Advances in Psychology 心理学进展, 2023, 13(11), 5422-5430
Published Online November 2023 in Hans. https://www.hanspub.org/journal/ap
https://doi.org/10.12677/ap.2023.1311686

Hans汉斯

基于保护动机理论的伦理风险型技术科普传播效果心理审视

徐煜晖，李雨洁

南京林业大学，人文社会科学学院，江苏 南京

收稿日期：2023年10月9日；录用日期：2023年11月14日；发布日期：2023年11月24日

摘　要

人工智能等涉及伦理风险的新兴技术正在推动社会生态发生改变，引发社会学、心理学和新闻传播学等多学科对环境、个人和行为三者的广泛讨论。与此同时，这类技术的科学传播过程正在从传统模式向大众深度参与转变，公众与技术的互动关系也将随之发生变化。文章借助人本主义心理学家代表罗杰斯基于健康信念模式提出的保护动机理论，结合伦理风险型技术的特性，从社会心理学角度深入理解伦理风险型技术科普传播过程中的公众心理逻辑，并讨论其科普效果公众接受的心理阻滞因素，在此基础上探讨文章提出的伦理－技术生态观作为提升伦理风险型技术科普传播效果的可行性，为优化新兴技术的科普实践提供有益的启示。

关键词

伦理风险型技术，科学普及，保护动机理论

Psychological Review of the Effectiveness of Ethical Risk-Based Technological Science Communication Based on Protection Motivation Theory

Yuhui Xu, Yujie Li

Faculty of Humanities and Social Sciences, Nanjing Forestry University, Nanjing Jiangsu

Received: Oct. 9th, 2023; accepted: Nov. 14th, 2023; published: Nov. 24th, 2023

Abstract

Emerging technologies involving ethical risks, such as artificial intelligence, are driving changes in

文章引用: 徐煜晖, 李雨洁(2023). 基于保护动机理论的伦理风险型技术科普传播效果心理审视. *心理学进展, 13(11),* 5422-5430. DOI: 10.12677/ap.2023.1311686

the social ecology, triggering extensive multidisciplinary discussions on the environment, individuals and behaviours in sociology, psychology and journalism and communication. At the same time, the process of scientific communication of such technologies is shifting from traditional models to deeper public participation, and the interaction between the public and the technologies will change as well. With the help of the theory of protective motivation proposed by Rogers, a representative of humanistic psychologists, based on the health belief model, and in combination with the characteristics of ethical-risk technologies, the article provides an in-depth understanding of the public's psychological logic in the process of ethical-risk technologies' science communication from the perspective of social psychology, and discusses the factors of psychological blockage in the public's acceptance of their science communication effects, based on which it explores the ethical-technological ecological view presented in the article. On this basis, we discuss the feasibility of the ethical-technological ecological view proposed in the article as a means to enhance the effectiveness of popular science communication of ethical risk-based technologies.

Keywords

Ethical Risk-Based Technologies, Science Popularization, Conservation Motive Theory

1. 背景与意义：伦理风险型技术的科普传播效果再审视

2022 年 3 月，中共中央办公厅、国务院办公厅发布《关于加强科技伦理治理的意见》，这一政策对我国进一步加强科技伦理治理、推动科技向善以及建设创新型国家和科技强国具有重大意义。随着数字生活空间中公众对信息公开的需求和社会参与意识的提高，公众正在通过社会化感知和参与新兴技术的科普解读，成为科学知识传播过程中的关键节点，从而影响和改变社会医疗、环境、文化等公共领域。

然而，我国在科技伦理治理方面的起步相对较晚，体制机制和制度体系尚不完善，这些因素共同导致其发展相对滞后。以人工智能技术为代表的新兴技术，其应用不断更迭创新，促进社会发展的同时，专业知识壁垒使其成为社会感知不对称的根源。公众在面对伦理风险型技术的传播时，由于全面理解人工智能技术的“知识需量”与自身有限的“知识存量”之间存在的差异，常常产生紧张、焦虑甚至恐惧的情绪。这种情况逐渐引发了主体责任不清、算法歧视、侵犯隐私等传统伦理价值之争，并进一步产生自主决策泛滥、责任归属分化等主体层面问题，以及社会不平等现象加剧、社会伦理反向脱敏、伦理问题多面化等实践层面症结(崔中良，卢艺，2023)。因此，如何保证伦理风险型技术的科普效果，如何引导受众触发有效行动提升伦理风险型技术的科学传播效果成为社会学、心理学和新闻传播学等多学科领域中的关键性问题。

