生成式人工智能时代的使用者伦理研究

闫宏秀　杨映瑜

（上海交通大学科学史与科学文化研究院，上海 200240）

[摘　要] 在既有生成式人工智能伦理治理讨论中，使用者层面的伦理长久以来遭到忽视。交互性与生成性作为生成式人工智能的重要属性为使用者打开了一条呈现自我的有效路径，生成式人工智能时代中使用者与开发者的融合早已超过生产者与消费者在经济活动中的融合，走向了人类社会形态的重构与人类未来的重塑。因此，使用者伦理的出场与前置应被视为数字素养的未来发展方向。本文将个人、社会两个层面与积极伦理、消极伦理两个维度相结合所构建的使用者伦理矩阵，为使用者在面对、应用生成式人工智能时所应遵守的行为道德准则和规范提供了基本框架。为了更好地实施这一框架，需要技术使用者、技术设计（开发）者、技术监管者三方共同协力，推动人与技术的关系走向紧密与向善，为人类的福祉与未来保驾护航。

[关键词] 生成式人工智能　使用者伦理　技术道德化　数字素养

[中图分类号] B82-057；TP18；N4　　[文献标识码] A　　[DOI] 10.19293/j.cnki.1673-8357.2024.02.003

近年来，以 GPT-4、Sora 等为代表的生成式人工智能作为一种科技创新，在推动新质生产力发展的同时，其隐含的不确定性后果和潜在伦理风险也越来越被社会所关注。交互性与生成性作为生成式人工智能的重要属性为使用者打开了一条呈现自我的有效路径，并与设计者共同推进技术的发展。特别是随着生成式人工智能技术的易用性不断提升，普通使用者的地位也日益凸显，并成为推动技术革新和应用普及的关键力量。他们不仅是技术的受益者，更是技术的塑造者和传播者。然而，面对日新月异的生成式人工智能，使用者自身应当建立并遵守何种伦理规则以应对技术的挑战却鲜有讨论。

根据《生成式人工智能服务管理暂行办法》，生成式人工智能技术指的是“具有文本、图片、音频、视频等内容生成能力的模型及相关技术”[1]。目前，在生成式人工智能伦理治理的讨论中，企业和国家层面的伦理议题受到了广泛关注。这些研究多集中于企业内部的伦理审查机制建设[2]，包括人工智能领域的立法以及政策性文件、行业技术标准、行业协会自律公约、行业协会发布的倡议书等[3]。对比之下，目前学界对于使用者层面的

收稿日期：2024-03-02
项目基金：上海市哲学社会科学规划项目“数智时代的价值对齐研究”（2023BZX001）。
作者简介：闫宏秀，上海交通大学科学史与科学文化研究院教授，研究方向：技术哲学、技术伦理学，E-mail：hxyan@sjtu.edu.cn。

伦理讨论却相对较少，这反映出生成式人工智能时代背景下的使用者伦理在一定程度上并未受到社会的关注与探讨。纵观人工智能发展史，使用者在生成式人工智能时代的重要性已然不可同以往等量齐观，这既是令人欣慰的技术路径改变，同时也带来了使用者治理的难题。

正如习近平总书记所指出："要加强人工智能同保障和改善民生的结合，推动人工智能在人们日常工作、学习、生活中的深度运用，创造更加智能的工作方式和生活方式。"[4] 随着未来生成式人工智能的普及，使用者作为技术的最终目的与指向，亟须重新审视自身，推动使用者伦理的出场与构建，以期与技术发展相协同，共同守护人类的福祉与未来。

1 使用者伦理出场的技术图谱：Web3.0背景下的"产消者"

1.1 "产消者"的新样态

针对科学技术的发展，阿尔文·托夫勒（Alvin Toffler）从经济活动的视角提出，"第一次浪潮社会的产消合一者在以高新科技为基础的第三次浪潮时代又将重新成为经济活动的中心"[5]。产消者是指那些参与生产活动的消费者，他们既是消费者又是生产者。产消者不以交换目的而进行生产，而是出于自己的兴趣、爱好、需求而无偿地生产产品或者服务。同时随着消费者对生产过程的介入越多，其在生产活动中的必要性也就越显著，从Web1.0时代到Web3.0时代的演变历程（见表1）来看，也恰好印证了这一判断。

表1 Web1.0时代、Web2.0时代、Web3.0时代对比表

分期	Web1.0时代	Web2.0时代	Web3.0时代
时间	1990—2004年	2004—2022年	2022年以后
内容创作方式	PGC：专业生成内容	UGC：使用者生成内容	AIGC：人工智能生成内容
技术支撑工具	HTML、HTTP、URI	JavaScript、AJAX	AI、ML、CNN、RNN、NLP
使用者交互性	使用者互动性有限的静态网站	使用者互动性增强的动态网站	智能且个性化的使用者体验

