



科学学研究  
*Studies in Science of Science*  
ISSN 1003-2053, CN 11-1805/G3

## 《科学学研究》网络首发论文

题目：元宇宙的主要特征、社会风险与治理方案  
作者：王文玉  
DOI：10.16192/j.cnki.1003-2053.20221102.002  
收稿日期：2022-06-30  
网络首发日期：2022-11-03  
引用格式：王文玉. 元宇宙的主要特征、社会风险与治理方案[J/OL]. 科学学研究.  
<https://doi.org/10.16192/j.cnki.1003-2053.20221102.002>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

## 元宇宙的主要特征、社会风险与治理方案

王文玉<sup>1,2</sup>

(1. 重庆大学法学院, 重庆 400044; 2. 河南工业大学法学院, 河南郑州 450000)

**摘要:** 代表最先进信息技术综合集成趋势的元宇宙具有虚实相融性、自治性、社会性三个主要特征。元宇宙的出现会引发社会生产结构、运行机制、关系状态的联动变革, 在变革中也将引发用户数据隐私被泄露、元宇宙平台权力被滥用、用户个体意识被操控、歧视风险加剧、意识形态安全无法保障等社会风险。元宇宙社会风险的治理方案应当遵循系统化原则和多元共治理念。在伦理方案上, 应当坚持以人为本、在元宇宙社区内设立伦理审查机构、开展价值敏感设计和价值熔断设计; 在法治方案上, 应当及时立法以澄清元宇宙相关法律争议、通过公权力监管和保障用户权利的方式规范元宇宙平台权力的行使; 在技术方案上, 应当通过以代码规制代码、以技术治理技术的方式优化风险应对方案的内嵌性和可行性。

**关键词:** 元宇宙; 社会风险; 多元共治; 虚实相融

**中图分类号:** TP391.9 **文献标识码:** A

随着 2021 年元宇宙第一股、游戏公司 Roblox 上市以及著名社交网络公司 Facebook 更名为 Meta 等事件的发生, 元宇宙这一概念迅速引发全球关注。此后, 微软、亚马逊、百度、字节跳动等网络巨头争相跟进, 元宇宙一夜间成为了互联网投资的新风口和信息技术发展的新趋势。代表最先进信息技术综合集成趋势的元宇宙掀起席卷全球的热潮有着一定的必然性: 新冠疫情下, 人们的学习、工作、社交、娱乐等活动被迫迁移到虚拟空间之中, 从而为元宇宙的形成提供了外部催动力。大数据、区块链、扩展现实、脑机接口、人工智能、云计

**收稿日期:** 2022-06-30; **修回日期:** 2022-09-08

**基金项目:** 司法部专项任务课题“数字治理法治化研究”(21SFB4004)

**作者简介:** 王文玉(1992-), 男, 河南商丘人, 讲师, 博士研究生, 研究方向为数字法治。E-mail: wangwenyu199201@163.com。

算等技术不断发展成熟并产生群聚效应（Critical Mass）则为元宇宙的形成提供了内在技术支撑。然而，当前无论是产业界还是学界并未就元宇宙的内涵、外延、社会风险、发展前景等得出一致结论，这使得拥有变革人类生活状态、生产方式、认知格局潜力的元宇宙的未来充满了不确定性风险。因而我们有必要遵循“技术社会风险反思的先导性原则”，在元宇宙还处于初始探索阶段的背景下，对元宇宙的主要特征以及可能引发的社会风险展开梳理和分析，并努力提出具有前瞻性、可行性的治理方案，从而为元宇宙的安全、有序、高质量、可持续发展提供保障。

## 1 元宇宙的主要特征

元宇宙并不是全新的概念，早在 1992 年，尼尔·斯蒂芬森在其科幻小说《雪崩》中就提出了元宇宙（Metaverse）的概念。然而直到近年来，相关信息技术集群取得突破性发展后，能够将现有通讯技术、人工智能技术、扩展现实技术、区块链技术、脑机接口技术等先进数字信息技术予以集成创新、深度融合、紧密嵌套的元宇宙概念才被重提和重视。综合现有学者的总结和论述，元宇宙往往被视为基于现今数字信息技术，通过对现实世界的数字化改造搭建出的融合了经济系统、社交系统、身份系统的，能够为用户提供沉浸式体验的，并且是开放、允许每个用户进行内容生产和编辑的特殊空间形态<sup>[1]</sup>。虽然作为新兴互联网应用平台的元宇宙的内涵和外延仍处于一种不断变动的状态，但我们可以通过对元宇宙主要特征、发展趋势的总结，从而为元宇宙提供一个相对稳定的、具象化的认知描述。关于元宇宙的主要特征，诸多学者和公司都做出过探讨，如游戏公司 Roblox 在招股说明中就总结了元宇宙的八个主要特征，它们是身份性、互动性、沉浸性、场景多样性、低延迟性、内容丰富性、经济性和安全性<sup>[2]</sup>。结合现有探索和论述，本文认为元宇宙具有虚实相融性、自治性、社会性三个主要特征。