针对以上问题，文章借助人本主义心理学家罗纳德·罗杰斯(Ronald Rogers)提出的保护动机理论(Protection Motivation Theory, PMT)，从社会心理学角度对伦理风险型技术科学传播过程中的受众心理逻辑进行分析，以期通过探讨伦理风险型技术的科学传播效果和受众接受程度，为保证伦理风险型技术的科普效果和引导受众触发有效行动提供理论支持，为优化新兴技术的科普实践提供有益的启示。

2. 理论基础：“科学传播的科学”框架与保护动机理论

2.1. 科学干预：“科学传播的科学”框架

近年来，“科学传播的科学”这一框架对国际科学传播领域的影响不断扩大，包括心理学在内的认

知科学备受学界关注(贾鹤鹏，2020)。《科学传播的科学手册》一书于2017年由牛津大学出版社出版，进一步扩展了该理论框架所带来的影响，同时传递了一个重要信息：在众多议题中，科学传播的有效性问题是“科学传播的科学”框架下最为关键的问题。从认知角度来看，科学传播的主要目的就是帮助公众真正了解并接受科学知识，但这种认知是否有效取决于受众的接受程度，即受众能否真正参与到知识建构过程之中。对于伦理风险型技术在科普传播方面的研究，鲜有以此为基础的探讨，即从心理学出发，在微观视角下考察阻滞伦理风险型技术科普传播成效的产生因素。

一方面，科学技术的科普实践导向中必然会引入伦理议题的介入。在此意义上，科学普及不再仅仅只是学界埋头苦思的问题，更是一种社会“干预”的行动(刘鹏，2023)，具有强烈的实践导向。另一方面，科技的不确定型也增加了伦理介入的不确定性。从实践层面来说，科学技术的不确定性确实导致了大量风险事件的发生。因此，承认科学技术的不确定性，并非对其加以全盘否定，相反，这种承认基于对科学不完备性的认可，进而要求科学技术在社会普及过程中保持一种面向未来认识论的开放性，同时也是伦理层面上的开放性。在此意义上，即便科学技术的不确定性难以提前测定，公众也可以对其保持一种开放心态，并对其可能带来的伦理风险的不确定性保持警惕。

2.2. 公众行动：保护动机理论

伦理风险型技术科普传播并非单纯的技术信息传递，而是一个带有社会心理性相互作用和交流的过程，其特殊性体现在科普主体是伦理参与，深层次目的在于转变大众对于这类科技的态度与行为，说服科普对象，从而引导公众做出行动。今天的科学传播，主要包括“科学共同体”内的沟通以及科学组织(包括科学家)与普通公众之间的沟通，媒体在这些交流形式中所起的作用是必不可少的(张迪等，2021)。因此，此类技术的科学普及过程更加需要结合社交媒体特性，并深入理解公众心理逻辑。Rogers等(Stainback & Rogers, 1983)心理学家以健康信念理论模型(Health Belief Model, HBM)为基础提出的“保护动机理论(Protection Motivation Theory, PMT)”为分析伦理风险型技术科普传播效果提供了理论依据和结构表达工具，干预措施也更具操作性和可行性。现有学者的研究也证实了保护动机理论在风险干预中能够进行有依据的评估，基于健康信念模式对于大众健康行为的提升具有很好的说服作用，已经被用于推广婚检行为，艾滋病(刘彩，2015)，糖尿病(杨青，2010；李莉莉等，2019)，胃癌(韦丽鹤等，2022)及青少年控烟(王芸等，2009)等科学普及工作中，能够更加全面和深刻地剖析公众行为变化的内在机制及过程，已在公共卫生风险相关研究领域得到了广泛运用。

保护动机理论根据行为产生的三种模式将其划分为：信息源，认知中介过程和行为应对模式(见图1)。理论上，感知严重性(perceived severity)、感知易遭受性(perceived vulnerability) 、反应效能(response efficacy)是影响信息采纳的重要因素，后又将自我效能(self efficacy)补充进去，前两者为风险评估，后两者为效能判断，其中风险评估包括内部奖励、外部奖励、感知严重性和感知易遭受性；效能判断包括自我效能、反应效能和反应成本(response cost)。只有通过激发人们对技术风险的四种认知，方能激发其积极应对的意愿。