Web1.0时代始于20世纪90年代，使用者主要通过互联网网页（浏览器）来获取信息，但彼时的网页由HTML代码编写，网页内容是固定的，不会因为使用者的交互而进行动态变化。同时，网页内容是由网站开发人员以及专业的内容创作者提供的，即专业生成内容（Professional Generated Content，PGC），普通使用者在这一阶段只能作为读者阅读信息，而无法创造内容。在这一阶段，生产者和使用者（消费者）是壁垒分明的，且相对生产者而言，使用者的话语权和地位都较为低微。长此以往，Web1.0时代在"作者少、读者多"的情况下，仍然采取中心化的数据存储方式，导致使用者所能享受到的资源十分匮乏，因此Web1.0时代很快被Web2.0时代所取代。

2004—2022年大致处于Web2.0时代，使用者不再仅仅是读者，他们可以通过各类动态网站参与在线编写和内容修改而成为创作者，即使用者生成内容（User-Generated Content，UGC）。在这一时期，UGC的创作途径为人们通过各类社交平台分享自己创作的内容。通过去中心化的数据共享和协作，不同的组织或个人之间进行共享和协作使群体的内容创造力不断提升。不过在这一阶段，UGC的创作主力仍然是拥有专门技能的意见领袖，普通使用者的潜力仍待发掘。

2022年之后进入Web3.0时代，通过运用生成式人工智能，使用者可以与机器协作，共同生成内容，即人工智能生成内容（AI-Generated Content，AIGC）。不同于前两个阶段，生成式人工智能的颠覆性在于引发了生产方式的变革，而非仅仅局限于托夫勒所指的商品活动过程。

1.2 生成式人工智能视角的“产消者”

在生成式人工智能中，使用者与设计者的融合早已超出生产者与消费者在经济活动中的融合，走向了人类社会形态的重构与人类未来的重塑。

一方面，生成式人工智能开始替代人类非物质形式的劳动。自从第二次工业革命引发的工业化、机械化劳动大规模地取代了人类的体力和物质劳动以后，生成式人工智能开始冲击人类对于脑力和非物质劳动的“独占”地位。目前，AIGC 已在各类内容创作领域布局，如文本生成有 ChatGPT，图像生成有 Stable Diffusion、Midjourney，视频生成有 Sora 等。通过分布式的数据存储和分享，以及在先进算法和强大算力的加持下，未来人类的生产力和创造力或将难以离开生成式人工智能的支撑。与此同时，这也引发了人类对于自身地位的怀疑，或许未来从事非物质劳动的主体不再仅仅是人，机器也不再是被动的工具，人机共生已是不可阻挡的趋势。

另一方面，在人机共生的进程中，使用者已然成为中枢。不同于 Web1.0 和 Web2.0 时代，AIGC 背景下的所有使用者从广义上而言已参与到生产行为中。“整个世界，我们每一个人，都在为人工智能做贡献。每当我们上网，使用我们的智能手机、信用卡或社交媒体，通过自动收费站，播放流媒体电影、看电视，都是在为人工智能做贡献。”[6] 目前，生成式人工智能的训练数据来自互联网的文本、音视频和聊天记录等，同时使用者在使用生成式人工智能所产生的数据和信息又“反哺”为训练数据，帮助大模型更新迭代。在这一过程中，用户的消费行为兼具了生产性质。在托夫勒那里，“产消者”仍需经过一道商品交换，且消费者并没有真正参与到劳动工具的制造中，但在 Web3.0 时代，使用者不仅直接参与到了劳动工具的制造中，他们的消费资料还源源不断转换为生产资料，进而推动生产力的发展。因此，生成式人工智能的消费者超越了托夫勒所处的时代，并引发了本体论意义上的变革。

综上，生成式人工智能推动了劳动形式、生产关系的转变，打破了传统的生产者和消费者的平衡，逐渐呈现出了使用者成为 AIGC 时代重要力量的趋势。从 PGC 时代的开发者（生产者），到 UGC 时代的部分使用者（拥有专门技能的意见领袖），最后到 AIGC 时代的全体使用者（产消者），这既是人机共生所带来的生产力解放，也是潜在的人类生存危机。

1.3 技术发展亟须使用者伦理的出场

从 Web1.0 到 Web3.0 时代，使用者的数量在不断增加，而这种数量上的无限扩大也将导致秩序的混乱。得益于人工智能使用成本的降低，早期 Web1.0 只能搭载在电脑终端上，而该时期的电脑售价昂贵，只有部分技术精英才能够承担，而到 Web3.0 时代，生成式人工智能已经可以在手机上运行，并且无须使用者掌握编程语言，只需自然语言即可使用。