### 1.1 虚实相融性

元宇宙本质上是通过数字技术建构的虚拟空间，但和原有互联网虚拟空间的高度独立性不同，元宇宙是虚拟空间和现实世界深度融合的产物，对现实世界的映射以及和现实世界之间的交互性是元宇宙的重要特征。其一，元宇宙的基础形态来自于通过数字孪生技术对现实世界的映射，其是在对现实世界数字镜像移植的基础上衍变和延伸而来的。将现实世界数字化，并通过知识提取、空间孪生、场景投射等方式再造现实世界的镜像世界是打造元宇宙世界的第一步。因而，元宇宙世界并不是和现实世界完全平行或对立的，“其可以被视作现实世界在虚拟世界中的映射或翻版”<sup>[3]</sup>。其二，元宇宙是和现实世界深度交互和耦合的，其通过区块链技术搭建的经济体系能够实现与现实世界在经济、社交、身份系统上的融合。如在元宇宙游戏第二人生（The Second Life）中流通的林登币（Linden Dollar）就可以通过浮动汇

率实现和美元的直接兑换。可以预见，随着区块链技术以及以区块链技术为基础的通证技术的发展和大规模使用，元宇宙和现实世界之间的商品交易、资金流动等将变得更加便捷、安全和频繁<sup>[4]</sup>。其三，元宇宙的存在目的和最终意义仍然是为现实世界的人服务的。通过虚拟现实（VR）、增强现实（AR）、扩展现实（XR）等技术手段的运用，元宇宙可以摆脱现实世界的社会结构、物理规律、生理特质、资源种类等限制，为用户提供满足其身份、能力、自由、感官、想象力等需求的内容，从而能够丰富人的情感体验、扩展人的感知阈值、延伸人的创造能力、提升人的价值内涵。

## 1.2 自治性

元宇宙是一个去中心化的组织系统，在理想的元宇宙之中，各类主体可以摆脱现实世界经济和身份的限制、法律法规和道德伦理的约束以及中心权力的管控，从而获得高度的平等性和自主性。由于在元宇宙中，相关平台仅提供基础的开发架构，而不提供具体的产品内容，其需要用户利用开源代码开展社会共建、边界开拓、产品塑造、协议制定等具体活动，这就使得元宇宙平台和以往的互联网平台产生了差异性：用户在一定程度上获得了开发者的地位，其不但可以自主控制数据的生产活动，而且还拥有了资源开发、内容管理、规则制定和分配收益的权限<sup>[5]</sup>。

如，作为元宇宙底层技术的区块链技术的应用就会强化元宇宙高度自治和去中心化的特质。区块链技术通过确证数据的公开性、不可篡改性、可靠性，从而为元宇宙带来了“去中心化身份控制机制”（DID）：一方面，基于链上协议生成的“非同质化通证”（NFT）保障了用户在无需第三方认证和维护的情况下，就可以实现对不可拆分、不可复制、具有唯一性的数据资产绝对、排他的占有权利。另一方面，通过智能合约这一元宇宙主要决策执行机制，元宇宙用户还可以在没有任何第三方监督和强制的情况下，展开具有自主性的交易活动，从而保障了复杂社区自组织体的顺利运转。

## 1.3 社会性

不同于仅仅将互联网空间视为信息存储和传递工具的屏幕阅读时代，web3.0 时代元宇宙空间的功用中心已经由信息变为了人们在其中的虚拟化身，在元宇宙空间内活动的不再是一个被操控的对象化的角色，而是在不同虚拟角色间的交往中获得主体间性的角色，这个角色就代表着玩家或者“AI 数字原著人”。被赋予了主体性的元宇宙角色之间可以通过诸如合作、交往、交易等行为而形成除血缘、地缘、业缘之外的第四种社会关系，即虚拟体之间的社会关系。元宇宙社会关系本体论就是这种虚拟体之间在交往和实践中形成的虚拟体的社会关系形态，其既是元宇宙区别于传统技术工具性定位的重要特质，也意味着元宇宙建构和重塑了区别于现实社会但又与现实社会相互嵌套的社会关系新体系<sup>[6]</sup>。

虚拟性决定了元宇宙中的角色可以超越现实社会的身份、地位、等级、种族等限制，而建构出一套新的社会交往伦理准则、文化规范以及行为框架。摆脱了现实身份的限制，意味着元宇宙社会关系的形成会更加纯粹，玩家之间的交往可以基于共同的旨趣或利益，社区秩序规则的形成也更多地依靠全体玩家的民主表决。但元宇宙社会和现实社会相互嵌套的关系又决定了玩家的社会交往无法完全脱离现实社会的知识体系、价值理念与交往规则的枷锁，现实世界中的秩序意识、话语结构和信念认知都会被带入和移植到元宇宙的交往行为中。