这一理论强调风险评估与反应评估在预测个体意图与行为变化中的重要性。当个体认识到风险威胁非常严重，受众自身是风险易受群体，并且相信行为变化是有益的，此时行为变化的代价越小、行为变化的自信和能力越强；不良行为的内、外部回报越低，则保护的动机越大，从而推动个体采取行动或者采取抑制行为。同时风险评估与反应评估之间也具有交互作用。即当反应的有效性和自我效能感较高时，风险严重性和易感性认知的提高会积极地促进受众行为改变的意愿，否则无作用或者起反作用，这一观点在实证研究中也得到了证实(裴东超等，2021；Prentice-Dunn & Rogers，1986)。

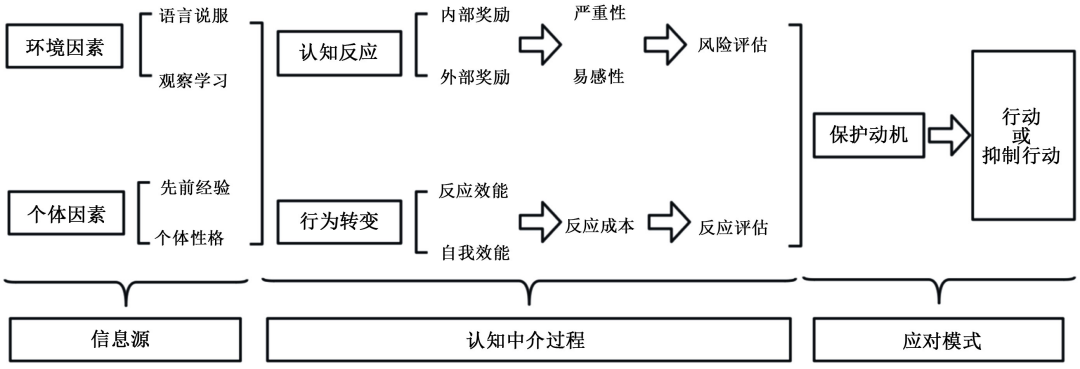

Figure 1. Structure of conservation motivation theory
图 1. 保护动机理论结构图

3. 现状与分析：伦理风险型技术传播效果的公众心理逻辑与阻滞因素

3.1. 伦理风险型技术严重性感知启动失谐

感知严重性(perceived severity)即个体认为某种事物或行为会为自身身心带来危害及其危害程度的一种判断，包括受众对技术风险引起的社会后果的知觉程度。不同受众对技术风险的严重性判断也有参差。据保护动机理论，唤起公众伦理风险型技术严重性感知是唤起伦理风险型技术科普过程中公众感知、态度和行动的起始点。在所有技术的科学普及中，技术的风险性被视为一种潜在的威胁信息，然而，它是否能够唤起公众对其严重性感知的适度认知，仍需进行更加深层次的思考。

“数据化”“智能化”和“算法决定论”等词汇的频繁出现昭示了一种以人工智能为驱动力的新型经济和社会形式的日新月著。当公众仍在疑虑、未能及时采取行动的情况下，智能算法已开始评估你的面试机会和犯罪率，以决定哪个广告和新闻应该被推荐。在海量数据的支持下，人工智能逐渐具备超越人类的预测和决策能力，例如自动驾驶可能比人类驾驶更具安全性，智能医疗或比医生具有更加精准的诊断效果，智能语音识别比速记员具有更低的错误率等等。

虽然高新技术确值得今后大力研究开发并不断投入，但凡是有望改变人类社会发展的新技术都必然会引起社会伦理争议。用户隐私、虚假信息、算法“黑盒”、网络犯罪、电子产品过度依赖等问题都引发了世界各国对于互联网技术本身及其所带来的冲击的思考与讨论，在这些论辩声中，尤以 2018 年岁末的基因编辑婴儿事件为辩争中心。“人类在很长一段时间扛着理性大旗，并将追求‘理性’视作个体和社会进步的表征，但事实上人类自身从未彻底摆脱‘感性’的制约和限制(陈相雨，2023)。”人人都有发言权的移动智能时代，个体化表达成为常态，情感和情绪相对于事实更能左右受众的风险感知以及观点行动。