根据第 53 次《中国互联网络发展状况统计报告》，截至 2023 年 12 月，我国手机网民规模达 10.92 亿人[7]。这也从另一方面启示我们，面对数量庞大且素质良莠不齐的使用者群体，建立并普及使用者伦理应当是必须且急迫的社会任务。

2 技术伦理的新面向：使用者伦理作为“将技术道德化”的补充

将技术道德化（Moralizing Technology），即将道德理念嵌入到技术设计中，从而使技术物能引导和规范人的行为[8]。在生成式人工智能的设计中，道德物化已有所体现。例如，近年来所提出的“为社会负责任的人工智能”（Socially Responsible Artificial Intelligence）

“为社会负责任的人工智能算法”（Socially Responsible Artificial Intelligence Algorithms，SARs）等均明确将社会责任与人工智能的发展置于一个共同的框架之中[9]。此外，各国所推出的相关倡议与框架，如中国发布的《全球人工智能治理倡议》，美国发布的《负责任的人工智能战略和实施路径》等，也都提及在算法设计、技术开发、产品研发与应用过程中，应避免歧视、偏见等问题。虽然在设计中嵌入道德是可行的，但仍不足以应对严峻的人工智能治理困境，“将技术道德化”在某些情境下仍然会有“失灵”的潜在风险。

2.1 技术调节的反调节与再调节

技术具有多稳态性，技术设计者与使用者二者之间的意向性并非是连贯的或一致的，甚至可能是对抗性的。例如，在设计ChatGPT、Bing Chat等聊天机器人时，设计者出于安全性能的考量会拒绝为使用者生成诸如贩毒、偷窃等有害的教学内容，但使用者通过各类“越狱”技巧即可绕过ChatGPT的安全屏障而获得答案，这类行为属于技术的反调节。就目前的情况而言，在社交媒体、论坛，一些含有不良倾向的“越狱”技巧、经验仍在被用户分享和转载，且少有监管，因此，这种反调节的伦理风险需要引起高度重视。

与此同时，技术调节的再调节，即不能预先假定技术设计者都是诚实而完美的，设计者的设计行为必须得到再调节。如广州地铁女子裸照事件，一名网红在地铁拍摄的正常照片被恶意“一键脱衣”成“裸照”，虚假照片的迅速流传严重损害了当事人的名誉权和肖像权。但与技术设计者初衷相背离的是，这类软件设计之初是为了帮助女性在网络上虚拟试衣或智能挑衣[10]。由此可见，反调节与再调节都超出了技术设计者的可控范围，二者均与使用者的意图、行为息息相关。

2.2 “将技术道德化”中的使用者问题

就将技术道德化而言，从《将技术道德化》一书的副标题“理解与设计物的道德”就可以看出，设计者被赋予了高度关注，但这并非仅仅指通过设计将道德予以技术化，进而将技术的道德问题置于设计端，而是力图通过对物的道德解码来开启对技术道德构成问题的探讨。就这种构成而言，设计者与使用者的意向性在将技术道德化的理论中均被关注。然而，鉴于设计者在技术环节中的重要性，技术设计者的伦理被高度重视。特别是，随着技术日趋自主与智能化的趋势，技术设计者的伦理素养、价值取向等逐渐被视为技术伦理问题的核心。

就技术设计者的现状来看，目前主流的生成式人工智能多由欧美发达国家的白人男性工程师开发，这些技术设计者的价值偏好也被映射到了生成式人工智能内容生成的各个环节。例如，ChatGPT所用的语料就主要来自维基百科（Wikipedia）、Common Crawl、Web Text等语料库，这些语料库语种类型多为英语，这也许会导致非英语国家的使用者在使用生成式人工智能时不免受到来自西方价值观念和文化思想的冲击，同样非白人群体、女性群体也因训练数据的局限性而备受人工智能的偏见。因此，事实证明技术设计者很难保证完全的价值中立，如果仅以技术设计者的“法则”来发展技术，使用者被剥夺参与技术民主决策的权利，那么技术就会成为技术专家专制的“利维坦”。大至国家安全、意识形态，小至弱势群体的利益，都应由多方群体协商参与，而非仅由技术设计者独揽决策大权。