## 2 元宇宙的社会风险

历史上的每一次技术革命都会引发社会变迁、权力变动和伦理革新，元宇宙的出现无疑也会引发社会生产结构、运行机制、关系状态的联动变革，在变革中新的社会风险也将随之产生。虽然元宇宙创造了超越现实世界的虚拟世界，但由于元宇宙的现实性和社会性以及元宇宙的主要行为主体仍然来自于现实世界，因而其所引发的社会风险并不是一个独立的存在论风险，而是仍然嵌套或溢出于现实社会的风险<sup>[7]</sup>。对此，我们有必要在元宇宙的发展还处于技术探索阶段，对相关社会风险展开前瞻性的梳理和反思，从而为推进元宇宙的良性、长远发展奠定基础。

其一，作为元宇宙底层架构技术的区块链技术的应用和用户数据隐私保护之间存在一定的冲突。由于区块链在技术平台、应用场景、协作机制等方面为元宇宙提供着身份认证体系、资产认证和交易体系、货币加密体系等技术参考模型的核心架构，因而被视为元宇宙的核心基础设施<sup>[8]</sup>。但区块链技术的应用和用户个人隐私之间存在一定的冲突：为了保障去中心化平台中相关数据的可靠性和可信赖性，区块链技术采取将链上节点信息向所有节点公开、并接受所有节点验证的方式对数据信息展开校验、存储和维护，从而解决了元宇宙缺失第三方中心机构的信任难题<sup>[9]</sup>。但当元宇宙用户的财产、交易等信息被收集和存储于区块之中后，这一信息也将会被其它参与验证节点共享，这也就意味着对用户而言，其个人信息对链上其它节点是透明的，进而用户的交易数据、账户信息、身份地址、存储的文件照片等个人数据信息都将面临被公开和泄露的风险。

除去数据收集、存储和认证过程中的风险，智能合约以及跨链操作等区块链技术在元宇宙中具体落地的相关应用也都可能产生用户数据隐私被泄露的风险。基于区块链技术而提出的智能合约被视为元宇宙的主要决策执行机制，但当用户发起调用智能交易合约指令后，区块链中的相关节点都会对这一交易展开认证和处理，此时和交易相关的操作流程以及数据信息便会对所有认证节点开放，而这一交易认证方式便存在诸如用户的身份地址、财产流向、投票方案等敏感信息被泄露的风险<sup>[10]</sup>。又如，为了实现不同元宇宙之间的沟通互联，需要利用跨链技术以完成不同元宇宙数据链之间的价值传递。但由于不同的链之间在系统架构和隐



私保护方式上存在差异，因而利用跨链技术对同构或异构的链之间进行数据传递和交换时，便可能引发用户隐私数据被跨链泄露的风险。

其二，通过技术和资本相结合的方式成为元宇宙权力中心的元宇宙平台存在滥用权力、侵害用户正当权益、压缩用户权利空间的风险。有学者基于元宇宙去中心化的特质认为元宇宙平台以一种精简化、高效化甚至无人化的制度体系取代了现实社会中复杂的信用担保和臃肿的官僚体系，从而有效推动了社会制度体系更加高效和平等地运转<sup>[11]</sup>。但由于元宇宙具有现实性、公共性和社会性，因而元宇宙的去中心化实质上却是一个权力中心的迁移和再造过程：元宇宙的基本秩序和发展预期需要开发者和建设者予以设计和明确；用户在利用开源代码和区块链开展活动之前，还需要元宇宙平台提供代码开源标准、社区治理结构、争议解决方案等底层架构内容；随着更多关键的组织机构入场，元宇宙平台会成为新的影响用户活动的核心中介；出于保障数据安全、提高数据利用效率的目的，元宇宙平台也将通过控制数据中心的方式垄断数据资源。

在元宇宙权力中心再造的过程中，元宇宙平台由于获得了更多行为规则制定、资源分配、秩序维护、个体行为控制等方面的支配能力，从而成为了元宇宙的实际秩序维护者和资源利益分配者。但不同于传统公权力经由民众授权，并以保障社会公平、正义等公共利益的实现为最终目的的特质，元宇宙平台的权力是通过资本和技术赋权而获得的，其最终目的是为了商业利润，这就使得这一权力获得和行使面临着正当性和公正性的质疑：由于拥有远超用户的技术优势、信息优势、资本优势，并且还能够以虚拟空间隔离现实法律、以技术黑箱隔离第三方监管、以自动决策隔离担责风险，元宇宙平台在权力的行使中便可以充分利用技术赋权和约束体系严重的不对称性，尽可能扩大和用户之间的权力势差。在权力扩张法则以及利益最大化需求的驱动下，元宇宙平台会充分利用其权力在实际运作中的隐蔽性、自动性、弥散性、脱域性和垄断性特质，打破现有限制公权力、保障私权利的权力制衡体系，努力营造出“权力—支配”格局，从而在外部竞争中通过设立交易壁垒、阻碍用户迁移等方式限制其它元宇宙的发展；在内部管理中通过全面提升准入限制、交易税率、虚拟资产价格等方式压缩用户的权利空间。例如，用户不仅是元宇宙的消费者，更是元宇宙空间内容的生产者，其在诸如建筑构造、游戏开发等模组的创作中所耗费的智力、精力、财力等往往会在消费者身份定位的遮蔽下被元宇宙平台无偿占有并用来盈利，最终，相关机构利用信息和管控能力的不对称性可能会将元宇宙异化为资本剥削的新媒介和新平台，用户则沦为了被平台任意剥削的“玩工”（Playbour）<sup>[12]</sup>。

