

引用格式：鲁晓, 刘慧晖, 李瑞, 等. 前沿科技的伦理风险：日益增长的威胁. 中国科学院院刊, 2024, 39(11): 1835-1844, doi: 10.16418/j.issn.1000-3045.20240516004.

Lu X, Liu H H, Li R, et al. Ethical risks of frontier science and technology: An escalating threat. Bulletin of Chinese Academy of Sciences, 2024, 39 (11): 1835-1844, doi: 10.16418/j.issn.1000-3045.20240516004. (in Chinese)

前沿科技的伦理风险： 日益增长的威胁

鲁晓^{1,2} 刘慧晖¹ 李瑞¹ 潘教峰^{1,2,3*}

1 中国科学院科技战略咨询研究院 北京 100190

2 中国科学院大学 公共政策与管理学院 北京 100049

3 中国发展战略学研究会 北京 100190

摘要 科技是发展的利器，也可能成为风险的源头。当今世界，新科技革命和产业变革加速演进，深刻影响和重塑人类生产生活，引发诸多已知和未知的伦理风险，攸关个人权益、社会稳定、国家安全和人类永续发展。文章凝练前沿科技伦理风险的特征，包括领域差异性、未来不确定性、社会重构性、终极挑战性。综合考虑前沿科技各领域的风险，从安全风险、生命风险、隐私风险、环境风险、道德风险、灾难风险、技术叠加风险等7个维度进行阐释，为化解前沿科技伦理风险，明确治理重点和安全护栏，有效构建前沿科技伦理治理的规则和机制，提出整体性对策建议。

关键词 前沿科技，伦理，风险，治理

DOI 10.16418/j.issn.1000-3045.20240516004

CSTR 32128.14.CASbulletin.20240516004

当前，人工智能、基因编辑等前沿科技以前所未有的速度推动产业变革、社会进步、增进人类福祉，愈发深入地融入人类社会，极大地改变人类传统的工作和生活方式，伦理风险挑战也相伴而生。例如，以CRISPR-Cas9为代表的基因编辑技术本身存在的“脱

靶效应”引发人们对其应用安全性的担忧；无人驾驶、医疗机器人等面临隐私数据泄露、责任归属不清等质疑。科技伦理治理旨在将前沿科技活动规范在人类可控的轨道上，对前沿科技可能造成的危害加以预警，防范和化解前沿科技创新及应用中引发的伦理风

*通信作者

资助项目：科技创新2030—“新一代人工智能”重大项目（2022ZD0116407），国家社会科学基金重大项目（23&ZD149）

修改稿收到日期：2024年10月28日

险挑战。这有助于引导前沿科技向着有利于人类文明进步、促进人类可持续发展的正确道路上前进，从而确保前沿科技在伦理道德框架下健康发展，最终实现科技创新高质量发展和高水平安全的良性互动。

相较于传统科学技术，前沿科技带来的伦理风险更为复杂，关涉个体自身、社会结构乃至人类永续发展。前沿科技使人类有能力对自身进行更深入地改造和干预，对社会结构和运行方式构成根本性的挑战，甚至可以在某种程度上实现对自然界的操控、重塑人与自然的关系。同时，随着全球化的发展，前沿科技的研发与应用已超越单一国界和地域的限制，对全球产生深远影响。由此来看，前沿科技所产生的伦理风险是底层性、全局性、结构性的，需要对其进行机理性、整体性、系统性的深刻审视与考量，这是进一步的规则制定和有效治理的前提条件和理论基础。

一些前沿科技领域的发展仍处于尚未成熟的阶段，伴随着较大的不确定性；其可能造成的伦理风险将引致广泛而深刻的社会争议，对人类社会道德边界带来严峻考验，“基因编辑”婴儿就是一个典型例子。传统上的“先发展后治理”的方式已然不适应前沿科技的伦理治理，而是需要一种更为前瞻性的策略，提前对前沿科技的伦理风险进行深入研判，识别其可能产生的多种维度的伦理风险，从而提出更为科学合理的预防之道。为此，本文对前沿科技发展带来的伦理风险进行系统性的前瞻识别，提出加强前沿科技伦理治理的对策建议，以期抓好科技向善的“方向盘”，建好科技事业健康发展的“瞭望塔”，推动我国前沿科技领域向善向好、健康有序发展，同时为我国前沿科技发展开辟新的空间，增加更多可能性。

