促进公众理解生成式人工智能的集体责任分析

——基于集体行动的视角

李健民

（山东理工大学马克思主义学院，淄博 255000）

[摘　要] 生成式人工智能的快速迭代和应用存在的种种风险使“人工智能威胁论”在社会公众视野中盛行。本文基于集体行动的视角，将人工智能应用问题解释为集体责任问题，分析公众在生成式人工智能应用中所具有的能动性和反身性的集体责任定位，以及公众在生成式人工智能治理中的集体义务。本文以结构性非正义理解生成式人工智能应用的问题，阐述促进公众理解生成式人工智能的科学传播困境，从信任角度进行出路分析。生成式人工智能应用问题的集体本质决定了促进公众理解生成式人工智能的集体意义，生成式人工智能应用的责任伦理亟待从个人伦理走向集体伦理。

[关键词] 公众　生成式人工智能　集体责任　集体行动

[中图分类号] N4；B82-057；TP18　[文献标识码] A　[DOI] 10.19293/j.cnki.1673-8357.2024.02.006

生成式人工智能有着无限前景与风险，其应用问题可以从两个不同角度被解释为集体责任（collective responsibility）问题：生成式人工智能应用本质上是一种集体问题，或集体实体（collective entity）承担生成式人工智能应用的集体责任。在后一种解释中，使用生成式人工智能服务的社会公众是作为集体能动者（collective agent）的众多集体实体之一。同时，促进公众理解生成式人工智能是人工智能风险治理的重要内容之一。本文从集体行动视角切入，以公众在生成式人工智能应用中的责任关系为对象，分析公众的集体责任定位和公众在人工智能风险治理中的集体义务，就促进公众理解生成式人工智能所面临的问题给予解决进路。

1 生成式人工智能应用问题的集体本质

加速发展的生成式人工智能重塑了人类与技术的交互模式，将全球社会中的更多人群纳入人工智能应用中，大模型、大数据、大算力支撑的生成式人工智能技术呈现“涌现”（emergence）特性，带来一系列风险挑

收稿日期：2024-03-25
基金项目：山东省社科规划研究青年项目“社会科学哲学视域下社会制度的构造研究”（23DZXJ02）。
作者简介：李健民，山东理工大学马克思主义学院讲师，研究方向：集体意向性、社会本体论等，E-mail：logos1992@sdut.edu.cn。

战。同时，生成式人工智能的应用与发展进程中涉及不同群体的大量能动者（agent），造成风险的因素在时间和空间上是分散的，人工智能风险治理不能仅靠单个能动者。生成式人工智能风险本质上是集体造成的“涌现”，它已成为一个真正的社会性、全球性问题。

1.1 生成式人工智能应用问题是一种全球性问题

生成式人工智能应用问题已经成为除气候变化、公共卫生、战争等全球性问题以外人类社会面临的又一个共同问题。作为新一代人工智能技术，生成式人工智能基于大模型的内容生成，是推动全球数字生产力变革的重要技术力量。相比以往的人工智能技术，生成式人工智能呈现更强的内容创新、人机协作、数据依赖等特性，在众多应用领域产生深刻影响，也带来一系列风险挑战，生成式人工智能正在改变人类世界。

与传统人工智能技术不同，生成式人工智能可以在已有数据基础上生成新内容，这令人工智能变得更加智能化，呈现通用人工智能的特征，生成式人工智能的此特性引发人们对人类社会未来是否会真正进入通用人工智能时代这一问题的思考。同时，生成式人工智能的强人机协作会导致人机界限模糊等问题，如何分清虚拟与现实也将成为全人类面临的共同难题。另外，生成式人工智能的大模型训练需要以互联网为依托收集海量数据，这使得生成式人工智能应用问题真正成为全球性问题，人类社会正面临前所未有的数据信息安全风险挑战。OpenAI 公司为训练其自然语言处理模型 GPT–4 使用了庞大的数据集，其数据收集以在全球互联网上抓取公开网页内容为主，其中涉及众多社交媒体、论坛、书籍、论文等原始文本数据。