其三，元宇宙可能引发用户个体意识被操控，进而危及人类自主性的风险。个人的自主性一直被认为是最基本的人权之一，其是理性的个人获得尊严的前提。控制自己的意识、按照自己的意愿做出决定、独立地做出选择的权利和自由是几乎其它所有自由的基础<sup>[13]</sup>。而作为人脑和元宇宙主要连接通道的脑机接口技术的应用则对人类意识的自主性和思想的自由

性提出了挑战：一方面，为了提升元宇宙的吸引力以及更多地营销元宇宙内的商品，元宇宙平台会利用脑机接口接收和分析用户大脑的神经信号并绘制出反应个体情感波动和审美体验的“神经地图”，从而精准地向用户提供可以刺激其情感愉悦和审美偏好的内容，最终实现对用户的不当诱导和隐性操控<sup>[14]</sup>。例如，神经营销学逐渐受到实践重视，它们可以利用技术直接或间接捕捉神经信息中隐藏的个人偏好以及个体的情绪状态，从而通过调换音乐、更换阅读内容、改变照明环境等方式实现对用户情感状态的调动和操控<sup>[15]</sup>。另一方面，更为严重的是，新一代“脑深部电刺激”（Deep Brain Stimulation）技术还可以利用电极刺激脑神经电波的方式，直接对个体的记忆、情感、意向等展开生理性改良和操控。这一原本用于治疗精神疾病的脑机接口技术在元宇宙中的应用存在着改变使用者个性特征、行为目标和情绪反应进而威胁个体的能动性、自主性以及精神完整性的风险。而当个体的行为和思想被外力操纵和扭曲，无法按照自己的生理喜好和动机生活时，也就意味着其将会丧失自决和尊严，沦为技术操控者的“奴隶”<sup>[16]</sup>。

其四，由于用户可以在元宇宙空间内自由编订和更换角色的性别、肤色、年龄、种族、社会阶层等可能引发歧视的个体特征，因而元宇宙的出现被认为是重塑用户之间平等身份的新契机。但实际上，元宇宙的出现不但无法使用户摆脱身份歧视的问题，还可能会进一步加剧数字弱势群体（Data Vulnerable Groups）的不利处境。一方面，由于元宇宙的基础架构来自于对现实世界的映射，元宇宙的参主体主要是现实世界的用户，因而现实世界的歧视问题也会被映射入元宇宙之中。例如，研究表明，在元宇宙游戏第二人生（The Second Life）中，现实生活中的沟通机制、行为规则和制裁方案仍然会被玩家沿用，将其做为虚拟社区中展开沟通交往和社会控制的主要方案<sup>[17]</sup>。如此，现实世界沟通、交往规则中内涵的歧视倾向自然也会被元宇宙用户所践行，甚至在虚拟身份的掩盖下，个体反而可能会更加肆无忌惮地将原有的歧视态度和倾向付诸实践<sup>[18]</sup>。另一方面，元宇宙还存在进一步加剧数字弱势群体不利处境的风险。由于技术的高度复杂性、先进性、集成性，元宇宙不但对相关数字基础设施，而且对参与用户的经济能力、技术认知能力、数字学习能力和数字敏感度都提出了较高的要求，这无疑会进一步加剧个体间的“纵向数字鸿沟”<sup>[19]</sup>。无法担负元宇宙进入成本、缺乏元宇宙参与能力的用户将会被排斥于元宇宙之外；数字生产能力、经济能力不足者将无法全面、深度参与元宇宙世界的设计；数字信息获取能力不足者难以及时、充分知悉关涉用户发展权利的相关事项；少数族群在元宇宙空间规则的制定中被忽视等都是元宇宙时代数字弱势群体需要面临的不平等风险。

其五，元宇宙还存在危及意识形态安全的风险。元宇宙具有鲜明的自由、民主特质，无论是社区交易、规则制定、政治决议、信息流动等都是全体用户在技术的辅助下自主、独立、自由地完成的。但这一自由的背后，潜伏着少数元宇宙平台借助元宇宙的脱域性以及数字自由、数字民主之名，利用元宇宙社会和现实社会嵌套的特质，对主权国家的意识形态展开解