1 前沿科技伦理风险的特征阐释

当前，随着基础科学研究不断向宇观拓展、微观深入和极端条件方向加速纵深演进，物质结构、生命起源、宇宙演化等重大前沿问题有望取得突破。信

息、生物、能源、先进材料与制造等前沿科技领域呈现多技术交叉融合和群体性跃升态势。以高速移动通信、物联网、人工智能、大数据、区块链、量子信息等新一代信息技术、智能化技术为引领的新技术体系，加速技术和产业质的嬗变，推动人类社会向着人机物三元融合发展。整体上看，前沿科技是指高技术领域中具有前瞻性、先导性和探索性的重大技术，是具有突破科学探索边界、引领技术发展方向和形成新质生产力潜力的技术，是新科技革命和产业变革的核心驱动力量。前沿科技探索和开拓未知领域，从科技伦理视角看，传统的伦理边界日益突破，伦理风险种类增多、潜在隐患成为现实，识别难度急剧增大。前沿科技伦理风险的特征体现为领域差异性、未来不确定性、社会重构性和终极挑战性。

1.1 领域差异性

前沿科技建立在知识密集突破和新颖创造的基础上，其伦理风险因技术本质属性的不同而存在领域差异。科学技术是人类为自身及其社会文明进步与发展而创造的文明化方式或工具，其发展属于客观现实的范畴。而科技伦理是建立在科学技术客观实在基础上的价值关系，具有技术依赖性^[1]。现代科学技术高度发达的重要标志之一是科学技术知识的专业性、密集性，更重要的是其在专业领域和知识生产体系中的快速分衍与更新^[2]。

当前，前沿科技正在大规模、多点式、群簇性爆发，形成了差异化的多元领域，由此使得前沿科技的发展演化呈现明显的领域差异性。在此基础上，科技伦理问题因不同的科学知识和技术应用而存在领域差异。一方面，这要求对于前沿科技伦理问题讨论诉诸当代科学技术发展本身，探微特定领域的科学技术最新理论成果和技术方法，并在识别特定领域的科学技术本质规律和演化特征的基础上，深入细致探讨其独特的伦理问题。另一方面，对现代科技伦理的探讨需要保持多元领域的开放性及前瞻性，对多元交叉领域

的科技伦理问题保持敏感性，促进建立在专业化基础之上的跨学科、跨领域的交叉互联。

1.2 未来不确定性

前沿科技因其探索路径、应用场景的多样性，以及人类社会对技术进步的复杂反应，使得伦理风险更加难以预测、识别和应对。前沿科技快速应用及发展具有技术路线复杂多变、应用场景多元拓展的特性，尤其是在科学技术尚未成熟时期，知识产生和技术演化方向都具有很大的不确定性。同时，伴随着研发过程的推进，将会产生如何评估未来产品、用途、应用场景和后果的不确定性，以及如何解决随之而来的种种原则和规范的不确定性^[3]。

此外，由于人类应用技术的目的及途径具有不确定性，前沿科技伦理风险具有未来不确定性。技术本身是中性的，可以用于为善也可以用于作恶。技术能产生怎样的影响，取决于使用它的主体出于何种目的，或使用过程中存在的特定逻辑。人类对技术应用目的的“设计”贯穿于一切科技活动的始终，甚至已经深刻影响到了人的心理（如行为控制技术）和生理（如基因工程）的活动。因此，前沿科技引发的伦理风险也必须从技术设计和创新的初始阶段加以考虑^[4]。

总之，前沿科技的爆发式增长及扩张，以及技术无所不在的实施运用，都会给现代人类社会带来不确定性的伦理风险和挑战。因此，需要采取事先预防、前瞻性、敏捷性的治理方式，将技术的可能性风险事先考虑、研判，并将技术的伦理道德价值嵌入研发的全过程。

1.3 社会重构性

前沿科技领域的发展是一个包含着科学技术突破、人类思想认知重塑、社会文化革新的广泛而深刻的整体性变革。前沿科技可能会影响甚至重塑人类的认知模式、思维方式、教育理念等，催生新思想、新价值观、新文化，带来社会结构变革。

(1) 以人工智能、量子科技、生物科技、太空探

索等为代表的前沿科技在时空范畴拓宽了人类对于客观世界的认知，这使得人的思维习惯也将从线性、因果模式向非线性、非直观方式转变。同时，伴随着前沿科技对知识信息获取和传播逻辑的重塑，教育理念也要求同步变革，将更加凸显自主性、交互性和创新性^[5]。学校不再是一个封闭系统，而成为知识生产与创造的中心，成为与其他社会主体协同创新的重要力量。更为重要的是，以人工智能为代表的前沿科技成为新的知识发现工具，乃至有可能成为知识发现的主体，将会重塑人类知识体系，加速现实构造的变化，并深刻影响人类政治和社会。