美国符号论美学家苏珊·朗格(Susanne K. Langer)认为表征性符号常常依附娱乐化元素构成感官刺激，情绪化表达正是表征性符号的突出特点。对于未曾意识到涉及伦理风险的技术的公众来说，他们对于技术风险严重性感知本身就并不强烈，而部分自媒体对于伦理风险型技术的解读和普及过度娱乐化更是加剧了这种趋势。社交媒体平台自身所处的娱乐化环境也容易造成伦理风险型技术破坏严重性感知。一旦充满娱乐特征的表征性符号占据社交媒体传播较大比例，泛娱乐化环境便会由此得到形塑。当狂欢与围观成为人们推广伦理风险型技术时所面对的行为特点后，伦理风险型技术负面可能性便无限渲染，并和其他娱乐化信息串联在一起，激活并沉淀大众在伦理风险型技术面前应有的情感，真正应有的理性情绪便会受到挑战。伦理风险型技术在科普过程中形成的所谓“视觉奇观”，极易诱发大众对伦理风险型技术严重程度的过度认知，从而促使心理逃避或者防御性拒绝，而科普的过度奇观化亦屏蔽了大众对

理性文字或者图像背后含义的反思，符号的真实意义让位于视觉感知最终导致观看者丧失了对伦理风险型技术深层次的认知和理解。

3.2. 伦理风险型技术易感性感知唤起困乏

感知易遭受性(perceived vulnerability)指向公众对自身遭受某种威胁或某种消极后果可能性的评估(杨英新，2022)。人们的风险意识与其是否具有预防性行为之间存在联系。当个体认为某种技术产生的风险威胁极可能对自身造成影响时，便会在其应用和普及过程中倾注更多的关注并付诸行动。伦理风险型技术科普中技术易感性启动效果不佳造成其可预测性不强，难以向大众群体进行传播与推广是限制科普效果的主要原因之一。第一，信源的可信性问题。一些伦理风险型技术在科普过程中仅仅是对技术应用场域进行笼统的介绍，科普者的权威性以及技术和公众之间关联程度模糊性的表述难以把公众引入到可信易感性风险判断之中。互联网发展的“下半场”，以互联网和智能算法为代表的数学媒介下沉为社会“操作系统”，根本性地重构着社会(喻国明，2022)，内容与传播生态的演进将人工智能的“可知与不可知”以话语权的形式分散给普罗大众，个体由此实现自身的话语表达与资源调配。长久与人工智能相伴的人可能会错把“智能”当成“智慧”(喻国明，李钒，2023)，从而导致伊莉莎效应，个体过度解读机器意义，拯救自我的意识逐渐消失，致使自我价值的消解和狂欢之后人类对机器的屈服。

第二，说服效果错综复杂，公众存在“乐观偏差”。数字空间中说服技巧繁多，在论及涉及伦理的风险型技术时，“两面提示”较“一面提示”对受教育程度更高的受众效果更好，而普通受众摆脱不了人云亦云、随波逐流和乐观偏差的误区。很多公众感到，与自己相比，他人更易遇到更为严重的伦理风险型技术事件。最为典型的是，随着交互频次不断加深，算法作为决定用户信息议程的因素开始对用户“想什么”产生影响，对用户信息范畴，内容偏好产生影响、使用频度多少受算法的影响较大，甚至存在某种程度的驯化(domestication)现象。用户在通过标签设定、搜索、屏蔽、点击、分享等交互行为进行技术的反向驯化无果后，逐渐认清现实，并放弃抵抗，转而享受算法推荐的偏好性服务，在无损自身利益的前提下享受技术风险的“乐观偏差”。已经习惯于技术服务的用户会倾向于认为，自己对伦理风险型技术事件的认知比别人多，重大伦理风险型技术事件发生在自己身上的可能性不大，公众自我卷入度不够，无法及时评估风险感知的易遭受性，伦理风险型技术科学普及易被判定为无关紧要的信息。同时技术和数据主义大背景的双重数据化控制也使得个体在认识到感知技术风险和发展反控制意识和能力的必要性后成为“余数生命”(吴冠军，2020)的风险和自我抵抗的有限，公众逐渐发展为宁愿掩耳盗铃，默认现有的技术风险，也不愿被数据化了的技术体系拒之门外，从而丧失自身某种权益。这也使得行动与结果之间以及数据与人员之间的关系难以分辨，加剧确定伦理风险感知易遭受性唤起的难度。

3.3. 伦理风险型技术反应效能供给不足

反应效能(response efficacy)是指个体对于自己采取的某种保护性行为能否发挥作用的感知。通常情况下，公众之所以会采取某种行为，是由于认为自己会从中受益，而这种受益对于个体来说是合理的。比如戒烟可以减少孩子哮喘发作的范围和数量，但是如果抽烟的家长认为戒烟并不能减少哮喘发作时或出于社会交往的考虑不愿戒烟的话，对孩子来说戒烟并不会有太大的好处；其结果为：相对于认为戒烟对控制儿童哮喘发作有好处、认为戒烟可改善儿童呼吸道症状者，戒烟概率较低。伦理风险型技术的反应效能就是指只有在公众相信某种信息能够有效地减少威胁的情况下才会起到说服效果。伦理风险型技术科普技术反应效能弱化的原因主要有三个方面：