此外，生成式人工智能的多领域应用导致的伦理风险、产业风险、价值风险等挑战，正逐渐影响人类社会的发展进程。笔者认为，生成式人工智能应用本质上是人类群体、社会集体共同面临的问题，理解这一全球性问题的集体本质是生成式人工智能风险治理的重要理论前提与决策依据，具有重要的理论和现实意义。2023 年，我国发布的《生成式人工智能服务管理暂行办法》明确鼓励对生成式人工智能技术平等互利开展国际交流与合作，参与生成式人工智能相关国际规则制定 [1]。

1.2 生成式人工智能应用问题是一种集体行动和集体伤害问题

集体行动问题（collective action problem）指“由独立的个体行动聚合而产生的结果，而这些行动并非旨在产生该结果”[2]。全球各国政府、科学界、企业、公众等能动者共同行动，旨在推动生成式人工智能技术发展，引领新一轮科技革命与产业变革，但生成式人工智能可能带来的社会、伦理乃至安全风险并非预先存在于相关能动者的行动意图中。生成式人工智能的社会应用是一种集体行动问题，其自身的复杂性“涌现”机制令其结果难以预测和控制。

在集体行动问题中，个体理性往往导致集体非理性。生成式人工智能的开发者、使用者、管理者等集体能动者依照其自身领域的理性规范采取行动，而能动者间缺乏合作性质的交互行动，更可能导致非理性结果。如在教育领域中，运用生成式人工智能为学生教育提供个性化支持是开发者和使用者的共同初衷，但对生成式人工智能的过度依赖将损害人类认知能力，这将成为教育领域的潜在风险 [3]。另外，能动者活动的规范性准则可能存在非理性成分，这将导致行动整体走向非理性方向，如生成式人工智能的技术开发者对底层算法有较大决策权，存在算法歧视风险。

作为集体行动问题的子概念，生成式人工智能应用问题还是一种集体伤害问题

（collective harm problem）。集体伤害问题涉及以集体造成伤害或未能防止伤害的方式行动，但相关的个人行为本身似乎没有区别[4]。面向公众提供服务的生成式人工智能产品拓展了公共生活空间，不仅使人机关系的互动与协作更加紧密，还将连接整合、协同利用更多领域的资源。在生成式人工智能塑造的更具互动性、更紧密关联的社会世界中，存在更多以间接方式伤害社会成员的风险，这些伤害通常是“多手”（many hands）[5]造成的。许多人以各种方式参与到大数据的信息收集中，特别是在使用生成式人工智能产品时，若包括个人信息等在内的输入内容被用来迭代训练，个人将面临数据与隐私安全风险。这类因相关个体行为本身的无差别性所导致的集体伤害，在生成式人工智能变革社会认知层面上尤为突出：生成式人工智能一方面可以增强个体认知，另一方面人们过度依赖人工智能可能会对社会认知产生结构性的负面影响[6]。事实上，该问题的本质是，当以某种方式行动似乎对结果没有影响或影响较小时，很难辨别我们如何有理由行动或选择不以此方式行动，普通民众显然更多关注生成式人工智能应用所带来的眼前收益，进而采取相关行动。这表明在生成式人工智能风险治理方面，开发者、使用者、管理者等单方面将降低风险视为自身的职责或义务都是错误的，应该寻求一种合作方案来解决这类集体行动问题。

1.3 生成式人工智能风险的集体责任定位

人工智能应用的责任归属是人工智能伦理的核心问题，一旦人工智能在应用进程中出现问题，能够找到责任承担者十分重要。以往的哲学分析通常基于单子化个体主义视角讨论责任的归属与分配，包括对人工智能自身能否成为道德能动者（moral agent）的讨论。然而这种分析思路忽略了人工智能应用行动的集体本质。同其他技术行动一样，人工智能行动涉及时间线上的众多能动者。人工智能的使用者、管理者可能并不知道谁曾参与技术的开发和使用，这使得风险责任的明确归属与分配变得困难。