(2) 前沿科技的涌现将人类从传统上依赖自然物发展到依赖智能物或者人造物，人与智能机器可能形成新的生产生活关系，人与人、人与社会、人与自然的关系将会发生重大变革。与之相对应，人类对于生命、人性、社会、自然的理解和共识将会重构，从而形成新的关于生命本质、社会运行规律的精神思想，也将冲击人类现有的关于善恶和道德的价值体系，进而形成新的社会规范和文化思想。

(3) 当这些技术组成的技术群汇聚融合并使得形成新的产业体系成为可能，会形成新的利益相关方格局，由此使得传统的社会关系、收入分配、就业形态等面临巨大冲击，形成人类社会新的权力来源。更重要的是，当前沿科技突破改变了已经形成的社会规则、社会稳态，可能打破原有的社会公平公正理念。技术获取的不对等性将导致社会贫富两极分化，加剧社会矛盾。快速发展的科技及应用所带来的社会矛盾可能在某个科技事件上集中释放，产生“黑天鹅”事件，继而对社会结构、社会稳定性带来伤害。

1.4 终极挑战性

前沿科技使设计和改造生命成为可能，重构人类生命自然本体；创造新的智能机器物理世界，改变人类生存的自然情境，构成对人类生存与发展的终极挑战。

(1) 从人类生命的自然本体而言，随着前沿科技的突破式进展，人类在某种程度上已经超越了达尔文进化论以来“格物致知”的基本范式，进入到人造生物体、“造物致知”的新阶段^[6]。人类发展出可以理解和操纵生物系统的能力，改变了人类内部世界（身体和精神）成长发育的自然轨迹。由此带来了对人的自然权利、公平公正问题的重新理解，进而引发了诸如生命的定义及价值、合成生命的道德地位、物种界限等一系列关涉人类本身生存发展的问题。

(2) 从人类生存的自然情境而言，以新一代信息技术为代表的前沿科技应用已经将人类从人机或物机二元世界融合推向了人机物三元融合的世界^[7]。物理空间、信息空间、能源空间、社会空间源源不断地进行着信息交互和资源的有机融合，并在人类智慧和机器智慧的共同作用下推动人类社会向前发展。三元融合在增强了人类的感知、共情、行为和应变能力，为人类生存带来个性化产品和服务的同时，也引发了诸如“人类在多大程度上接受人工情感、虚拟社会关系”“人的专注力、感知力和精神情感是否应该商品化”“关于人的尊严情感、人的社会价值、人与机器关系等一系列影响人类作为‘智人’生存方式和意义”的伦理拷问。

科技是把“双刃剑”，在前沿科技伦理挑战面前，如何平衡科技发展与伦理风险是当下需要深入讨论的重要议题。一方面，前沿科技具有巨大潜力，可以深刻改变人的生存情境。另一方面，伦理风险的警钟需要长鸣，需要人类社会共同努力，以确保科学和技术的应用是负责任的，构建和谐向善的人类命运共同体。

2 前沿科技伦理风险的类型识别

前沿科技发展对人类社会带来深刻变革和长远影响，这些伦理风险不能通过简单的伦理原则加以规范，而是需要在对其进行系统性识别和认知的基础

上，探寻具有全局性、前瞻性的应对策略。本文将前沿科技伦理风险分为安全风险、生命风险、隐私风险、环境风险、道德风险、灾难风险、技术叠加风险七大类型，剖析各类风险的成因及影响，并以风险最突出的前沿科技为例进行深入分析。

2.1 安全风险：前沿科技发展及应用的不确定性对人类造成安全隐患

当前前沿科技处于探索阶段时，技术本身仍然不成熟、不完善，其未来发展及应用也存在极大的不确定性，将对人类生命健康、社会有序发展造成安全隐患。安全风险主要源于前沿科技本身的特点。例如，基因编辑技术存在“脱靶效应”，可能会对非靶位点位置的基因进行切割，用于人体将带来不可预测的副作用；若母体子宫内植入脱靶胚胎细胞，可能使后代患上难以根治的遗传性疾病甚至导致先天性畸形。深部脑刺激、运用超声和光遗传学对脑区进行调控、药物神经增强等新技术对大脑功能的长期影响存在着不确定性。再生医学和干细胞研究存在技术不确定性，细胞可能增殖失控，并在植入部位产生肿瘤。无人驾驶汽车本身技术的不成熟和应用场景的不确定性，可能出现难以定责的事故，涉及电车难题的伦理困境，更加难以解决。3D打印技术可能被滥用于打印枪支器械、化学毒品等，带来安全风险。据有关报道，枪击日本前首相安倍晋三的嫌犯所使用的枪支可能是3D打印手枪^[8]。一些纳米材料会损害人体内的细胞，一旦被吸入肺脏，可能引发致命性病症。