构的风险。如少数元宇宙平台可能利用元宇宙不受国界限制的特质，创造出一个剥离了用户对民族国家的历史、文化、价值观认同感的，由数据想像共同体构成的，契合其政治目的和经济利益的数字帝国<sup>[20]</sup>。受元宇宙空间内信息茧房、回音室效应的影响，一旦元宇宙中的数据历史、数据价值观、数字文化建构成熟并获得用户的内在认同，那么用户的自主判断能力将会受到严重干扰，甚至在平台的诱导下还会对现实国家的主流意识形态产生排斥和质疑。例如，有学者在社交平台 Facebook（现更名为 Meta）上展开的一场关于 68 万多人的实验结果就表明，研究者甚至都不需要和实验对象进行互动以及做出任何非言语暗示，而只需控制用户在 Facebook 上所能够接触到的朋友表达情绪的贴文，就足以达成其所想要的引导对象情绪的目的<sup>[21]</sup>。一旦元宇宙平台对元宇宙内信息资源的支配能力被用来输出特定的价值观，那么其将会对现实国家的主流意识形态安全产生极大的冲击。

### 3 元宇宙社会风险的治理方案

面对智能社会的加速到来，我们将不得不适应在虚拟数字社会和现实物理社会的双重空间内交叉进行学习、工作、社交、娱乐等活动的局面。面对智能社会中生活秩序的重塑以及社会风险的叠加，运用单一的法律或伦理等手段展开风险治理的周全性、可行性、有效性将面临挑战。对此，莱斯格就指出，对于智能社会中网络空间的规制而言，我们应当建构起多元规制体系，探索形成将社会规范、法律、市场和架构相统一的多元规制框架<sup>[22]</sup>。在借鉴相关学者研究的基础上，本文认为，对于元宇宙社会风险的治理，我们应当遵循系统化原则和多元共治理念，以“公权力—平台权力—私权利”相平衡的权力制约思路为指引，沿着元宇宙空间的内部运行控制和外部影响控制相结合的双重路径，探索出“伦理—法律—技术”共治的风险治理方案，从而破解元宇宙社会风险的系统性、复杂性、交叉性难题，保障元宇宙能够朝着稳定、可预期的方向良性发展。

#### 3.1 伦理方案

其一，坚持以人为本，防范人的主体性被消解。科技发展的最终目的是为了地提升人类福祉，正如爱因斯坦所言，“我们发展所有技术的主要奋斗目标就是关心人的本身”<sup>[23]</sup>。在人和元宇宙相关技术的关系中，人类始终是主体和目的，相关技术则应当被定位为实现目的的工具。因而，元宇宙的一切运行机制都应当服务于人的目的，满足人的需求，始终处于人类的可控范围之内。为了避免人类的主体性被消解，元宇宙的发展应当始终遵循以人为本、技术向善的要求，对于那些可能会损害人的尊严和自由意志的技术，则应当严格审查甚至予以禁止。例如对于可能会消解人类自主性、危及人类尊严的脑机接口技术尤其是反向神经刺激相关技术的使用，我们就应当予以严格限制甚至禁用：只允许其在医学领域内的有限应用，而对于商业领域，则应当对相关技术的使用予以禁止<sup>[24]</sup>。



其二，在元宇宙社区内设立伦理审查机构从而对元宇宙的运行展开伦理审查。由于元宇宙和现实世界的隔离性，相比外在监管的方式而言，在元宇宙内部设立相关伦理审查机构对于及时发现、全面预防和有效应对相关伦理风险具有积极意义。在伦理审查机构主体的构成上，可以事先建立包括政府机构代表、元宇宙相关平台代表、社会组织代表和民众代表组成的统一伦理审查委员会，而后再根据不同元宇宙类型增加相关专业技术代表、元宇宙社区成员代表，从而保障审查机构的专业性和代表性。在职能上，伦理审查机构既需要负责制定符合不同类型元宇宙特质的基础伦理准则，还需要审查元宇宙平台的管理规则、技术应用规范、交易规则、社区自治规则等技术和秩序规范是否契合基本伦理要求。对于审查启动的方式而言，应当通过清单列举的方式确立起关键内容的备案审查制度。此外，伦理审查机构还应当主动开展日常巡视审查，积极接受元宇宙用户、相关机构提出的审查申请，对有争议或可能引发社会风险的相关技术的应用、内容的设计以及决议的形成等展开伦理审查<sup>[25]</sup>。

其三，通过价值敏感设计和价值熔断设计的方式保障元宇宙的运行符合社会基本伦理价值准则<sup>①</sup>。价值敏感设计（Value Sensitive Design）是指借助技术手段将社会公众和元宇宙用户形成的基本价值共识嵌入到元宇宙的设计、开发、建设、维护之中，从而为元宇宙锚定基础价值，保障其运行符合人类基本价值准则。价值敏感设计主要遵循三个步骤：在概念研究层面，需要通过对元宇宙运行中的主要伦理价值如平等、安全、自由等展开识别与判断，以确保元宇宙的设计和运行不偏离这些核心价值；在经验研究层面，通过实验、调研、测试等定性和定量方式针对不同类型的元宇宙以及不同类型的技术应用展开科学评估，以确定具体场景下的价值排序；在技术设计层面，则需要相关设计者对不同价值嵌入方式展开对比，从而选出最优设计方案<sup>[26]</sup>。价值熔断设计也是保障元宇宙的运行符合社会伦理准则的重要方式。为了避免元宇宙的设计或相关决策出现危及社会基本正义理念和民众基本权利的风险，我们可以在元宇宙的运行、相关技术的应用以及元宇宙社区自治协议的形成等过程中的核心节点设置“伦理安全阀”，当相关设计或决议偏离或违背诸如社会核心价值理念、用户决策自主性准则、人类共同福祉目标等基本价值设定时，便会因无法满足触发条件而被中止。此时，相关决议将会自动交由政府监管部门展开审查，从而及时阻断重大社会风险发生的可能<sup>[27]</sup>。