第一，伦理风险型技术不可控归因。技术风险具有客观实在性和主观建构性(闫坤如，2016)，现代技术风险寄生于影响当代人类生存的全球自然风险，由于技术进步而引发了人类社会发展的不确定性成为

“风险社会”(毛明芳，2009)的主要成因和重要特征。当前，技术风险是风险的一个子集合，按其产生的可预测性分为准确定性与不确定性、技术知识不确定性与技术研发等、应用与其他实践在逻辑上的错位影响了技术风险客观实在性的发挥，技术主体依据其知识结构，使用意向，生活背景，利害关系，风险距离，风险性质，风险态度，风险偏好和风险倾向构建风险。当科技理性，经济理性超过价值理性与社会理性时，科学理性与社会理性就会出现“断裂与缺口”(Beck, 2004)，科学技术与人文伦理失去平衡，便会为现代技术风险的产生提供温床。新兴技术的风险感知是基于个体之间风险性信息扩散后的耦合效应所产生的，客观上技术风险可能会与主体意愿相悖，损害个体利益，危害社会稳定和国家安全；而主观上，个体对技术风险的感知和担忧会影响公众的社会与政治上的重要判断(孙典等，2021)。如果大众认为伦理风险型技术产生与发展不可避免、无法控制，那么逃无可逃、避无可避的风险心理逻辑和规避思想就会根深蒂固，面对伦理风险型技术的科学普及就会采取消极的态度，伦理风险型技术科普信息便会无法发挥作用。

第二，科普作品质量良莠不齐，部分内容可信度不高。就科普短视频而言，明确的行业规范尚未形成，平台机构在内容审查，准确性和专业性都存在疏漏。在以社交媒体为代表的网络科学与健康传播场景中，“网络科学家”和“科学网络用户”构成了网络科学与健康传播场景的“去中心化”，这反映了他们对“自上而下”的技术科普路线的补充。值得注意的是，专业身份并没有对其知识科普和信息接受产生显著差异化影响。科学领袖的说服力更注重“内容”本身的质量，而非他们的“专业地位”，但同时，部分自媒体内容缺乏科学性，也在一定程度上弱化了伦理风险型技术科普权威。为了科普技术相关知识，科学媒体将晦涩难懂的理论和项目转化为通俗易懂的文化产品，以便公众通过科学媒介了解技术的发展趋势。部分科学共同体与科普基地开始尝试通过科学故事讲述自己的技术研究成果或发现过程。然而，有些媒体所创作技术相关的故事并未真实地展现技术的真实面貌，为了让这些故事更具吸引力，科普者对技术研究进行了筛选和改造，却忽略了其不确定性和风险性，同时也放大了技术研发和应用的局部片段。随着数字媒体时代的兴起，碎片化的阅读方式使得公众对于技术的全面真实了解变得更加困难，这种“碎片化”的技术知识内容在很大程度上削弱了受众对于如人工智能这类高新技术的认知能力和兴趣。

第三，伦理风险型技术指导欠缺专业化和细分化，未能聚焦细分人群和科普情境，使得涉及伦理风险的技术科普成为公众不学便知的“基本常识”。有学者基于风险感知理论，构建了风险的社会放大框架，即风险经由信息传递而被不同社会要素所影响，并经由心理机制与社会机制而被放大或者削弱公众对风险感知。具体表现为，在各种传播渠道(包括社会个体本身)放大镜的作用下，传播的可用信息通常作为带有倾向性的、经过加工的主观信息，当带有偏见的信息散布于个体之间时便会增强偏见，公众对于某些危险的认知被放大而个体对于采取某种保护性的行为的认知被缩小，进而使人们对于现实中的威胁产生一致低估。

3.4. 伦理风险型技术自我效能培植失焦

在应对评估时，自我效能(self efficacy)是指个体感知到自己有一定保护性行为能力后，在实施某种行为操作之前，对于自己能在何种程度上完成这种行为活动的信念，判断以及主观的自我感受等，这也是保护动机理论研究的核心内容。自我效能在行为发生与改变过程中极为重要，自我效能感越高，行为产生和变化的概率就越大。除此之外，由于公众间的科学素养参差不齐，不同公众也会产生截然不同的自我效能感。部分公众会仅将未成年人及常识欠缺者视为伦理风险型技术科普对象，这一部分公众由于现实或虚幻的自我效能过高，普遍选择忽略风险信息的传递。而对于另外一部分公众而言，尽管新兴技术多样化普及已经促进了信息通达性的提高，但是其仍然难以理解专业内容，并且由于低自我效能而产生