生成式人工智能风险的责任类型涉及三个层面：因果责任、道德责任和法律责任。其中道德责任和法律责任都是规范性责任，这意味着存在作出此类规范性判断的主体，虽然这类判断不需要因果责任，但却受到因果考量的影响。因果关系存在于发生的事情与引起它发生的事情之间，原因并非必须是一个能动者，例如可以是一种自然或意外现象。然而，即便雷雨天气可能会导致电力系统损坏，但并不能追究其造成损坏的道德责任。一般而言，道德责任意味着道德能动性（moral agency），只有道德能动者才能对电力系统老化背后的经济、政治或其他相关因素负责，这些因素加大了损坏程度。就人工智能应用而言，尽管新一代生成式人工智能技术已具有更强的自主性，但笔者认为，关于人工智能风险的责任伦理讨论仍应以人类自身为焦点，至于将诸如“人机混合能动者”作为道德能动者的混合进路尝试也依然无法脱离人自身。生成式人工智能的强智能自动化会产生难以预测的有害后果，但人类却可以控制哪些任务被自动化，因此可能结果的因果责任和道德责任仍在于我们人类个人和集体。

因果责任在生成式人工智能风险责任中难以判定，呈现集体特性，同时也带来集体道德责任。以生成式人工智能应用可能导致的知识产权侵权责任为例，生成式人工智能从迭代数据到内容输出，再到发生权利侵害，整个侵权过程的因果关系追溯困难[7]。在生成式人工智能应用过程中，开发者、使用者甚至人工智能自身各自是否应承担伤害的因

果责任或应承担哪些因果责任难以确认。这种复杂的因果关系同时也导致相关法律责任难以被追究，处于行动链条的相关行动者甚至难以确定自己是否参与了行动。卢恰诺·弗洛里迪（Luciano Floridi）认为，能动者间的道德中立或道德无关交互活动是一种分布式道德行动（distributed moral action），这类道德行动的道德责任也分布于各成员间，形成分布式道德责任（distributed moral responsibility）[8]。生成式人工智能应用行动的道德责任并不能单独指向某一个体，生成式人工智能风险是一种集体伤害问题，这种伤害应分散给各类相关能动者。

道德责任不仅限于对过往行为的指责，也是一种规范性责任。由于造成生成式人工智能风险的道德能动者并不唯一，因此需要集体责任概念，这一方面包括造成伤害的因果责任，另一方面也涉及造成伤害后所受的指责，集体责任概念将道德责任的根源定位在这些群体所采取的集体行动中。科技伦理治理将政府、科技人员、企业及利益相关方、社会公众等作为参与人工智能行动的主要集体实体，涉及应对人工智能风险应该做什么及由谁做的问题，这些集体实体承担人工智能风险的集体责任，在有关生成式人工智能的责任讨论上涉及哲学、法学、管理学等多个学科领域。这种规范性集体道德责任将各类能动者在结构上定位于不同的集体实体范围内，避免分布式道德责任面临的责任主体缺失问题[9]。

大多数哲学家在讨论集体责任时主要围绕回溯性集体责任（backward looking collective responsibility）展开，即在伤害发生后追溯道德能动者的责任，近年来他们对所谓的前瞻性集体责任或前瞻性集体道德责任（forward looking collective moral responsibility）的关注度逐渐增加，前瞻性集体责任主要关注能动者应该为弥补伤害或预防伤害发生做什么。生成式人工智能风险的集体责任很大程度上是在讨论一种前瞻性集体责任，毕竟这类技术的应用在人类社会中刚刚起步，还并未造成重大伤害。对于人工智能技术的开发者和管理者而言，这显然是一种集体职责（collective duty），同时开发者有义务让人工智能技术走向为人类服务的善的方向。对于使用者而言，存在一种更为积极的集体义务（collective obligation），不仅包括对未来科技发展的信任与信心，同时也涉及主动提高自身科学素质、参与并支持人工智能风险治理的集体行动，这都表明人工智能应用本质上是一种集体问题。