一旦任由技术设计者成为类似的“统治阶级”，技术治理可能异化为技术统治，而避免这种命运就需要落实再治理的制度设计[11]。易言之，采取将技术道德化的方式虽能在一

定程度上用技术引导使用者做出道德行为，但技术调节仍然会面临诸如反调节与再调节、技术统治论等问题，这些问题超出了技术设计者的客观能力范围，且与使用者层面息息相关。究其原因，技术本质上仍然有社会属性，它指向的另一端是使用者，因此，作为实际伦理受益者和风险受害者的使用者问题也亟待解决。

2.3 “将技术道德化”与使用者伦理的出场

使用者伦理的出场作为“将技术道德化”的补充，意味着更多元的技术伦理，重视使用者与技术的关系，能推动多方利益群体的协同，为将技术道德化的内在主义进路进行补充，进而推动生成式人工智能更好地走向善与发展之路。

生成式人工智能作为一种尚处于研发升级阶段的新兴技术，与之相关的伦理分析往往面临“科林格里奇困境”问题，即过早且过严的控制容易导致技术难以获得爆发与突破；过晚控制则易导致技术发展失控，再治理起来将耗费更多的时间、金钱、人力。为此，有学者指出，为破解这一难题，AIGC 的治理应建立“从决策层、技术层到应用层，从标准规范、管理制度到工具支撑，自上而下贯穿 AIGC 发展底层数据治理的完整生态体系。在 AIGC 运行的各个阶段以及基于数据合规治理框架规制的各个环节之间需要对规制目标和宗旨取得共识，以确保合规高效、科学动态地完成治理目标”[12]。而就现状来看，将技术道德化已从决策层、技术层引导使用者如何应然地使用生成式人工智能，但具体到使用者如何实然地使用层面则进一步呼唤使用者伦理的出场。

实证研究表明，公正、责任、隐私与透明度是使用者在实践中难以察觉的伦理问题，并且不同身份的使用者存在个体利益和社会影响感知上的明显差异[13]。这意味着，使用者视角的伦理层次存在着被遮蔽的现象，应在使用者的实然层面有针对性地建构相应的、可操作的伦理框架。而就将技术道德化而言，其自身作为一种前瞻性的技术伦理分析，仍然会面临对技术未来的猜测同真实使用行为不符的情况。易言之，这种方法忽视了生活经验中的基本规范性问题。为弥补这一缺憾，以汤姆·波尔森（Tom Brsen）为代表的技术人类学派提出构筑以专家、使用者、人工物三方互动的人类学模型，“技术哲学更多的是从工程师、专家的角度思考问题，而奥尔堡学派的技术人类学方式则更倾向于对用户、顾客等技术的使用者进行研究”[14]。就理论方法而言，技术人类学强调，“微观层面个体与实体之间的具体互动；制度层面应有成文的规范和具体化的实践；社会层面强调支撑微观层面和制度层面的意识形态和政治进程”[15]。因此，使用者伦理作为将技术道德化的补充，将从一种技术人类学的视角为使用者提供“实然”的伦理框架。

3 基于数字素养新发展视角的使用者伦理：从唯技能论到全民伦理启蒙

随着人工智能的浪潮逐渐蔓延至更多人群，传统的数字素养（Digital Literacy）概念已经不足以应对生成式人工智能时代的挑战。在此基础之上，拓宽数字素养的内涵与边界，使数字素养能为全体使用者所共享是确保全社会享有数字福祉、规避数字风险的必由之路。

3.1 技能培养视角下的数字素养

理查德·兰纳姆（Richard Lanham）于 1995 年提出数字素养的概念，他认为“素养”在数字时代的涵义应当从理解文本拓展到图像、声音、视频等多媒体信息形态[16]。

在 20 世纪与 21 世纪之交，数字素养主要指使用者对数字技术的学习能力与技术技

能。这种观念的形成是因为自 20 世纪 70 年代计算机应用程序诞生以来，早期使用者必须通过专门、系统的学习才能使用与特定任务或工作相关的计算机系统。因此，数字素养一词的出现正是用来评估与计算机相关的基本概念和技能的。随着计算机技能在人们当代社会生活及工作中的重要性不断增强，数字素养已逐渐进入幼儿教育、中小学教育领域。在 K–12 基础教育阶段，数字素养的教育更偏向于学习人工智能的知识和技能，如计算机科学基础、数据挖掘、机器编程等，从而为学生未来的工作需求打下基础。

值得注意的是，以色列学者约拉姆·埃谢特 – 阿尔卡莱（Yoram Eshet–Alkalai）在 2004 年依据功能将数字素养划分为图像素养（视觉识读）、创新素养（创造性生产）、分支素养（新思维）、信息素养（辨别真伪）及“社会—情感”素养（共享与情感交流）五个层次[17]。

