很可能触及其他群体利益。因此，科研共同体和利益相关者有权了解研究目的与研究动态。透明公开原则意味着研究者需要把ChatGPT应用中所收集的自己无法保证真实的信息、来源不明的信息以及可能的伦理风险告知科研合作者、成果发表机构和利益相关主体。当然，该原则也许会受到来自隐私权和知识产权的质疑。但是，ChatGPT 作为一种公共的信息资源，对其使用并不会侵犯研究者的隐私权和知识产权，对其使用信息的公开并不会伤害研究者的权益。

其次，公平正义原则。科学研究诚信始终以善为前提，而公正是“一切德性的总括”[16](130)。公平正义原则意味着对ChatGPT的使用必须服务于道德目的，有助于人类福祉的实现。任何有损于人类生命、健康、安全、利益的技术必须严格禁止。在 ChatGPT 使用中履行作为研究者的法定义务和职业操守，是公正原则的根本要义。在该类型软件的使用中，研究者不仅要考虑自身的利益，还必须顾及他人和社会的利益。如亚里士多德所言，“因为具有公正德性的人不仅能够对自身运用其德性，而且还能对邻人运用其德性”[16](130)。

此外，公平正义原则要求我们创造公平的科研环境。ChatGPT 存在扩大研究者群体差距的可能性。我们所生活的世界广泛存在着技术资源分配所产生的不平等。基于网络技术、人工智能技术的 ChatGPT 则可能加剧这种不平等。人工智能领域发达的地区可以借助该软件提高科研效率，这些地区也通常是科技发展优势地区。在 ChatGPT 的科研应用中，研发地区掌握着软件的话语权。他们决定这款软件向谁开放，以及如何授予权限。因此，类似软件究竟会打破学术垄断，还是会固化学术垄断，事实上存在很大疑问。有学者指出，虽然 ChatGPT 擅长用英语进行回应，并且熟悉美国文化背景，但对其他语言和文化的反馈则存在一定障碍[13](4)。显然，英语世界的研究者在 ChatGPT 使用中更加得心应手，而其他语言地区的研究者则面对更大的困难。由于 ChatGPT 以人类的知识体系作为其储备，在应用中表现出明显的公共性特点。在某种意义上，ChatGPT 可以被视为一种公共的学术研究平台。

就此而言，它必须满足作为公共性资源的两个正义条件：一是机会向所有人开放，二是有助于缩小不同研究群体在资源分配中形成的差别。机会向所有人开放对于 ChatGPT 应用而言，除了无差别地开放登录入口，还要充分考虑信息的多样性以及为所有人提供便利的操作。除了语言、文化的差异，还有对自然生理差异的弥补。即便它似乎向所有研究者都开放了使用通道，但如果有的研究者因为生理缺陷或者生理能力限制无法使用，实质上 ChatGPT 向他们关闭了大门。因此，ChatGPT 需要针对残障群体和行为不便群体（比如老年群体）进行操作设计，使这部分研究群体也能平等获得使用 ChatGPT 的机会。对于第二个条件，我们要充分考虑不同地区研究者的境遇。ChatGPT 目前并不是免费软件，而是提供了多种收费模式。经济发达地区的研究者可能在承担费用时不会承受太大压力，但经济落后地区的研究者则面临困境。ChatGPT 在费用收取中应该考虑经济落后地区的研究群体，采取差异化的方式降低经济 成本。