信息淡漠现象。

技术哲学视域下，技术作为人利用自然、改造自然的产物，是人的本质力量的彰显。但技术一经广泛应用，就会按照其内在的结构和作用形成自主性力量。这种自主性力量极有可能把技术带离预设轨道而脱离人的支配，造成对人的主体性的遮蔽。感知到技术风险的利益受损群体和既得利益群体为了达到诉求目标而“在表达诉求时往往热衷于展示诉求主体和诉求客体的身份特征，以迎合和激发某些领域中尚存的社会负面情感(陈相雨，丁柏铨，2018)”。与此同时，信息交互不充分，公众对于伦理风险型技术的诸多质疑与诉求未能得到及时回馈，继而导致伦理风险型技术自我效能的提高举步维艰。周敏，郅慧(2023)通过实验研究验证了感知信息过载对社交媒体用户隐私披露意愿影响并认为，用户感知到的信息威胁程度同样是导致隐私疲劳，影响其进一步行动的要素之一。人工智能与一般技术不同，是一种旨在模拟、延伸并超越人的颠覆性技术。它深刻地改变着人类社会组织与生活方式时，却又以内在智能技术范式形成一种对人类开放或隐伏的宰制，使人类成为马尔库塞笔下科技的奴隶、受操控、受支配的“单向度人”。比较典型的例子是外卖骑手的“困在系统”。在算法系统的“指挥”之下，他骑手们越来越沦为智能社会复杂体系中微不足道的“附庸”，无法有效培植自我效能，做出解放自身的正确行动。这有悖于以人工智能为代表的新兴技术以人为本的核心宗旨。

4. 科普效果提升路径：伦理－技术生态观的心理学探索

科学技术在给我们带来巨大便利的同时也正在作为一种社会性力量重组着人的“生存方式”(彭兰，2022)，当“数字化生存”逐步深化为“数据化生存”，人类的身体和生存方式也受到了技术化的影响。例如不管是出于协助还是治疗的目的，涉及伦理层面的医疗技术制品在延伸了人类身体性能的同时也使得人们在技术及市场制约下的自我量化中深陷个体赋权和外界制约的张力之中，并在某种程度上改变了人的生物学界定。由于伦理风险型技术在科学普及过程中的高度不可预见性和偶然性，伦理秩序应作为一种制度化参与方式内嵌到此类科学技术的发展过程中。这同时也要求了我们亟须从心理学角度出发，重视公众在伦理风险型技术科学传播中的心理逻辑，建立起不同以往的科普－伦理学的传统认知和建构模式，并强调伦理介入科普的必要性。

正如科技部科技监督与诚信建设司副司长冯楚建指出，要将伦理问题具体化，伦理－技术生态观的提出也需要“主动前瞻研判、强化监测预警、促进科技伦理风险预防关口前移、及时在规制方面予以回应，以实现科技创新与高水平良性互动高质量发展。”在未来媒体伦理风险性技术科普传播的策略方面，更应注重对于个体主体性心理感知的考量，合理使用展现公众诉求的传播策略，才能在实践中更好地提升公众的自我保护行为，达到伦理风险型技术的科普效果。美国国家标准与技术研究中心(NIST)给出了可解释人工智能系统须遵循的四项指导性原则，即：解释原则、意义原则、解释精确性原则与知识限度原则(Angelov et al., 2021)。将这些 NIST 给出的仅从技术层面出发的原则，置于伦理－技术生态观的这一大背景下，或许下列提出的幸福、信任、可持续发展性原则能够更加符合我国科普受众心理对伦理风险型技术的科普效果提升路径。

第一，幸福(happiness)：以人民需求和幸福为根本出发点。

互联网、云计算和人工智能等前沿技术的融合与发展加快了更成熟信息社会的来临，人类正步入更激进的技术型社会，这要求人们对智能社会中如何人机共生进行深入的思考(human-computer symbiosis)。一方面公众有权寻求数字福祉(digital well-being)，每个人都可以享受互联网技术所带来的方便与红利，但是现在技术鸿沟与数字鸿沟仍然存在，部分弱势群体无法完全享受数字技术所提供的便捷，继而造成伦理风险型的技术严重性与易感性认知的启动无力。因此，伦理风险型的技术科普应当进一步细分，拥有生活背景、受教育程度等不同特征的受众群体应采用不同的科普手段。另一方面，人人都有幸福工作