2.2 生命风险：前沿科技改造人类自然本体，颠覆生命伦理认知

新生物技术作为前沿科技的重要领域，其快速发展使人类从认识生命走向改造生命、合成生命甚至设计生命，颠覆人们对生命和自然的认知，模糊生命与非生命、自然物与人造物的界限，对人类的生命价值和生命尊严造成威胁。生命风险是由于一些前沿科技的出现打破了生命与非生命的天然界限，

对传统生命含义、起源、本质、价值和意义等观念产生冲击而引发的，这种冲击将影响、撼动人们对生命界定乃至人类生命尊严的基本理解。例如，基因编辑涉及人类基因组的修改，人为干预人类遗传多样性引发关于人类是否可以干预生命进化的争议。人-非人灵长类嵌合胚胎的研究成为干细胞与再生医学研究的前沿，同时也带来人们对于神经系统、生殖系统等组织嵌合的担忧。合成生物学将工程学理念和策略引入新生命形式的研究和创造，在分子水平实现对生命系统的重新设计和改造，重塑自然生命形态甚至打破物种界限，将对生命本质的认知及相关认识论提出挑战。

2.3 隐私风险：前沿科技带来的个人隐私数据泄露和滥用问题凸显

随着人工智能、大数据等新一代信息技术的发展，个人隐私数据成为一种重要的生产资源。隐私数据有可能被非法收集、使用、泄露甚至滥用，对个体的隐私权、信息知情权与选择权造成挑战。隐私风险主要是由于个人隐私数据在采集和使用过程中的不当使用和处理所造成的，对于个人信息、个人空间、个人自由都将带来极大的伤害。例如，人工智能、物联网等技术发展下的各种移动终端、摄像头、传感设备、可穿戴设备可以实现对用户在线、离线行为的全记录，使得个人隐私的获取方式更为便捷和隐蔽，引发大量与隐私有关的伦理与法律监管问题。有研究表明，在虚拟现实（VR）模拟中停留 20 分钟将被记录近 200 万条独特的肢体记录、动作信息^[9]。

通过大数据技术对多元数据及其足迹进行信息整合和融合，可用于推测用户的行为、情绪、偏好等信息，造成对个人隐私权利的严重侵犯。厂商利用大数据技术分析用户隐私数据和预测用户行为偏好，并广泛运用于信息推送、广告推介，弱化人们的独立思考能力与行动自主权。以 ChatGPT 为代表的大语言模型，以人类现有的知识库、数据库为训练语料，常常

伴随着隐私侵犯的风险，并可能带来数据滥用、深度伪造的问题。

2.4 环境风险：前沿科技发展及应用带来新的环境问题

新材料、新能源、智能技术、基因驱动等新技术在其应用过程中，往往伴随着一系列新的污染物的产生；这些污染物难以降解，有新的污染作用和毒性机理，新技术带来的能源消耗也引发环境问题。例如，太阳能电池板的生产过程中会产生大量废水、废气和废渣，其中含有多种有害物质，若没有经过妥善处理直接丢弃或排放，将严重污染环境。又如，人工智能高算力需求引发大量能源消耗问题。据估计，ChatGPT 每天推理计算需要消耗超过 50 万千瓦时的电力，是美国普通家庭每户每天用电量的 1.7 万倍^[10]。电子产品的快速迭代引发电子废物处理难题，据 2024 年联合国训练研究所和国际电信联盟共同发布的《全球电子废物监测报告》显示，2022 年全球电子废物量高达 6 200 万吨，但仅有 22.3% 的电子废物能被妥善处理^[11]。基因驱动技术可能使改造基因在自然种群中迅速扩散，导致一些物种的减少或灭绝，干扰原有的自然生态平衡。

2.5 道德风险：前沿科技影响改变人与人、人与社会间的伦理道德关系

前沿科技在人类生产生活中的应用，不断影响、改变着现实社会，给人与人、人与社会间的传统伦理道德关系造成影响甚至带来颠覆性的改变。道德风险的产生，根源在于前沿科技及其应用对于人类社会价值理念、伦理道德的撼动，引致与传统伦理道德的矛盾凸显，乃至出现激烈冲击、尖锐对立，带来伦理道德秩序的紊乱甚至失范。例如，线粒体置换技术使后代同时拥有亲生父母和女性线粒体捐赠者的遗传物质，这种“一父两母”的遗传模式对传统“一父一母”的亲子关系造成冲击。克隆技术为干细胞研究提供细胞供体，能够通过复制胚胎干细胞制造与原胚胎