最后，共同责任原则。所有研究者都是学术共同体的成员，在研究中必须承担共同责任。在对 ChatGPT 的应用中，研究者必须尊重共同体成员权利，尊重共同体成员的成果和做出的贡献。其一，研究者有责任澄明、保证 ChatGPT 应用内容的真实性。以 ChatGPT 为代表的生成性人工智能使我们产生了主体性担忧，即人工智能是否会成为研究者之外的另一个主体。如果从人工智能的功能以及发展趋势而言，我们对其主体地位的辨别将越来越困难，这也使得对此问题的探讨充满争议。但如果从责任主体的视角分析，我们则可以清晰地断定 ChatGPT 并不具有研究主体身份。之所以众多期刊都反对将 ChatGPT 作为共同作者，并不是否认 ChatGPT 以及其他人工智能产品在研究中做出的贡献，而是这些产品不可能对研究成果担负任何责任。可以肯定，研究者是科研诚信的唯一主体。所以，无论研究者使用何种技术，都必须保证研究数据、信息的客观性、可靠性。对于 ChatGPT 所提供的信息，研究者负有真实性检验的义务。其二，研究者要在ChatGPT应用中自觉维护包括知识产权在内的其他研究者权利。对 ChatGPT 提供的理论、方法和结论，研究者必须标明出处，防止 ChatGPT 复制、编辑、篡改其他学者的成果。ChatGPT 可能会提供潜在的研究合作者，在 ChatGPT 生成作用下，也许研究成果的框架和思路来源于某一位同行，这位同行就将成为研究实际的合作者。研究者在溯源信息的同时要尊重其他学者的贡献，在成果中完整表现所参照、借鉴理论、方法在研究中发挥的作用。其三，研究者要担负研究的社会责任。ChatGPT 无法准确判断研究的风险及产生的社会影响。研究者在应用 ChatGPT 及类似人工智能软件的过程中要准确把握研究的价值取向，充分考虑研究可能带来的收益与风险，保证收益的最大化、同时确保可能收益必须远大于风险，在风险不可避免的情况下将风险降至最小。

三、ChatGPT 应用的伦理治理路径

促使研究者在ChatGPT应用中秉承科研诚信精神，遵循相应的伦理原则，是防止 ChatGPT 滥用的根本保障。对 ChatGPT 应用，我们必须建立行之有效的伦理治理路径。

首先，强化研究者的道德自觉。伦理治理与其他治理的关键区别在于，伦理治理注重主体内在德性的培养，关注外在规范的道德内化。要从根本上确保 ChatGPT 应用合乎社会的伦理期待，就必须提升软件使用者的道德水平，形成自觉的道德意识。Haven 等学者指出，科研公开、确保科研数据的可塑性和可靠性、科研透明、选择合理的科研方式提高研究质量和可信度等科研诚信的核心要素最终都掌握在科研主体手中，促使研究者自觉遵循科研诚信就显得尤为重要[17](302−303)。内在道德修养与规范制度建设对于科研诚信而言犹如车之两轮、鸟之双翼，缺一不可。与后者相比，研究主体的道德培育更有利于及时规避可能出现的问题，满足伦理治理的个体性需要。构筑研究者的道德堤坝，是一种预防性的主动治理机制。就研究者而言，需要不断强化自身的科研诚信责任观念，涵养科研诚信品格。研究者实际扮演着三种伦理角色：一是作为普遍意义上的道德个体，应该在道德理性的指引

下追求善的价值，塑造个人美德；二是作为科研工作从业者，应该遵守职业伦理要求，恪守职业道德；三是带有公共性特点的社会成员，应该关切公共利益，发挥正面道德示范作用。就研究机构和科研管理部门而言，应该定期组织研究人员进行伦理培训，详细讲解 ChatGPT 使用过程中可能出现的科研诚信问题，明确科研诚信的道德原则，强化研究者的道德认知。同时，相关部门还可以利用情境虚拟等方式增强研究者在使用 ChatGPT 过程中对于科研诚信的道德情感与道德意志。就研究团队而言，形成健康的伦理文化也是提升研究者道德意识的关键环节。人们都希望获得所在团队的认同，因此，团队文化深刻影响成员的行为模式。如 Shimizu 所言，人们对于所在团队声誉的关切对于行为有着决定性作用，为了获得团队的认同，人们甚至愿意承担个体利益的代价[18](429)。研究者违反科研诚信利用 ChatGPT 的根本动机在于试图获取不正当的额外利益，而且容易在团队内部引发道德破窗效应。在研究团队中形成科研诚信的风尚，对以非正当方式使用 ChatGPT 采取排斥的道德态度，将有助于提升团队成员的道德自律。