3.2 对唯技能论的反思

无疑，上述这种以掌握具体技能操作为目标的数字素养，过度强调实用性和现实性的功能主义导向或将难以应对生成式人工智能对人类社会造成的具有颠覆性质的潜在伦理风险。当下，这种唯技能论的数字素养教育思潮的弊端也逐渐显现出来。实证研究发现，学生很少关注伦理问题，如人工智能和法律责任的偏见（8%）、人工智能和知识产权问题（9%），因为这对于他们的就业并无直接益处[18]。因此，数字素养的教育不仅应该关注提高学生的人工智能能力和兴趣，还应该帮助学生认识到作为使用者应如何去正确应对人工智能带来的社会影响和伦理问题。

基于此，学界对这类定义进行了反思和批判，“数字素养”的外延也逐渐扩大。数字素养教育的核心内容应该集中在数字技术素养、针对动态文本的思辨能力、创造性生产与自我表达能力、互动交往能力以及数字媒介伦理[19]。同样，在各国及各大国际组织的官方文件中，伦理也作为重要因素被纳入数字素养框架。例如，《欧洲终身学习核心素养框架》（*European Commission：Key Competences for Lifelong Learning*）明确要求使用者自身需要以批判性的态度看待从数字渠道所获信息及数据的有效性、可靠性和影响，并且了解有关数字技术使用的法律和伦理原则；同时要以符合伦理的、安全的、负责任的方式来使用数字工具[20]。我国的《提升全民数字素养与技能行动纲要》则将数字素养明确定义为，数字社会公民学习工作生活应具备的数字获取、制作、使用、评价、交互、分享、创新、安全保障、伦理道德等一系列素质与能力的集合，并提出推动全民数字道德伦理水平大幅提升的发展目标[21]。

3.3 数字素养的新意涵：全民伦理启蒙

纵观近几十年数字素养概念与战略的发展，其自身已突破唯技能论的桎梏，走向了伦理层次的启蒙。因此，不能把数字素养看作简单的数字获取与使用等技术性技能，还要看到与之相关联的思维和行为模式。换言之，公众应具备的数字素养与技能，不仅指公众具备与信息通信等数字技术相关的知识和技能，还包括他们在数字社会中的价值观、伦理、行为和思维方式等方面的素养。而从使用者的角度来看，包含伦理启蒙的数字素养应当惠及全体公民，不应仅局限于 K–12 教育阶段的学生。

在使用者群体中，我们尤其要关心“银发一族”对于人工智能的接受度和学习能力。诸如其文化程度、经济能力的不足，以及网络安全隐患催生的“数字恐惧”、从现实世界到虚拟空间的过渡所引发的“数字排斥”等问题，导致部分老年群体不愿学、很难学习人工智能的相关知识[22]，这也使得老年群体

成为不法分子利用人工智能犯罪的主要目标。在技术的裹挟下，这类使用者如果不具备基础的人工智能素养，不仅无法享受到生成式人工智能所带来的福利，反而很容易被不法分子利用人工智能加害。这类情况也启示公众，在已有的数字素养框架中，总是被首先强调的那些数字技术与技能在现实生活中并不一定真正能为使用者所有，因此将使用者伦理前置、倡导伦理先行或许是更适宜数字素养未来发展的新方向。

4 生成式人工智能时代的使用者伦理框架

4.1 使用者伦理的涵义及其矩阵框架

广而言之，使用者伦理即是在生成式人工智能时代，使用者在面对、应用这项技术所应该遵守的行为道德准则与规范的总称。使用者伦理取向可分为积极伦理与消极伦理两个维度。所谓积极伦理，是指使用者以积极的态度来释放技术给人带来的正面价值，如促进人类解放与人类的全面自由发展等[23]。消极伦理，则是指强调使用者的道德义务，节制、禁止某些行为。同样，使用者伦理还有个人伦理层与社会伦理层两种层面。将它们与积极伦理以及消极伦理两个维度相结合，则可得到一个基本的使用者伦理矩阵框架（见表 2）。