具有同样遗传基因的个体，这颠覆了原有的有性繁殖方式，对传统的由血缘确定亲缘的伦理道德规则带来挑战。人胚胎干细胞研究中使用的胚胎多为辅助生殖捐赠的“剩余”胚胎，但实验室研究往往涉及对胚胎的干预甚至破坏，这引发破坏人胚胎是否道德的争议。智能机器人可能会导致人类个体对其产生情感和依赖，影响其与他人的正常交流，导致社会关系的疏离和淡化，甚至引发与传统婚姻家庭观念相冲突等道德问题。

2.6 灾难风险：前沿科技发展的失控及滥用可能对人类带来毁灭性打击

前沿科技在为人类带来利益和生产力的同时，其“双刃剑”属性也愈发凸显，一旦人类不能阻止前沿科技的失控、无法限制前沿科技的滥用，将给人类带来巨大灾难。灾难风险主要是由于前沿科技发展偏离人类控制所造成的，特别是当前前沿科技发展的方向偏向于破坏性科技时，灾难风险尤为严重，危及人类生存，甚至对人类造成毁灭性打击。利用人工智能的智能解码能力，可入侵、操控、破坏国家关键基础设施或公共服务系统，如操控交通系统导致大规模交通事故、操控医疗系统影响患者救治、操控金融系统引发经济恐慌和社会动荡，这些都将引发社会系统的大规模功能失效，产生人类难以承受的后果。人工智能的武器化应用，展现出高效的杀伤能力，一旦被大规模投放到战争中，将使人类面临类似核武器一样的毁灭性威胁。合成生物学在病毒“功能获得性”研究的运用，如复现已灭绝的“西班牙”流感病毒、生成天花病毒“近亲”马痘病毒等，若被恐怖分子滥用于制造生物武器、开展恐怖袭击，将严重威胁人类生存。

2.7 技术叠加风险：多项前沿科技耦合叠加带来的复杂风险挑战

当不同领域的前沿科技相互交织、叠加应用时，往往产生不同于单一前沿科技领域的新的伦理风险。技术叠加风险主要源于多项技术的叠加耦合和技术应

用场景的多元化，这些情况将加剧伦理风险的不确定性和复杂性，并使影响受众和覆盖范围急速扩张。具体而言，脑机接口、医疗机器人、人脸识别等属于生命科学和信息技术的交叉领域应用。例如，侵入式脑机接口技术走向临床试验和应用阶段，直接作用于人脑，具有安全风险；特别是脑机接口技术干扰了受试者的大脑意识，如何获取明确的知情同意成为一个难题。国际学界也在讨论神经权利是否应作为一种新的人权。医疗机器人涉及设计者、开发者、供应商、医院、医生等多方责任主体，其在自主决策过程中发生的医疗事故，难以确定责任归属和明确界定各主体的责任分配，面临责任归属不清的问题。欧盟《人工智能法案》禁止在公共场所使用实时远程生物识别系统，主要是避免公共场所实时识别的人脸数据被滥用于跟踪、监控个人行为。

3 关于加强我国前沿科技伦理治理的对策建议

为应对前沿科技伦理治理面临的挑战，参与全球科技治理，我国急需树立“以人为本、科技向善”的伦理立场，实现科技创新与科技伦理的良性互动，为此提出6个方面的建议。

3.1 树立鲜明国际立场，平衡好伦理治理和科技发展之间的关系

伦理治理既关乎前沿科技创新之发展，也涉及国家竞争力之打造，需要平衡好伦理治理与科技发展之间的关系，实现良性互动、共同促进。

(1) **建立鲜明的国际伦理立场。**积极倡导“以人为本、科技向善”的价值观及原则，向国际表明我国在前沿科技研发、应用、监管等方面的伦理立场和态度。我国应尽快明确伦理方面的原则立场，对外融入全球科技治理体系，引领相关规则制定，增加政策影响力；对内规范科技和产业领域的发展，保持市场活力，提升技术影响力。

(2) **确立合理的伦理治理规则。**针对我国必争、必抢、必占的前沿科技领域，前瞻性研判伦理风险、确定伦理治理规则，使其能够在合乎伦理道德、符合国际伦理规范的道路上稳步前行，持续保持我国在这些前沿科技领域上的“领跑”优势。特别是针对一些勇闯“无人区”的前沿科技领域，其窗口期非常短，需要快速制定相应伦理规制措施，确保伦理规则能够跟上前沿科技发展步伐。