其次，建立 ChatGPT 伦理审查机制。伦理审查是保护研究相关者权利、维护研究规范性、道德性的重要机制，围绕 ChatGPT 使用建立伦理审查机制将为该软件辅助科研提供及时有效的诚信监管。当前，我国对于科技伦理审查的呼声日益高涨，国家层面对于科技伦理予以高度重视。2019 年，中央全面深化改革委员会第九次会议审议通过《国家科技伦理委员会组建方案》，提出“要加强统筹规范和指导协调，推动构建覆盖全面、导向明确、规范有序、协调一致的科技伦理治理体系”[19](51)。由于 ChatGPT 等人工智能辅助手段的特殊性，各级科研部门及科研管理部门都应针对类似软件使用建立伦理审查委员会，编写伦理审查委员会章程，设计伦理审查流程，推动伦理审查的制度化、规范化、常态化运转。伦理委员会主体包括科研管理人员、研究人员、人工智能领域的专家以及公众，形成多元协同的审查网络，使 ChatGPT 使用得到全方位的监督。研究者在利用人工智能辅助研究时，应该撰写研究报告，向伦理审查委员会说明使用人工智能软件的路径设计与目的，以及科研诚信的保障措施。伦理委员会对于辅助科研项目和成果具有否决权。相应地，各级科研部门和科研管理部门应该组织实施伦理审查委员的培训，设置相关系统课程，建立持证上岗制度，通过培训的人员才能获得伦理审查资格。伦理审查机制的建设一是要保证中立性与客观性，避免由于利益纠葛影响审查的公正性。二是要统一审查标准，防止不同机构、不同级别伦理审查质量出现明显差异，特别是要注意伦理审查规则与其他规范性文件的适配性。李建军等学者在分析科技伦理审查存在问题时就发现，美国“机构伦理审查委员会和联邦法规应用不一致的证据加剧了对该审查机制的不满”[19](56)。三是要凸显 ChatGPT 等人工智能软件应用的特点，准确对接科研诚信的伦理要求。作为专项伦理审查，人工智能辅助科研伦理审查委员会的运作必须反映 ChatGPT 的技术特征，而且既要恪守科研诚信的底线，又不能在审查过程中扩大对象范畴，影响科研工作的正常开展、限制科研自由。

最后，构建ChatGPT科研诚信伦理控制体系。ChatGPT 科研诚信伦理控制体系主要涉及两个方面。一是对 ChatGPT 软件的伦理控制。ChatGPT 的强大之处在于它独特的算法与知识生成方式。作为一款软件，ChatGPT 本身更多体现工具性价值，如何填补工具潜在的科研诚信漏洞，在其技术结构中体现科研诚信的伦理规范，是实现 ChatGPT 道德化的必然要求。技术的道德化包含两条路径，“一种是隐性调解设计，尽量避免不合理的调解作用，避免技术的不道德化；另一种是显性调解设计，直接设计具有特定调解作用的技术物，使某种具体的道德规范成为技术功能的一部分，实现技术的道德化”[20](88)。此外，软件的运行模式会直接影响用户的使用行为，人机交互过程也是使用者的行为塑造过程。假设 ChatGPT 的运行要求用户以遵守科研诚信的方式操作，必将提高 ChatGPT 使用的可信度。所以，需要对 ChatGPT 的运行程式进行伦理修正，将科研诚信要求转化为 ChatGPT 的结构、运行要素输入软件之中，使之成为软件程序的有机组成部

分。二是对研究者使用行为的伦理控制。确立标准化的行为规范是诚信行为控制的内在诉求。研究机构和科研管理机构应该撰写关于 ChatGPT 等人工智能软件使用的科研诚信规范文件，制定标准化的规范体系，为人工智能辅助科研提供明晰的伦理指导。由于科研的差异性，不同研究机构可以根据主要研究方向特点编制使用 ChatGPT 等人工智能软件辅助研究的伦理手册，规范 ChatGPT 应用过程。同时，科研机构与科研管理机构要强化ChatGPT应用过程中的科研诚信问责制度，以负面惩戒的方式提高科研失信成本。在人工智能技术飞速发展的今天，ChatGPT 等人工智能软件不断更新换代，这就要求我们对用户进行实时动态的行为引导和约束。以往的行为约束更多来自外部的单向监管，被监管者处于被动的地位。此种形式显然在拥有广泛灵活度的 ChatGPT 使用中难以发挥以往的效力。有学者针对 ChatGPT 等新兴产业管理提出了“敏捷管理”的模式。如薛澜等学者所言，该模式的优势在于“实现多目标间平衡”“追求过程动态优化”“促进工具灵活转化”。在该模式中，“监管方与被监管方的双向动态互动是保障监管思路与创新相匹配的重要方式”[21](33)。引入类似的管理模式，让 ChatGPT 使用者参与到规范体系建设之中，形成双向甚至多向互动的态势，将提高 ChatGPT 用户的主体性，有利于提升科研诚信行为的伦理控制效果。