2 公众在生成式人工智能应用中的集体责任定位

联合国教育科学及文化组织发布的《人工智能伦理问题建议书》（*Recommendation on the Ethics of Artificial Intelligence*）强调了公众参与在人工智能伦理治理中的重要作用[10]。分析公众在人工智能应用行动中责任定位的前提是理解公众如何作为集体来承担责任、其能动性的基本内涵和公众参与人工智能风险治理的途径及集体意义。

2.1 公众的能动性和反身性

责任离不开能动性，有能动性的实体往往是有组织的实体，如政府中相关工作人员的规范性角色嵌入组织结构中，使政府能作为集体能动者行动。克里斯蒂安·李斯特（Christian List）和菲利普·佩蒂特（Philip Pettit）认为能动者是一种系统，有描述其周围事物如何的表征状态、确定事物应该如何的动机状态，以及处理这两种状态以在两者不匹配时便于干预的能力[11]。另外，一定程度的自主性和理性也是成为能动者的必要条件。在人工智能应用及风险治理行动中，将公众理解为具有能动性的集体实体具有重要

意义，他们成为承担人工智能应用责任的重要一环，将使用者纳入责任归属的讨论中，作为人工智能的使用者，公众既是能动者，同时也是受动者，呈现一种反身性特征。

作为集体实体，公众的反身性特征首先体现在其参与生成式人工智能应用的集体行动中。公众群体中的个体成员不应仅指大部分人工智能普通用户，还应包括参与人工智能技术发展的开发者及管理者，因为即便是开发者和管理者也并非完全能够掌握关于人工智能的所有知识，这些人彼此之间呈现一种非独立的关系，因而一旦出现伤害，也都应承担相应的因果责任和道德责任。

另外，公众也是人工智能应用责任关系中的“责任受动者”（responsibility patients），受能动者行动的影响，同时要求能动者能负责任地行动，即能动者被期待并被要求给出其行动理由[12]。公众在人工智能应用中兼具能动者（并非唯一能动者）和受动者身份，一方面，公众参与生成式人工智能的应用行动；另一方面，生成式人工智能应用导致的可能风险会影响公众。值得注意的是，公众受到的风险危害并非总是完全负面意义的，如生成式人工智能中的数据爬虫问题反倒可能引发公众信息隐私意识的觉醒。

2.2 义务承担者：公众在生成式人工智能行动中的集体意义

自乌尔里希·贝克（Ulrich Beck）首次提出“风险社会”概念以来，伴随各个领域内技术的极速发展，“风险社会”一词已被大众所熟知并接受，“相比于其他个别因素，技术可能性的巨大拓展对吸引公众关注风险的贡献巨大”[13]。生成式人工智能技术一经诞生就引发社会公众的关注，这很大程度上是由公众对该技术是什么及其未来可能的发展方向不够明确所致。笔者认为，公众作为承担生成式人工智能应用集体责任的集体实体，有主动理解生成式人工智能、提升自身科学素质的集体义务，同时在参与人工智能风险治理行动中，也有促成集体实体采取集体行动的义务。成为承担责任的道德能动者的前提是知道人们在做什么或已经做了什么，这是风险社会中公众价值的重要体现。

生成式人工智能技术未来将成为推动社会各领域智能化升级的关键动力，将更加深度地融入社会公众的日常生活中，促进个性化服务、智能制造、虚拟现实等多个领域的进展，这项技术的最终目的是服务于人类的社会生活。阿尔文·托夫勒（Alvin Toffler）曾指出，快速的技术变化将超出正常人的接受能力，必须学会理解和控制变化速度，成为技术进化的主人[14]。社会公众使用生成式人工智能技术时，无论是直接使用抑或被动参与，都应具备一定的相关知识，这不仅由于面对生成式人工智能技术的发展，公众对其理解产生分歧继而导致“人工智能威胁论”盛行，更深层次的原因是公众本质上既是人工智能应用的能动者，也是人工智能风险的责任承担者和受动者。笔者认为，公众理解生成式人工智能的基本内涵是作为集体的公众对什么是生成式人工智能、生成式人工智能的应用与发展前景、生成式人工智能应用的风险具有一定的认知，认同管理者主导的人工智能风险治理的基本策略。