(3) **明确科技伦理的目标定位。**科技伦理不应通过固定、刻板的伦理规则来约束、阻挠前沿科技的发展，而应成为前沿科技发展保驾护航的工具，使前沿科技给社会带来的巨大冲击“软着陆”。同时，要将科技伦理作为开展科技创新活动的“基线”，科技立法作为不可逾越的“底线”，伦理价值观则作为科技创新中追求卓越的“高线”。

3.2 识别前沿科技伦理风险，明确治理重点，设置安全护栏

前沿科技领域的伦理风险具有多样性、泛在性、动态性、复杂性的特征，需要在风险识别和研判的基础上，构建分级分类的治理体系，并对重点领域设立安全护栏。

(1) **建立前沿科技领域风险识别体系。**对前沿科技领域进行风险研判，并以此进行分级分类治理。建议前沿科技风险识别体系包括2个维度：① **前沿科技本身的风险**，主要针对科学知识和安全信息的充分程度、技术成熟度、操作难易程度、安全性、有效性、可控性进行分析；② **前沿科技引发的伦理风险**，包括安全风险、生命风险、隐私风险、环境风险、道德风险、灾难风险、技术叠加风险等。对每种类型风险发生的可能性、严重程度、影响范围等进行分析。在此基础上，充分考虑前沿科技活动的必要性、目标人群、目标应用场景等，对不同领域、不同类别、不同应用场景的前沿科技伦理风险进行前瞻研判和分级分类治理。

(2) **明确治理重点领域并构建安全护栏。**要在综合考虑国家安全、经济社会发展需求、我国比较优势和国家科技战略重点领域的基础上，结合技术本身风险及其可能引发的社会风险，为人工智能、大数据技术、脑科学研究、合成生物学、基因编辑、干细胞和再生医学、脑机接口技术等伦理风险突出的科技领域构筑安全护栏。安全护栏应按照技术风险的影响范围、应用场景、受众和实验对象等不同要素设置“高一中一低”的风险阈值，并根据技术的发展和对其反应进行动态调整。需要注意的是，要对前沿科技的不确定性有一定的宽容度，对技术的快速发展不宜设置过多障碍。例如，大模型的优化需要持续的数据反馈流动才能实现。如果治理标准过高，部分企业可能无法上线大模型应用，进而阻碍技术的发展进程。

3.3 结合不同前沿科技领域特点，开展差异化伦理治理

针对信息技术、生命科学、信息生命交叉领域，准确把握领域特征，结合实际情况差异化开展伦理治理。

(1) **信息技术领域，重点关注当前快速发展的生成式人工智能和大模型的应用研发。**从技术应用的场景、功能、对象、影响等维度进一步落实分级分层监管举措，细化可执行的风险监测指标。强化技术人员伦理嵌入的教育引导，从研发到部署的全流程坚持技术伦理风险可控；通过价值对齐、过滤对抗测试等技术手段，使大模型与人类社会价值观相符合。完善大模型底座的安全可控和伦理嵌入。重点关注高质量中文数据库的建设。

(2) **生命科学领域，加强伦理审查委员会建设和资质评估，优化伦理审查制度。**加强基层伦理审查委员会建设，建立伦理审查委员会认证制度，落实地方和相关行业主管部门监督管理和复核审查职责。跟进优化伦理审查制度，合理调整风险等级划分，对低风险等级的研究精简审查流程。避免“多主体”和“再

申报”情况下的重复性审查。例如，借鉴美国人类研究保护办公室和美国食品药品监督管理局的联合审查规定，针对多机构的合作考虑采取联合审查。

(3) 信息生命交叉领域，持续跟踪前沿动态，加强伦理监管和标准建设。加强对人工智能技术介入医疗的监管控制，明确医疗机器人的法律定位、人工智能系统与医生的交互关系、医疗事故后的责任归属问题等。加强对通过信息技术收集指纹、虹膜、面容、DNA、神经信号等人体生物数据的监管，避免隐私侵犯。加强国内外脑机接口发展的动态跟踪监控，制定统一的市场准入安全认证标准。

3.4 加强统筹协调和责任主体的权责统一，建立重大科技伦理事件的快速响应机制

各部门、各主体在伦理治理方面的统筹协调及权责统一至关重要，需要平衡好各部门、主体之间的关系，明确责任划分。

(1) 统筹协调各部门伦理监管和审查程序。明确各部门在伦理监管方面的权责，并结合已有监管体系开展伦理治理。建立跨部门的工作组织或委员会，负责伦理审查方面的协调工作，及时分享相关信息和数据，避免重复性的伦理监管和审查。