的权利。从长远来看，机器学习所代表的人工智能技术将会给人类社会、经济及工作带来深刻影响，但是人类的角色与功能并不会减弱，反而会随之增强。例如人工智能辅助诊疗工具能够帮助医生在医疗诊断中提高效率和准确性，制定出更加高效和个人化的治疗计划，同时医生还有更多的时间去从事和其他医生进行沟通以及安慰患者等、制定诊疗计划等等要求判断力、创造力、同理心、沟通、情商等能力的非结构性任务。《英国数字战略》显示，未来二十年内 90%以上的工作或多或少都需要数字技能。技术的采用与渗透通常要经历几年乃至几十年的时间，要在生产流程、组织设计、商业模式、供应链、法律制度以及文化期待各个方面进行调整与变革，中国作为未来的机器人大国，这些改变会对目前就业及经济结构产生重大影响。为保障劳动者应对这一科技巨变，有效地为劳动者提供数字技能和提升技术风险的社会化感知是关键。

第二，信任(trust)：以正确价值观为导向，保障伦理风险型技术的可利用，可依赖，可获知和可控制。

如前所述，互联网二十多年来不断引发的问题与近期人工智能等新技术对伦理与社会的冲击一起，说到底就是对技术信任的质疑。如今人们还不能充分相信人工智能的存在，这一方面是由于人们相关信息获取不充分，对于这些关系到生活生产的技术发展没有充分理解；另一方面是由于人们缺乏预见能力，他们既不能预测公司将用他们的数据来做些什么，也不能预测这些高新技术系统将采取什么行动。因此，目前急需塑造包含伦理风险型的技术信任——首先是可用性(available)信任。可用性要求以人工智能为代表的新兴技术发展要遵循以人为中心的原则，使科技真正能被人类所利用并赋予其价值，对每个个体都要公平平等，以免产生不公平歧视与技术鸿沟，激化或固化社会不公平。

其次是可靠性(reliable)。涉及伦理风险的技术应安全可靠，达到安全，稳定，可靠的目的。以数字网络安全为例，可靠性意味着人工智能应当遵守隐私法律要求、加强隐私保护与数据安全，并确保个体对于数据的管控权利，防止数据滥用。三是可知性(comprehensible)。伦理风险型技术应具有透明性，可解释性和人类可理解性，以免技术“黑盒”左右公众对人工智能的信任度。四是可控性(controllable)。伦理风险型技术的开发和应用应该被放置在人的有效支配下，以免损害人的个人利益或者整体利益。此外，“可控”还意味着在人机关系中，对于机器的决策和结果都必须考虑到由人来承担最终责任，以确保人机共生社会不会出现失范现象。从长远看，尽管目前尚不能预测通用人工智能与超级人工智能是否能达到及怎样达到，而且不能充分预测它们所产生的效果，在涉及伦理的风险型技术的科普过程中可适当运用说服技巧中的警钟效果，但也应遵循预警原则(precautionary principle)以预防未来风险，确保技术向善。

第三，可持续(sustainability)：践行“科技向善”，形塑一个健康、包容和可持续发展的智慧社会。

技术创新是促进人类及人类社会发展最为重要的因素，一方面，第四轮技术革命中蕴藏着极大的“向善”潜能，必将给人类生活和社会进步以突破性改善。换而言之，这类技术本身就是一种“向善”的手段，可成为一种“向善”的力量来解决人类发展所面临的各种挑战与难题，为实现可持续发展目标作出贡献。另一方面，联合国制定的《2030 可持续发展议程》确立了 17 项可持续发展目标，实现这些目标需要解决生态环境、人类健康、社会治理、经济发展等方面相应的问题和挑战，将技术正确运用到这些领域中，才是正确的“技术向善”之道。因此企业不应仅仅关注经济效益而忽视社会责任和社会效益，应自觉地、有针对性地设计、开发和应用技术，以应对社会挑战，在伦理风险性技术的科普过程中做到利益和价值相统一。如谷歌 2018 年 10 月曾启动了“人工智能向善”(AI for Social Good)工程，通过向人工智能开放，与社会机构共同解决世界上最为严重的社会、人道和环境问题。近年来，在野生动物保护、洪水预报、野火防范和婴儿健康等方面已经有了行之有效的解决方案。

基金项目

中国科协 2022 年度研究生科普能力提升项目“伦理风险型技术的科普解读与社会感知研究——以人

工智能医疗技术为例”(KXYJS2022064)阶段性研究成果；江苏省研究生实践创新计划“伦理风险型技术科学传播中的社会感知研究”(KYCX23_1097)研究成果。

参考文献

(德)乌尔里希·贝克(Ulrich Beck) (2004). *风险社会*(何博闻 译). 译林出版社.