在生成式人工智能行动中，公众责任的集体性还体现在促成集体实体采取相关行动上。生成式人工智能技术导致风险伤害的一般原因之一是缺乏相关的责任规范。以生成式人工智能对深度伪造技术领域的影响为例，更加进步的技术模型有能力生成与现实难以区分的视频、图像和音频，越来越多相关的人工智能危害社会现象引发公共舆论、造成社会损失。一方面，政府及监管部门有职责出台相关法律法规；另一方面，由于技

术迭代与应用周期短，较大可能出现规范滞后情形，因此公众有集体义务促成集体实体采取行动以规避风险。第一，公众有集体义务促成公众自身包含的社会成员合作，从群体走向有组织的集体。弗吉尼亚·赫尔德（Virginia Held）提出，有理性的人清楚需要采取什么行动并且当该行动的预期结果明显有利时，一群人的集合可能因未能合作形成有组织的群体以防止伤害而承担责任[15]。更加有组织的、对生成式人工智能应用达成更多共识的社会公众将成为人工智能风险治理的重要力量。第二，公众有集体义务努力促成政府和其他集体能动者在还未存在有效集体能动性的情景中采取集体行动。公众既直接处于人工智能应用场景中，也是风险伤害的亲历者，应更为积极地向政府、开发者等能动者反馈伤害，以促进技术和规范责任体系的完善，为多元参与协同互动的“敏捷治理”贡献力量。

3 促进公众理解生成式人工智能

作为生成式人工智能应用的集体能动者，社会公众一方面有集体义务积极参与人工智能风险治理，包括主动加强对生成式人工智能知识的认知；另一方面，由于生成式人工智能应用以非正义的社会结构为背景，社会公众缺乏对生成式人工智能的理解和认知，其根源是信任问题。

3.1 生成式人工智能的结构性非正义问题

生成式人工智能应用问题不仅是单个能动者或集体行动的问题，还是一种社会结构问题。导致生成式人工智能风险的因素涉及社会整体的多个层面，在复杂社会网络中，个人和集体都与权力和利益产生关系。艾丽斯·杨（Iris Young）认为，当社会进程使大批人受到统治或剥夺其发展和行使能力途径的系统性威胁，同时这些进程使其他人能够统治或拥有发展和行使能力的广泛机会时，就会存在结构性非正义（structural injustice）[16]。社会中的所有人都因对社会进程作出因果贡献而负责。

生成式人工智能的主要结构性非正义在于，人工智能技术应用的成本和收益在全社会及全球不同群体中分布不均，使用生成式人工智能的成本与门槛较高。一些社会群体更容易受到人工智能风险的影响，不同群体对人工智能技术的认知与使用程度将影响其适应能力。在全球企业纷纷布局生成式人工智能的境况下，受益于生成式人工智能的工作者将提高生产力和收入水平，而无法受益的工作者将面临落后及失业风险。OpenAI 公司曾就生成式人工智能的语言大模型对美国劳动力市场的影响予以研究，得出的结论是约 80% 的美国人在其工作任务中会受到大模型的影响[17]。因此，可大胆预测，这种结构性非正义未来将进一步加剧全球贫富差距，一些从未使用或参与人工智能技术行动的群体可能受到最严重的影响，尤其是社会结构中的一些边缘性群体可能遭受人工智能技术影响下社会结构所产生的大部分风险伤害。另外，生成式人工智能还可能加剧社会结构中原本存在的不正义内容，如算法歧视可能进一步加剧社会性别偏见、种族刻板印象，模型训练消耗大量能源加剧全球碳不平等。