### 3.2 法治方案

法治方案是防范元宇宙平台权力异化、维护用户正当权益、推动元宇宙朝向有序化、规范化发展的重要保障。由于现实世界是元宇宙的母体，元宇宙的主要发起者和建设者都是现

---

<sup>①</sup> 虽然本文将相关应对方案依照重点关注方向的不同而划分为了伦理方案、技术方案和法律方案，但这三种方案之间实际上是相互融合、交叉而不是相互独立的。例如伦理嵌入方案是伦理方案和技术方案的结合、法律代码化的方案是法律方案和技术方案的融合，这也充分体现了本文所希望实现的三种方案共治的系统化风险应对理念。

实社会中的人或组织，因而通过立法、政府规制、权力制约等方式规范元宇宙的发起者、建设者的行为，从而避免元宇宙社会风险的产生是法治方案的主要目标。其一，元宇宙的出现引发了诸多新的法律问题，对于这些问题亟需通过立法的方式予以明确，如此，才能将元宇宙的发展纳入法治的确定性轨道之中，有效避免因规则缝隙而引发的社会风险。从立法方式上来说，一些问题可以通过融入数字平台立法规制的进程之中予以解决：如元宇宙相关数据的归属和保护就可以在数据立法进程中予以规定；元宇宙平台铸币权问题可以在虚拟货币定位和规制的立法中予以明确。还有一些专属于元宇宙引发的立法问题，则可以采取类型立法、地方立法、实验立法等方式展开规制，待元宇宙发展定型以及相关立法成熟之后，再制定统一的法律规制方案。如对于元宇宙最重要组成部分的去中心化自治组织的法律规制问题，美国的怀俄明州就通过制定《怀俄明州分布式自治组织法案》的方式，将去中心化自治组织 DAO 定义为有限责任公司，从而确定了这一组织的法律地位，保障了这一组织的顺利运转以及相关纠纷的顺利解决<sup>[28]</sup>。

其二，通过以公权力监管平台权力，以用户权利制约平台权力的方式，推动元宇宙中公权力、平台权力和用户权利之间的平衡发展。为了防范元宇宙平台攫取过多权力、逃避合法监管、侵害用户权利的风险，我们可以通过打造新的权力制约体系的方式，将平台权力关在制度的笼子里。一方面，应当强化公权力对平台权力的监管，从而界定元宇宙平台权力的界限。我们可以通过政府和元宇宙平台共同制定元宇宙的基础性规则、政府助力技术攻关和技术标准建设、政府推动元宇宙基础设施完善等方式提升政府在元宇宙建设、引导、监督中的参与度，从而应对元宇宙平台滥用垄断权力可能会带来的诸如操控意识形态、泄露机构和用户数据信息、侵害用户自主权、歧视数字弱势群体等具有全局性、易发性和普遍性的风险。另一方面，还应当通过积极赋权用户的方式，以制衡元宇宙平台权力。赋权元宇宙用户是一项复杂的、系统性工程：我们应当通过立法的方式以明确用户在元宇宙中的创作收益权、数据隐私权、知情权、决策自主权等权利内容；通过加强信息基础设施建设和元宇宙技术的用户友好程度等以保障数字弱势群体的平等参与权；通过强化宣传、教育等方式帮助用户了解自身权利内容，提升用户维权的积极性；通过畅通元宇宙内部和外部救济渠道，从而保障用户权益能够得到及时救济。

### 3.3 技术方案

随着技术应用对社会行为、社会关系的影响日益深刻，技术哲学出现了价值论转向，技术不再单单作为中立性的生产工具，其开始负载起更多的社会调整、风险治理、价值导向功能。对于元宇宙这一信息技术汇聚而成的平台而言，通过技术设计和技术应用的方式，以落实相关伦理、法律治理方案，提升元宇宙社会风险应对方案的内嵌性和可行性无疑十分必要。因而，一方面，我们应当充分重视代码在元宇宙社会风险治理中的作用，树立代码即法律的观念，积极将伦理准则和法律规则转译为代码，从而以代码规制代码，以源头控制和过程控

制的方式化解元宇宙可能产生的社会风险。依赖数字技术架构而成的元宇宙实质上是一个代码构成的世界，无论是平台垄断利益的获得、对用户自主性的操控，还是平台和用户歧视意图的实现往往都是由代码架构完成的。对此，我们可以采取代码内嵌的方式，借助算法和自动化决策等新技术，实时评估相关代码的运行可能给社会带来的风险，从而穿透元宇宙相关决策形式上的自动性和同等对待性，直指相关决策行为背后的垄断、操控、歧视等意图，进而参与、干涉和改变主体的决策过程，及时调整可能引发风险的行为<sup>[29]</sup>。