陈相雨(2023). 在线共情叙事的情感影响及其风险预判. *江苏社会科学, (4),* 132-139.

陈相雨，丁柏铨(2018). 自媒体时代网民诉求方式新变化研究. *传媒观察, (9),* 5-12+2.

崔中良，卢艺(2023). 国外学者关于人工智能伦理问题研究述评. *国外理论动态, (2),* 160-168.

贾鹤鹏(2020). 国际科学传播最新理论发展及其启示. *科普研究, 15(4),* 5-15+105.

李莉莉，史云，李梦(2019). 基于保护动机理论的健康教育配合支持性心理干预对妊娠期糖尿病患者认知水平、负性情绪及妊娠结局的影响. *中国健康心理学杂志, 27(10),* 1478-1482.

刘彩(2015). 保护动机理论在健康行为解释、干预与预测中的应用研究. *医学与社会, 28(7),* 77-79.

刘鹏(2023). 科技·伦理·治理“三位一体”推动科技与善同向而行. *唯实, (4),* 27-30.

毛明芳(2009). 应对现代技术风险的伦理重构. *自然辩证法研究, 25(12),* 55-60.

裴东超，傅国惠，尹玥，等(2021). 保护动机理论的健康教育对脑梗死患者自我效能、心理状态及生活质量的影响. *中国健康心理学杂志, 29(11),* 1636-1641.

彭兰(2022). “数据化生存”：被量化、外化的人与人生. *苏州大学学报(哲学社会科学版), 43(2),* 154-163.

孙典，薛澜，张路蓬(2021). 新兴技术风险感知扩散机理分析. *科学学研究, 39(1),* 2-11.

王芸，肖霞，郑频频，等(2009). 保护动机理论在个体行为改变中的应用和发展. *中国健康教育, 25(11),* 853-855+870.

韦丽鹤，王纳，薛娜，等(2022)基于保护动机理论的健康教育对进展期胃癌患者服药行为、负性心理反应和疾病感知的影响. *中国健康心理学杂志, 30(3),* 372-377.

吴冠军(2020). 健康码、数字人与余数生命——技术政治学与生命政治学的反思. *探索与争鸣, (9),* 115-122+159.

闫坤如(2016). 技术风险感知视角下的风险决策. *科学技术哲学研究, 33(1),* 73-78.

杨青(2010). *基于保护动机理论的综合护理干预对减轻 2 型糖尿病患者足底压力的研究*. 硕士学位论文，上海：复旦大学.

杨英新(2022). 防灾科普短视频传播效果的心理审视. *青年记者, (20),* 48-50.

喻国明(2022). 元宇宙就是人类社会的深度“媒介化”. *新闻爱好者, (5),* 4-6.

喻国明，李钒(2023). 内容范式的革命：生成式 AI 浪潮下内容生产的生态级演进. *新闻界, (7),* 1-8.

张迪，童桐，施真(2021). 新媒体环境下科学事件的解读特征与情绪表达——基于新浪微博“基因编辑婴儿”文本的框架研究. *国际新闻界, 43(3),* 107-122.

周敏，郅慧(2023). 感知信息过载对社交媒体用户隐私披露意愿影响的实验研究. *新闻大学, (5),* 12-28+118-119.

Angelov, P. P., Soares, E. A., Jiang, R. et al. (2021). Explainable Artificial Intelligence: An Analytical Review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11,* e1424. https://doi.org/10.1002/widm.1424

Prentice-Dunn, S., & Rogers, R. W. (1986). Protection Motivation Theory and Preventive Health: Beyond the Health Belief Model. *Health Education Research, 1,* 153-161. https://doi.org/10.1093/her/1.3.153

Stainback, R. D., & Rogers, R. W. (1983). Identifying Effective Components of Alcohol Abuse Prevention Programs: Effects of Fear Appeals, Message Style, and Source Expertise. *International Journal of the Addictions, 18,* 393-405. https://doi.org/10.3109/10826088309039356