改善生成式人工智能的结构性非正义有赖于集体行动的形成。人工智能应用问题本身是一种集体行动问题，应对该问题的关键在于形成更具有集体性的行动，即集体能动者的行动，或形成有共同目标的合作性联合行动。政府、企业等集体能动者的行动及合作行动将在改善生成式人工智能的结构性非正义问题上扮演重要角色，同时社会公众中的个体为实现规避风险的共同目标联合行动，不仅能形成集体行动，还令社会公众成

为人工智能应用行动中的重要集体能动者。罗宾·郑（Robin Zheng）认为，个人通过其社会角色对结构性非正义负责，角色是结构和能动性的交汇处，理解这些角色有助于确定个人为什么要负责、对什么负责及应承担什么义务[18]。在人工智能应用中，个人可能是开发者、管理者、使用者等不同社会角色，抑或存在角色重叠的情形，明晰不同角色的责任定位将有助于改善结构性非正义问题，包括理解社会公众在生成式人工智能风险治理中的角色责任定位，并在此基础上促进公众理解生成式人工智能，确保生成式人工智能的开放包容与公平普惠，推动人类社会能最大限度地共享生成式人工智能带来的益处。

3.2 促进公众理解生成式人工智能的信任问题与出路

导致生成式人工智能的结构性非正义问题的重要原因之一是社会公众自身缺乏相关知识，这既是一个科学传播和教育问题，也涉及公众对科技的态度。笔者认为，促进公众理解生成式人工智能并不能简单停留于科普层面，更重要的是培养公众的认知主动性。负责任地行事不仅要求作为能动者的公众知道自己在做什么，同时公众也要对人工智能技术的未来发展持有自己的理性态度。

促进公众理解科学、提升公众科学素质一直以来都被视为科学传播领域的核心主题，笔者认为，促进公众理解生成式人工智能面临的主要科学传播困境是生成式人工智能的开发者、使用者和管理者间缺乏知识性互动，未能将个体的科学素质以一种集体行动方式表达。作为新一代人工智能技术，生成式人工智能的基础性知识还未在公众中普及，公众对人工智能的理解大多停留在语音识别等弱人工智能应用层面，人工智能科普教育亟待推进。以往对社会层面科学素质的讨论仅关注个体的聚合，未能有效检视社会结构，如不同社会群体科学素质的差异分布情况对科学素质水平的影响等。事实上，提升社会层面科学素质的关键不是使各类群体的科学素质达到统一水平，普通民众很难且不必精通人工智能的所有相关知识，而是不同科学素质水平的群体能协同行动，特别是增强开发者和使用者间的知识性互动，以超越个体科学素质之总和的方式展现国家或全社会的科学素质水平[19]。另外，生成式人工智能的监管与治理需要积极鼓励并推进开发者和使用者的参与跟互动，一方面向公众普及使用生成式人工智能服务的规范，另一方面鼓励科学家参与人工智能科普，特别是要积极解决人工智能科普资源的分配不均衡问题，避免导致或加剧未来生成式人工智能应用的结构性非正义问题。

促进公众理解生成式人工智能，增强开发者和使用者间的知识性互动，其中科学家能否与公众良好互动是重要内容，也是促进科技发展的关键。然而，面对科技问题，社会人群倾向于产生分歧而非在很大程度上达成共识，人工智能与人类关系就是典型例子。人工智能威胁论盛行的通常解释是公众的相关知识不足，如果能够弥补不足，比如将人工智能素质教育纳入基础教育至高等教育的各个环节，公众就更可能与科学界保持一致。迪特里姆·舍费尔（Dietram Scheufele）认为这类进路并未触及问题的根源，“根本问题不在于知识缺失，而在于信任缺失。民众只有更信任科学家，才能更信任科学研究”[20]。刘永谋指出：“要维护技术专家与大众之间的信任，警惕此种信任的衰落甚至消失。”[21]信任是维持科学传播向好态势的关键因素。同时，促进公众理解生成式人工智能亟待在社会公众与发展负责任人工智能间建立信任关系，让公众相信人工智能技术的良性发展图景。然而，信任负责任人工智能的发展并不等于盲目相信开发者和管理者的发展及应用