通过该矩阵框架，可初步梳理出使用者伦理的详细内涵并用其指导使用者正确面对、使用生成式人工智能。

表 2 使用者伦理矩阵框架

维度 / 层面	积极伦理维度	消极伦理维度
个人伦理层面	（1）认识到生成式人工智能的存在。了解人工智能在各个领域的应用，包括警惕、识别、判断某产品是否使用了人工智能技术，如智能电话机器人，AI 合成的图片、音频、视频等 （2）学习、利用人工智能的原理和前沿技术、优势和劣势。在此学习基础之上努力探索人工智能的运用之道，包括改善个人学习和生活的新方式 （3）批判地看待生成式人工智能。在了解生成式人工智能的原理之上，警惕人工智能幻觉；不独依赖生成式人工智能，必要时能主动人工核查事实的准确性，保持自我的认知、批判、创造能力	（1）不以个人隐私换取某些便利。许多生成式人工智能项目正处于“盲目圈地”阶段，如果使用者不加选择地全盘接受，则有可能导致自身的隐私变为模型养料。例如，一些应用采取模糊处理的个人隐私条款；一些用户投喂自己的隐私照片、聊天记录给模型等 （2）克制“越狱”式的使用行为。使用者需要加强自我的道德修养，节制恶意利用漏洞攻击系统的欲望
社会伦理层面	（1）遵守特殊公共领域的人工智能使用规则。在学术、文化产业等领域，对于生成式人工智能的使用另有专门的规定，这也是维护社会公序良俗、正常商业秩序的必然要求 （2）积极反馈、参与人工智能的相关决策。由于输入数据集的不完整或者人工智能开发者、算法本身有偏见，人工智能的输出内容不可避免地也会产生诸如性别、年龄、种族方面的偏见和歧视。虽然人工智能可以给出“带有‘偏见’的方案和建议，却无法直接对他者实施价值层面上的‘歧视’”[24]。因此，使用者利用生成式人工智能产出关涉价值选择的内容时，应当批判性地剔除包含歧视、不公的部分，并积极反馈，以帮助人工智能修正这类错误，进而能对社会整体发挥积极的作用	（1）不利用人工智能生成损害他人合法利益的内容。《中华人民共和国民法典》规定民事主体享有人身权利和财产权利。人身权利方面，公民的名誉权、肖像权、隐私权等人格权是“受灾领域”重区；财产权利方面，虽然生成式人工智能的独创性和可版权性目前已处于探索阶段，但人工智能的普及和应用同样也大大增加了知识产权的侵权风险。在上述权利领域，使用者需要更加谨慎地使用，避免侵权受到法律惩罚 （2）不轻信、传播深度伪造的信息，避免假消息、假新闻扰乱舆论秩序与社会格局 （3）避免给人工智能模型投喂低俗、猎奇的信息。作为产消者，使用者的使用行为同样也是生产的一环，大量垃圾信息的投喂只会恶化人工智能的整体生态环境，因此使用者的此类行为应当规避

4.2 使用者伦理框架的实施与践行

使用者伦理框架虽直接关涉使用者，但其实施与践行仍然依赖多方共同努力。

首先，就技术使用者自身而言，提高自我的道德意识，对于生成式人工智能的批判

认知、使用能力已是伦理的应有之义。进一步而言，如前文所述，产消者所带来的本体论意义上的变革也意味着使用者应当更积极地介入到“生产”环节中。在国内首个大语言模型治理开源中文数据集 100 PoisonMpts 中，环境社会学专家范叶超、著名社会学家李银河、心理学家李松蔚、人权法专家刘小楠等专家作为标注工程师，各提出 100 个包含诱导偏见、歧视回答的刁钻问题以帮助人工智能模型完成从“投毒”到“解毒”的攻防[25]。这类专家型使用者在标注数据的过程中，相较于纯技术架构开发者，能更好地识别出隐蔽的偏见歧视、潜在有害内容。由此可见，一方面生成式人工智能技术的设计者应当更多地吸纳不同学科、领域的专家型使用者，通过博采众长，技术设计者可以更好地理解不同行业的需求和挑战，从而更有针对性地优化技术，使其更好地服务于各类群体。另一方面，使用者自身也可借鉴消费者协会、微博社区志愿者等制度建设自律组织，推动构建共治、共享的智能空间。

其次，尽管技术设计者会面临技术的反调节与再调节问题，但就目前来看，将伦理道德嵌入生成式人工智能设计的程度仍然是远远不够的。布鲁诺·拉图尔（Bruno Latour）将技术物对个体行为的影响作用概括为“脚本（Script）”，但在技术实践的实际过程中，使用者并不一定必然遵循设计者的“脚本”，反而会“改写”。困难的是，在生成式人工智能时代，使用者的“改写”现象更加频繁，这反过来促使技术设计者在编写“脚本”时，应尽可能详细地考虑到使用者可能的意图，并以说明、引导的方式帮助使用者正确使用人工智能。诸如强制用户阅读用户须知、隐私保护条款，以醒目方式警告使用者不当使用的严重后果及法律责任，强制标注某段内容是人工智能合成等。当然，随着使用者的地位日渐提升，未来设计者应当加强与使用者的合作，包括但不限于吸收各学科专家型学者优化数据集、重视使用者的反馈等。