另一方面，我们还应当注重对数字技术的应用，从而以技术治理技术，防范元宇宙运行过程中可能引发的社会风险。例如，对于区块链相关技术应用可能造成的用户数据隐私被泄露的风险，研究人员提出了基于加密技术和基于混币机制的数据隐私保护方案：加密货币 **Zerocash** 就利用零知识证明这一基于同态加密技术而形成的区块链隐私技术确保了分类账上只显示是否存在交易，而交易的细节如金额、双方身份信息等都是隐藏的<sup>[30]</sup>。中心化混币机制 **Mixcoin** 则可以利用中心节点将多笔交易打乱和重新分配，从而避免了攻击者根据聚类分析和时间分析而判断出参与者资金流向的风险<sup>[31]</sup>。又如，通过使用风险感知挖掘技术，可以帮助监管机构发现纷繁复杂元宇宙系统内可能存在的诸如身份歧视、平台垄断、意识形态解构等风险，从而提示监管机构及时展开治理，避免风险的发生或扩大。实践中，广泛应用的“感知歧视的数据挖掘技术”（**DADM**）就能够以自动化运行的方式检测出在元宇宙中运行的相关算法是否存在身份歧视等有违公正伦理要求的情形，从而发出预警，有效化解元宇宙中潜在的歧视风险<sup>[32]</sup>。

## 参考文献：

- [1] 人民论坛“特别策划”组. 元宇宙的当下与未来[J]. 人民论坛, 2022(7): 14-15. “Special Planning” Group of People’s Tribune. The present and future of metaverse[J]. People’s Tribune, 2022 (7) : 14-15.
- [2] Roblox Corporation IPO Investment Prospectus Form S-1/A[EB/OL]. <https://sec.report/Document/0001193125-21-049767>.
- [3] 李黎玥. 元宇宙：虚拟与现实的高度融合 [N]. 北京日报，2021-11-24(13). Li L Y. Metaverse: High integration of virtual and reality [N]. Beijing Daily , 2021-11-24(13).
- [4] 陈永伟，程华. 元宇宙的经济：与现实经济的比较 [J]. 财经问题研究，2022 (5) : 3-16. Chen Y W, Cheng H. Meta-economy: A comparison with real economy [J]. Research on Financial and Economic Issues, 2022 (5) : 3-16.

- [5] 陈辉, 闫佳琦, 陈瑞清, 沈阳. 元宇宙中的用户数据隐私问题[J]. 新疆师范大学学报(哲学社会科学版), 2022(5): 107-115. Chen H, Yan J Q, Chen R Q, Shen Y. Privacy of user's data in metaverse[J]. Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences), 2022(5): 107-115.
- [6] 蓝江. 元宇宙的幽灵和平台—用户的社会实在性[J]. 华中科技大学学报(社会科学版), 2022(3): 10-17. Lan J. The ghosts and platforms of metaverse—The social reality for users[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2022(3): 10-17.
- [7] 赵汀阳. 假如元宇宙成为一个存在论事件[J]. 江海学刊, 2022(1): 27-37. Zhao T Y. If the metaverse becomes an ontological event[J]. Jianghai Academic Journal, 2022(1): 27-37.
- [8] 李鸣, 张亮, 宋文鹏, 吴美玉. 区块链: 元宇宙的核心基础设施[J]. 计算机工程, 2022(6): 24-32. Li M, Zhang L, Song W P, Wu M Y. Blockchain: Core infrastructure of metaverse[J]. Computer Engineering, 2022(6): 24-32.
- [9] Galloway A R. Protocol: How control exists after decentralization [J]. Journal of the American Society for Information Science and Technology, 2005, 57(1): 146-147.
- [10] Steffen S, Bichsel B, Gersbach M, et al. Zkay: Specifying and enforcing data privacy in smart contracts[C]. Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019: 1759-1776.
- [11] Kondova G, Barba R. Governance of decentralized autonomous organizations [J]. Journal of Modern Accounting and Auditing, 2019, 15(8): 406-411.
- [12] Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata[J]. Ethics and Information Technology, 2004, 6(3): 175-183.
- [13] Sententia W. Neuroethical considerations: Cognitive liberty and converging technologies for improving human cognition[J]. Annals of the New York Academy of Sciences, 2004(1013): 221-228.
- [14] Kontson K, Megjhani M, Brantley J A, et al. Your brain on art: Emergent cortical dynamics during aesthetic experiences[J]. Frontiers in Human Neuroscience, 2015(9): 1-17.