策略取向，公众作为集体能动者和风险受动者，就新技术的迭代和应用应当展开更为广泛的讨论，积极参与人工智能风险治理行动。

此外，由于社会公众的价值观或文化背景存在差异，建立公众对发展负责任人工智能的信任关系可能面临公众难以就“负责任人工智能应该是什么”达成共识这一情形。OpenAI公司发布的人工智能文生视频大模型Sora具有强大的图像视频生成能力，改变了人们“眼见为实”的传统观念，但不同价值观下的公众成员却对此产品持有不同态度，存在追求更多可能性和维护真实世界间的价值观冲突。因此，促进公众理解生成式人工智能的重要内容之一应该是促成社会公众就生成式人工智能的发展和应用达成较为一致的价值观，特别是在生成式人工智能全球大发展的背景下，促成此种价值观的一致性将更具有集体意义。2023年世界互联网大会发布《发展负责任的生成式人工智能研究报告及共识文件》，提出发展负责任的生成式人工智能共识，强调增进人类福祉，坚持以人为本，推动人类经济、社会和生态可持续发展的价值宗旨[22]。

促成公众形成较为一致的价值观，以保证公众对发展负责任人工智能的信任，并非完全不允许不同的价值观存在。闫宏秀认为，一种更为审慎的信任关系应当是在解析价值观差异的基础上，寻找不信任的价值观基础，继而构建关于负责任人工智能的共识[23]。笔者认为，不同价值观的存在有利于保证公众对人工智能持有理性态度，但需要建构信任负责任人工智能的基本价值观，即生成式人工智能的发展和应用应当为社会带来更多的益处。公众在确立这种基本信任关系的前提下，将更加主动地加强对生成式人工智能的认知学习，提升自身科学素质，不仅在有关人工智能的科学知识方面得到提升，还能借助对人工智能的学习，掌握基本科学方法，培养科学思维，形成崇尚科学的精神和科技向善的价值观，继而作为集体能动者积极承担参与人工智能风险治理的集体义务。

4 结论

传统个人伦理框架难以较好地呈现生成式人工智能应用问题中的集体内容，生成式人工智能应用问题的集体本质意味着该问题必须通过集体行动来解决，集体责任无法在完整意义上转化为个体责任。我国发布的《新一代人工智能治理原则——发展负责任的人工智能》将研发者、使用者和受用者“共担责任”作为人工智能治理原则之一[24]。其中，既是使用者也是受用者的社会公众不仅要承担人工智能应用的集体责任，还要积极承担人工智能风险治理的集体义务。生成式人工智能应用问题还是一种结构性非正义问题，促进公众理解生成式人工智能并在此基础上形成集体行动有利于改善这种情形。然而，促进公众理解生成式人工智能面临科学传播困境，解决该问题有赖于建立公众对负责任人工智能发展的信任关系。

整体来看，生成式人工智能涉及的伦理治理问题亟待从个人伦理走向集体伦理，一种生成式人工智能应用的前瞻性集体道德责任框架亟待建构。作为集体能动者和受动者的社会公众在生成式人工智能发展和应用进程中应当发挥重要作用，在主动提升自身科学素质的同时积极参与人工智能风险治理行动。正如刘大椿所言：“科技发展所带来的四海一家的情势，则促使人们进一步发展一种具有大同世界胸襟的新型集体伦理……新型集体伦理将更加强调人类普遍共识基础上的共同行动，只有这样，才可能实现整体的永续发展。”[25]

参考文献

[1] 中国网信网 . 生成式人工智能服务管理暂行办法 [EB/OL].（2023–07–13）[2024–01–07]. https://www.cac.gov.cn/2023–07/13/c_1690898327029107.htm.