最后，作为技术伦理治理的兜底，技术的监管者也应对伦理秩序予以最基本的保护。就目前来看，我国在这方面已取得了一定的成绩。《中华人民共和国民法典》将数据、网络虚拟财产纳入法律保护范围；对肖像权、著作权、名誉权等的规定，同样适用于人工智能领域，如明令禁止“深度伪造”侵害他人肖像、声音等。《中华人民共和国民法典》《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律法规从根本上明确了相关技术服务提供者、使用者的权利和义务，为其提供了基本的行动指南，并且在《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等文件中有更加详细的规定。而根据《提升全民数字素养与技能行动纲要》对提升全民数字道德伦理水平的要求，未来我国可在全民普及生成式人工智能知识、加强伦理宣传，在对弱势群体倾斜、关注等方面发力，以提高使用者的整体数字素养水平。

4.3 作为生成式人工智能发展必要条件的使用者伦理

与发展势头如火如荼的生成式人工智能相对的是，人类社会应对人工智能风险的滞后性。直到 2024 年，欧洲议会批准的《人工智能法案》（*Artificial Intelligence Art*）才成为全球首部全面的人工智能监管法规。总体而言，相关立法和政策的推出与人工智能的发展速度总是不匹配的；人工智能领域的人才需求与供给之间存在不平衡，相关教育和培训不够充分；普通公众对人工智能技术的认知和理解程度有限，社会对人工智能伦理问题的认识和讨论需要提上日程。

以上种种都在提示公众，必须注重伦理在技术发展中的重要意义。然而，作为基数最为庞大的使用者，他们应当承担何种伦理责任却仍不明晰，这显然并不符合生成式人工智能的发展趋势。因此，从生成式人工智能的技术特性来看，使用者伦理是其发展的必要条件，但更需认清的是，建设使用者伦理是一个长期且复杂的过程，需要个人、企业、社会、政府等多方共同协力。在现阶段，面对不同水平、能力的使用者群体，使用者伦理的践行应当从个人层面到社会层面循序渐进地推进。

参考文献

[1] 中央网络安全和信息化委员会 . 生成式人工智能服务管理暂行办法 [EB/OL].（2023–07–13）[2024–03–24]. https://www.cac.gov.cn/2023–07/13/c_1690898327029107.htm.

[2] 章诚豪，张勇 . 生成式 AI 的源头治理：数据深度运用的风险隐忧与刑事规制 [J]. 湖北社会科学，2023(11)：127–135.

[3] 曾雄，梁正，张辉 . 人工智能软法治理的优化进路：由软法先行到软法与硬法协同 [EB/OL].（2024–01–23）[2024–02–22]. http://kns.cnki.net/kcms/detail/11.5181.TP.20240122.1129.014.html.

[4] 新华社 . 习近平主持中共中央政治局第九次集体学习并讲话 [EB/OL].（2018–10–13）[2024–03–24]. https://www.gov.cn/xinwen/2018–10/31/content_5336251.htm?eqid=8e9a43b100048e97000000003646b28f8.

[5] 阿尔文·托夫勒 . 第三次浪潮 [M]. 黄明坚，译 . 北京：中信出版社，2018.

[6] 帕梅拉·麦考黛克 . 人工智能往事：精英、文化与思维 [M]. 虞晶怡，杨丽凤，译 . 上海：格致出版社，2023.

[7] 中国互联网络信息中心 . 第 53 次《中国互联网络发展状况统计报告》[R/OL].（2024–03–21）[2024–03–24]. https://www.cnnic.net.cn/n4/2024/0321/c208–10962.html.

[8] 彼得·保罗·维贝克 . 将技术道德化——理解与设计物的道德 [M]. 闫宏秀，杨庆峰，译 . 上海：上海交通大学出版社，2016.

[9] 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)：40–49.

[10] 董浩 . 生成式人工智能时代人机对话的传播伦理风险及其应对 [J]. 阅江学刊，2024，16 (1)：104–114.

[11] 刘永谋，李尉博 . 有限技术治理的理论建构与时代意蕴 [J]. 自然辩证法研究，2024，40(2)：3–10.

[12] 张春春，孙瑞英 . 如何走出 AIGC 的“科林格里奇困境”：全流程动态数据合规治理 [J]. 图书情报知识，2024(2)：39–49.

[13] 洪杰文，常静宜 . 用户关切视角下的“生成式 AI”伦理框架建构 [J]. 新闻与写作，2023(12)：44–55.

[14] 王皓 . 奥尔堡学派的技术 – 人类学探究 [J]. 青海民族大学学报 (社会科学版)，2019，45(4)：57–64.