(2) 充分发挥科学共同体和行业学会的作用。由学术组织、学术团体等牵头，研讨不同领域面临的前沿科技伦理风险，制定不同领域的伦理指南和标准，并根据前沿科技发展进行动态调整和修订，促进有关科学共同体的共识达成，以灵活应对具体领域的伦理风险，促进负责任的研究。

(3) 充分发挥企业等责任主体作用，有效开展伦理治理。鼓励人工智能和生物医药领域的企业联合制定行业技术标准，将伦理道德因素嵌入到技术标准中；这些标准应能有效应对所在领域技术前沿的伦理挑战，以指导企业的行为和实践。引导企业自主设立伦理审查机制，积极参与伦理讨论和政策制定，并公开展示其伦理实践情况，提高透明度，增强公众和利

益相关者的信任。

(4) 建立重大科技伦理事件的快速响应机制。确保重大科技伦理事件的公平和透明处理，加强国际国内社会的信任。政府机构、学术机构和行业协会联合建立响应机制，快速响应和处理违反伦理规则的重大事件。确保各部门在响应科技伦理事件方面有明确的职责，建立公开透明的报告和反馈渠道，及时向国际社会和国内公众公布有关事件的处理进展情况。

3.5 加强多元利益主体沟通，促进多方共治

前沿科技伦理的问题本质上是由于科技快速发展下多元主体的利益冲突和不平衡所带来的，需要考量多方利益主体的诉求，实现多元共治。

(1) 促进多元主体的沟通对话。重视来自各级政府、高等院校、科研机构、医疗机构、科技企业、科学共同体、社会公众、媒体等利益相关者的参与。加强多元利益主体的沟通、交流、对话，在反复的沟通协商中达成不同利益主体关于科技伦理的共识。

(2) 建立多元主体联动机制。加强政府层面统筹，强化有关主管部门联动，促进多元主体有效沟通、密切配合。压实高等院校、科研机构、医疗机构、科技企业等在多元共治中的主体责任，有效发挥科学共同体的自治作用，通过宣传、引导不断提高公众参与科技伦理治理的积极性，形成科技伦理治理的最大合力。搭建常态化交流平台，建立科学家与伦理学家、法学家的联动机制，对不断发展的前沿科技伦理问题进行定期研讨和动态更新。

3.6 积极开展国际合作，参与国际科技伦理治理建设

加强我国在科技伦理领域的国际合作，积极发出中国声音，提高我国在科技伦理方面的公信力和国际影响力，推动我国科技事业的健康有序发展。

(1) 举办国际科技伦理研讨会。积极寻找国际科技伦理领域的合作者，针对科技伦理问题进行深入研究，形成有影响力的研究成果。定期举办国际科技伦

理研讨会，涵盖信息、生命、信息生命交叉等前沿科技伦理领域的关键问题。

(2) 培育具有国际视野的伦理专家。在生命科学、医学和人工智能领域，培育一批具有国际视野的伦理专家，积极开展国际对话，在领域内有影响力的国际期刊发表学术评论文章。鼓励我国专家积极参与伦理方面的国际合作研究项目，与国际团队共同探讨伦理问题，了解国际伦理治理动向，及时发出中国声音。