[2] Schwenkenbecher A. Structural Injustice and Massively Shared Obligations[J]. Journal of Applied Philosophy，2021，38(1)：3.

[3] 李艳燕，郑娅峰 . 生成式人工智能的教育应用 [J]. 人民论坛，2013(23)：70.

[4] Nefsky J. Collective Harm and the Inefficacy Problem[J]. Philosophy Compass 14 (4)：e12587.

[5] Thompson D F. Moral Responsibility of Public Officials：The Problem of Many Hands[J]. The American Political Science Review，1980，74(4)：905–916.

[6] 段伟文 . 准确研判生成式人工智能的社会伦理风险 [J]. 中国党政干部论坛，2023(4)：76–77.

[7] 袁曾 . 生成式人工智能的责任能力研究 [J]. 东方法学，2023(3)：18–33.

[8] Floridi L. Distributed Morality in an Information Society [J]. Science and Engineering Ethics，2013，19 (3)：727–743.

[9] 闫宏秀 . 数据时代的道德责任解析 : 从信任到结构 [J]. 探索与争鸣，2022(4)：37–46.

[10] 联合国教科文组织 . 人工智能伦理问题建议书 [EB/OL].（2021–11–24）[2023–12–08]. https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi.

[11] List C，Pettit P. Group Agency：The Possibility，Design，and Status of Corporate Agents[M]. Oxford：Oxford University，2011：20.

[12] Coeckelbergh M. Artificial Intelligence，Responsibility Attribution，and a Relational Justification of Explainability[J]. Science and Engineering Ethics，2020，26 (4)：2061.

[13] 尼可拉斯・卢曼 . 风险社会学 [M]. 孙一洲，译 . 南宁：广西人民出版社，2020：127.

[14] 阿尔文・托夫勒 . 未来的冲击 [M]. 蔡伸章，译 . 北京：中信出版社，2018.

[15] Held V. Can a Random Collection of Individuals be Morally Responsible? [J]. The Journal of Philosophy，1970，67(14)：476.

[16] Young I. Responsibility for Justice[M]. New York：Oxford University Press，2011：52.

[17] Eloundou T，Manning S，Mishkin P，et al. GPTs are GPTs：An Early Look at the Labor Market Impact Potential of Large Language Models[J]. arXiv Preprint arxiv：2303.10130，2023.

[18] Zheng R. What is My Role in Changing the System? A New Model of Responsibility for Structural Injustice[J]. Ethical Theory and Moral Practice，2018(21)：869–885.

[19] 凯瑟琳・E. 斯诺，肯妮・A. 迪布纳 . 科学素养 : 概念、情境与影响 [M]. 裴新宁，郑太妍，译 . 北京：中国科学技术出版社，2020：74–85.

[20] 迪特里姆・舍费尔，游文娟 . 公众如何理解科学 [J]. 科学与社会，2023(5)：56.

[21] 刘永谋 . 专家与大众 : 人们为何对专家不满 ? [J]. 科学传播，2023(3)：24.

[22] 世界互联网大会人工智能工作组 . 发展负责任的生成式人工智能研究报告及共识文件 [EB/OL].（2023–11–09）[2024–01–07]. https://cn.wicinternet.org/2023–11/09/content_36952741.htm.

[23] 闫宏秀 . 负责任人工智能的信任模塑：从理念到实践 [J]. 云南社会科学 . 2023(4)：45–46.

[24] 国家新一代人工智能治理专业委员会 . 新一代人工智能治理原则——发展负责任的人工智能 [EB/OL].（2019–06–17）[2023–12–08]. https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html.

[25] 刘大椿 . 科技伦理建构的新路径 [N/OL]. 光明日报，2023–06–05(15) [2023–12–17]. https://epaper.gmw.cn/gmrb/html/2023–06/05/nbs.D110000gmrb_15.htm.

（编辑 颜 燕 荆祎澜）