[15] 王中贝，周荣庭 . 新兴技术伦理分析路径研究 [J]. 自然辩证法研究，2022，38(3)：29–35.

[16] Lanham R A.Digital literacy[J].Scientific American，1995，273(3)：160–161.

[17] Eshet–Alkalai Y.Digital Literacy：A Conceptual Framework for Survival Skills in the Digital Era[J]. Journal of Educational Multimedia and Hypermedia，2004，13(1)：93–106.

[18] Gong X Y，Tang Y，Liu X Y，et al. K–9 Artificial Intelligence Education in Qingdao：Issues，Challenges and Suggestions[EB/OL].（2020–11–04）[2024–02–22]. https://ieeexplore.ieee.org/document/9238087.

[19] 李德刚 . 数字素养 : 新数字鸿沟背景下的媒介素养教育新走向 [J]. 思想理论教育，2012(18)：9–13.

[20] 钟周 . 胜任数字变革：欧盟数字素养框架体系研究 [J]. 世界教育信息，2023，36(1)：46–57.

[21] 中央网络安全和信息化委员会办公室 . 提升全民数字素养与技能行动纲要 [EB/OL].（2021–11–05）[2024–03–24]. http://www.cac.gov.cn/2021–11/05/c_1637708867754305.htm.

[22] 黄晨熹 . 老年数字鸿沟的现状、挑战及对策 [J]. 人民论坛，2020(29)：126–128.

[23] 成素梅 . 智能革命与个人的全面发展 [J]. 马克思主义与现实，2020(4)：196–202.

[24] 孟令宇 . 从算法偏见到算法歧视：算法歧视的责任问题探究 [J]. 东北大学学报（社会科学版），2022，24(1)：1–9.

[25] Openbayes. 100PoisonMpts：中文大模型治理数据集 [EB/OL].（2024–02–23）[2024–03–24]. https://openbayes.com/console/hyperai–tutorials/datasets/493fA4LmWwA/1/overview.

（编辑 颜　燕　和树美）

centric viewpoint. It surpasses general intelligence with its immense computing power and breaks through traditional creativity with seemingly nonsensical combinations. This shift has elevated artificial intelligence from a mere support tool for humans to a relatively autonomous creative force. In the evolving landscape of human-machine intelligence interactions and integration，individuals must embrace a continuous and dynamic dual spiral reconstruction of both technological literacy and humanistic creativity.

Keywords：creativity；generative artificial intelligence；scientific literacy；technological and humanistic literacy

CLC Numbers：N4；TP18　Document Code：A　DOI：10.19293/j.cnki.1673-8357.2024.02.002

Research on User Ethics in the Era of Generative Artificial Intelligence

Yan Hongxiu　Yang Yingyu

（Institute of Science History and Science Culture，Shanghai Jiao Tong University，Shanghai 200240）

Abstract：In the existing discussions on the ethical governance of generative artificial intelligence（AI），user-level ethics have long been overlooked. Interactivity and generativity，as key attributes of generative AI，open an effective path for users to present themselves. The fusion of users and developers in the era of generative AI has long surpassed the integration of producers and consumers in economic activities，leading to the reconstruction of human social forms and the reshaping of humanity's future. Therefore，the emergence and prioritization of user ethics should be seen as the future development direction of digital literacy. By combining individual and societal levels with the dimensions of positive and negative ethics，the constructed user ethics matrix provides a fundamental framework for the moral guidelines and norms that users should adhere to when facing and applying generative AI. To better implement this framework it requires the joint efforts of technology users，technology designers（developers），and technology regulators to promote a closer and more benevolent relationship between humans and technology，safeguarding human well-being and the future.

Keywords：generative artificial intelligence；user ethics；moralizing technology；digital literacy

CLC Numbers：B82-057；TP18；N4　Document Code：A

DOI：10.19293/j.cnki.1673-8357.2024.02.003

On Popularization of Artificial Intelligence

Yang Qingfeng[1,2]

（Institute of Science Ethics and Human Future，Fudan University，Shanghai 200433）[1]

（School of Philosophy，Fudan University，Shanghai 200433）[2]

Abstract：Artificial intelligence has become a driving force for national development and competition in the new era. Given this premise，greater emphasis must be placed on the popularization of artificial intelligence（AI）. Currently，societal perceptions of AI are constrained by traditional views，such as seeing AI merely as a tool or a daily object，which leads to a lack of comprehensive understanding of it. This Research explores the relationship between AI，the environment，and humanity，using concepts such as transparency，relevance，and augmentation. It proposes that AI popularization should be stratified and that traditional notions