参考文献

- 1 马克思, 恩格斯. 马克思恩格斯文集(第一卷). 北京: 人民出版社, 2009.
Marx K, Engels F. Collected Works of Marx and Engels: Volume I. Beijing: People's Publishing House, 2009. (in Chinese)
- 2 万俊人. 理性认识科技伦理学的三个维度. 光明日报, 2022-02-14(15).
Wan J R. A rational understanding of the three dimensions of technology ethics. Guangming Daily, 2022-02-14 (15). (in Chinese)
- 3 李磊, 鲁晓. 新兴科技的伦理治理: 概念、框架与实践路径分析. 自然辩证法研究, 2023, 39(9): 84-90.
Li L, Lu X. Ethical governance of emerging science and technologies: An analysis of concepts, frameworks and practical approaches. Studies in Dialectics of Nature, 2023, 39(9): 84-90. (in Chinese)
- 4 刘大椿. 科技与伦理何以内在统一. 社会科学文摘, 2023, (6): 21-23.
Liu D C. Why science, technology and ethics are internally unified?. Social Sciences Digest, 2023, (6): 21-23. (in Chinese)
- 5 王继新, 郑旭东, 黄涛. 非线性学习: 数字化时代的学习创新. 北京: 高等教育出版社, 2012.
Wang J X, Zheng X D, Huang T. Nonlinear Learning: Learning Innovation in the Digital Era. Beijing: Higher Education Press, 2012. (in Chinese)
- 6 刘陈立. 未来已来! 合成生物学将掀“定量”热潮. 中国科学报, 2021-05-10(8).
Liu C L. The future is here! Synthetic biology set to usher in a “quantitative” Boom. China Science Daily, May 10, 2021 (8). (in Chinese)
- 7 潘教峰. 新科技革命与三元融合社会——关于雄安新区建设的宏观思考. 中国科学院院刊, 2017, 32(11): 1177-1184.
Pan J F. New science and technology revolution and ternary fusion society—Macroscopic views on Xiongan new area construction. Bulletin of Chinese Academy of Sciences, 2017, 32(11): 1177-1184. (in Chinese)
- 8 Kageyama Y. Abe's killing haunts Japan with questions on homemade guns. (2022-07-11) [2024-05-16]. <https://apnews.com/article/shinzo-abe-japan-crime-tokyo-gun-politics-6ef3aa271e147bf2426363448ecd9f1b>.
- 9 Bailenson J. Protecting nonverbal data tracked in virtual reality. JAMA Pediatrics, 2018, 172(10): 905-906.
- 10 Kolbert E. The obscene energy demands of AI. (2024-03-09) [2024-11-06]. <https://www.newyorker.com/news/daily-comment/the-obscene-energy-demands-of-ai>.
- 11 Baldé C P, Kuehr R, Yamamoto T, et al. The Global E-waste Monitor 2024. Geneva: International Telecommunication Union, United Nations Institute for Training and Resources, 2024.

Ethical risks of frontier science and technology: An escalating threat

LU Xiao^{1,2} LIU Huihui¹ LI Rui¹ PAN Jiaofeng^{1,2,3*}

(1 Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;

2 School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China;

3 Chinese Association of Development Strategy Studies, Beijing 100190, China)

Abstract Science and technology (S&T) are not only powerful tools for development, but also a source of risk. The new round of scientific and technological revolution and industrial transformation are accelerating in current times. It deeply influences and fundamentally reshapes human production and life, which can raise a range of known and unknown ethical risks, involving individual rights and interests, social stability, national security, as well as sustainable development of mankind. This study summarizes the characteristics of ethical risks in frontier S&T, including the dependence on related fields, high uncertainty, social reconstruction, and ultimate challenge. Considering the risks in various fields, this study illustrates the risks from seven aspects, including security risk, life risk, privacy risk, environmental risk, moral risk, disaster risk, and technological overlay risk. Based on these, this study provides holistic suggestions for constructing S&T ethical governance system, coping with the ethical risks of frontier S&T, and clarifying the key points of governance and the safety barriers.

Keywords frontier science and technology, ethics, risks, governance

鲁 晓 中国科学院科技战略咨询研究院研究员, 中国科学院学部科学规范与伦理研究支撑中心执行主任, 中国科学院大学公共政策与管理学院智库科学与工程系主任。《中国科学院院刊》青年编委。研究方向为科技政策与管理、新兴科技治理。
E-mail: luxiao@casisd.cn

LU Xiao Professor of Institutes of Science and Development, Chinese Academy of Sciences (CAS), Executive Director of Research Center for Scientific Norms and Ethics, Academic Divisions of the Chinese Academy of Sciences, Dean of the Department of Think Tank Science and Engineering of the School of Public Policy and Management, University of Chinese Academy of Sciences (UCAS), and Youth Editor of *Bulletin of Chinese Academy of Sciences*. Her research interests include science policy and management, and the governance of new emerging technologies. E-mail: luxiao@casisd.cn

潘教峰 中国科学院科技战略咨询研究院院长、研究员, 中国科学院大学公共政策与管理学院院长。中国发展战略学研究会理事长,《中国科学院院刊》编委。主要从事科技战略规划、创新政策和智库理论方法研究。E-mail: jfpan@casisd.cn

PAN Jiaofeng Professor, President of the Institutes of Science and Development, Chinese Academy of Sciences (CAS), Dean of the School of Public Policy and Management, University of Chinese Academy of Sciences (UCAS), Chairman of Chinese Association of Development Strategy Studies, and Editor of *Bulletin of Chinese Academy of Sciences*. His research focuses on S&T strategic planning, innovation policy, and think-tank theory and method research. E-mail: jfpan@casisd.cn

■ 责任编辑：岳凌生

*Corresponding author