参考文献：

[1] YATOO M A, HABIB F. ChatGPT, a friend or a foe? [J]. Materials Research Society, 2023(48): 310–313.

[2] PAN C D, BANERJEE J S, DEBASHIS De, et al. ChatGPT: A open AI platform for society 5.0[M]// Intelligent human centered computing. Singapore: Springer Nature Singapore Pte Ltd, 2023: 384−397.

[3] Montreal statement on research integrity in cross-boundary research collaborations[EB/OL]. (2013–05–08) [2023–06–17]. https://www.wcrif.org/downloads/main-website/montreal-statement/123-montreal-statement-english/file.

[4] Amsterdam agenda [EB/OL]. (2017−08−09) [2023−06−17]. https://www.wcrif.org/guidance/amsterdam-agenda.

[5] Hongkong principle[EB/OL]. (2019−06−21) [2023−06−17]. https://www.wcrif.org/hong-kong-principles.

[6] Cape-town statement on research integrity [EB/OL]. (2022−05−21) [2023–06–17]. https://www.wcrif.org/guidance/cape-town-statement.

[7] Singapore statement on research integrity [EB/OL]. (2010–07–21) [2023–06–17]. https://www.wcrif.org/downloads/main-website/Singapore-statements/223-singpore-statement-a4size/file.

[8] DEEPAK P. ChatGPT is not OK! That’s not (just) because it lies[J/OL]. AI & Soc (2023−04−25) [2023−06−19]. https://doi.org/10.1007/s00146-023-01660-x.

[9] GORDIJN B, HAVE H T. ChatGPT: Evolution or revolution?[J]. Medicine, health care and philosophy, 2023(26): 1−2.

[10] 吴静，邓玉龙. 生成式人工智能前景下的公共性反思[J]. 南京社会科学, 2023(7): 19–36.

[11] 姜宇辉. 作为谎言机器的 ChatGPT[J]. 贵州大学学报(社会科学版), 2023(4): 38–47.

[12] STROWEL A. ChatGPT and generative AI tools: Theft of intellectual labor?[J]. IIC, 2023(54): 491–494.

[13] ZHOU J, KE P, QIU X P, et al. ChatGPT: Potential, prospects, and limitations[J/OL]. Frontiers of information technology & Electronic engineering eng (2023). https://doi.org/10.1631/FITEE.2300089, pp1–6.

[14] YADAVA O P. ChatGPT: A foe or an ally?[J]. Indian journal of thoracic and cardiovascular surgery (May—June) 2023, 39(3): 217–221.

[15] 贾婷, 陈强, 沈天添. 基于 LDA 模型的人工智能伦理准则体系研究[J]. 同济大学学报(自然科学版), 2023(5): 652–659.

[16] 亚里士多德. 尼各马可伦理学[M]. 廖申白，译注. 北京：商务印书馆, 2013: 130.

[17] HAVEN T, GOPALAKRISHNA G, TIJDINK J, et al. Promoting trust in research and researchers: How open science and research integrity are intertwined[J]. BMC Research Notes, 2022(15): 302–306.

[18] SHIMIZU H. Social cohesion and self-sacrificing behavior[J]. Public Choice, 2011: 427–440.

[19] 李建军，王添. 科研机构伦理审查机制设置的历史动因及现实运行中的问题[J]. 自然辩证法研究，2022(3): 51–57.

[20] 王钰，程海东. 人工智能技术伦理治理内在路径解析[J]. 自然辩证法研究, 2019(8): 87–93.

[21] 薛澜，赵静. 走向敏捷治理：新兴产业发展与监管模式探究[J]. 中国行政管理, 2019(8): 28−34.

(下转第 198 页)