- [15] Andujar M, Crawford C S, Nijholt A, et al. Artistic brain-computer interfaces: The expression and stimulation of the user's affective state[J]. Brain-Computer Interfaces, 2015, 2 (2) : 60-69.
- [16] Baylis F. "I am who I am" : On the perceived threats to personal identity from deep brain stimulation[J]. Neuroethics, 2013 (6) : 513-526.
- [17] Gabriels K, Bauwens J, Verstrynge K, et al. Virtual Worlds and Metaverse Platforms[M]. Hershey: IGI Global, 2012: 306-320.
- [18] Bailenson J N, Nass C, Groom V. The influence of racial embodiment on racial bias in immersive virtual environments[J]. Social Influence, 2009, 4 (3) : 235-240.
- [19] 单勇. 跨越“数字鸿沟”: 技术治理的非均衡性社会参与应对[J]. 中国特色社会主义研究, 2019(5):68-75. Shan Y. Crossing the "digital divide": The response of unbalanced social participation in technological governance[J]. Studies on Socialism with Chinese Characteristics, 2019 (5) : 68-75.
- [20] Owen T. Disruptive Power: The Crisis of the State in the Digital Age[M]. New York: Oxford University Press, 2015: 1-21.
- [21] Kramer A D I, Guillory J, Hancock J T. Experimental evidence of massive-scale emotional contagion through social networks[J]. Proceedings of the National Academy of Sciences, 2014, 111 (24) : 8788-8790.
- [22] 劳伦斯·莱斯格. 代码 2.0: 网络空间中的法律[M]. 北京: 清华大学出版社, 2018: 89. Lawrence L. Code: And Other Laws of Cyberspace, Version 2.0[M]. Beijing: Tsinghua university press, 2018: 89.
- [23] 爱因斯坦. 爱因斯坦文集第 3 卷[M]. 北京: 商务印书馆, 1979: 73. Einstein. The Collected Works of Einstein Volume 3[M]. Beijing: The Commercial Press, 1979: 73.
- [24] Sandel M J. The Case against Perfection: Ethics in the Age of Genetic Engineering[M]. Cambridge: Harvard University Press, 2007: 5-9.
- [25] 王文玉. 算法嵌入政府治理的优势、挑战与法律规制[J]. 华中科技大学学报(社会科学版), 2021 (4) :26-36. Wang W Y. The advantages, challenges and legal regulations of algorithm embedding in government governance[J]. Journal of Huazhong University of Science and Technology (Social Science Edition) , 2021 (4) :26-36.



- [26] Cummings M L. Integrating ethics in design through the value-sensitive design approach[J]. Science and Engineering Ethics, 2006, 12 (4) : 702-704.
- [27] 柳亦博. 人工智能阴影下: 政府大数据治理中的伦理困境[J]. 行政论坛, 2018 (3) : 97-103. Liu Y B. Under the shadow of artificial intelligence: Ethical dilemma in government big data governance[J]. Administrative Tribune, 2018 (3) : 97-103.
- [28] 程金华. 元宇宙治理的法治原则[J]. 东方法学, 2022 (2) : 20-30. Cheng J H. The rule of law principle of metaverse governance[J]. Oriental Law, 2022 (2) : 20-30.
- [29] 金韬. 歧视可以“间接”吗? [J]. 法制与社会发展, 2021 (6) : 51-67. Jin T. Can discrimination be “indirect” ? [J]. Law and Social Development, 2021 (6) : 51-67.
- [30] Sasson E B, Chiesa A, Garman C, et al. Zerocash: Decentralized anonymous payments from bitcoin [C]. Proceedings of the 2014 IEEE Symposium on Security and Privacy. Berkeley: IEEE, 2014: 459-474.
- [31] Valenta L, Rowan B. Blindcoin: Blinded, accountable mixes for bitcoin [C]. Proceedings of the International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2015: 112-126.
- [32] Binns R. Algorithmic accountability and public reason[J]. Philosophy & Technology, 2018, 31 (1) : 543-556.

## **Main features, social risks and governance schemes of the metaverse**

WANG Wen-yu<sup>1,2</sup>

(1. Law School, Chongqing University, Chongqing 400044, China;

2. Law School, Henan University of Technology, Zhengzhou 450000, China)

**Abstract:** The metaverse, which represents the comprehensive integration trend of the most advanced information technology, has three main characteristics: the fusion of the virtual and the real, the autonomy and the sociality. The emergence of the metaverse will trigger a linkage change in social production structure, operating mechanism, and relationship status. During the transformation, it will also lead to social risks such as the leakage of user's data privacy, the abuse

of power of the metaverse platform, the manipulation of individual consciousness, the increased risk of discrimination, and the ideological Safety cannot be guaranteed . Governance schemes of the social risks should follow the principle of systematization and the concept of multi-governance. In the ethical plans, we should adhere to people-oriented, establish an ethical review organization within the metaverse community, and carry out value sensitive design and value circuit breaker design; in terms of the rule of law plans, legislation should be made timely to clarify legal disputes, and the power of metaverse platform should be regulated by means of public power supervision and protection of user's rights; in terms of technical plans, we could optimize the embeddedness and feasibility of risk responses by means of regulating code by code and govern technology by technology.

**Key words:** metaverse; social risk; pluralistic co-governance; fusion of virtuality and reality