



中国人工智能学会
Chinese Association for Artificial Intelligence

中国人工智能学会系列白皮书 ——元宇宙技术(2025版)

中国人工智能学会
二〇二五年十二月



中国人工智能学会系列白皮书 ——元宇宙技术（2025版）

中国人工智能学会
二〇二五年十二月

《中国人工智能学会系列白皮书》编委会

主 任：戴琼海

执行主任：马华东

副 主 任：赵春江 何 友 王恩东 郑庆华 刘成林 周志华 孙富春

庄越挺 胡德文 杜军平 杨 强

委 员：陈松灿 董振江 付宜利 高新波 公茂果 古天龙 何 清

胡清华 黄河燕 季向阳 蒋田仔 林浩哲 梁吉业 刘奕群

潘 纲 石光明 孙茂松 孙长银 陶建华 王海峰 王熙照

王 轩 王蕴红 吴 飞 于 剑 余有成 张化光 张学工

章 毅 周鸿祎 周 杰 祝烈煌

《中国人工智能系列白皮书 元宇宙技术》编写组

陶建华 杨小康 何 斌 王茜莺 鲍虎军

王琼华 祝烈煌 何晓冬 李 琳 杨 艺

目 录

第 1 章	元宇宙技术概述	4
1.1	元宇宙技术的发展历程	5
1.2	元宇宙产业发展现状	10
1.2.1	海外发展格局	10
1.2.2	国内发展情况	13
1.2.3	国内外发展路径对比	15
1.3	元宇宙未来发展趋势	15
1.3.1	新信息形态	15
1.3.2	新数字器官	16
1.3.3	新经济模式	17
1.3.4	新社会图景	18
1.3.5	新沉浸业态	18
1.4	元宇宙面临的机遇和挑战	18
1.4.1	元宇宙发展机遇	18
1.4.2	元宇宙发展面临的挑战	19
第 2 章	元宇宙关键技术	21
2.1	元宇宙技术架构	21
2.2	元宇宙计算	24
2.2.1	分布式并行计算	24
2.2.2	元宇宙典型计算场景	26
2.3	元宇宙感知	31
2.3.1	微透镜阵列光场感知	31
2.3.2	计算光场成像	32
2.3.3	全景成像	35
2.3.4	多视图立体成像	36
2.3.5	音频感知	39
2.3.6	触觉感知	43
2.4	元宇宙生成	45
2.4.1	三维重建	46
2.4.2	数字人分身重构	51
2.4.3	三维内容生成	53
2.4.4	世界模型	56
2.5	元宇宙协同	57
2.5.1	区块链技术原理	58
2.5.2	区块链关键技术	61
2.5.3	元宇宙协同安全	68
2.6	元宇宙交互	77
2.6.1	信息显示	78
2.6.2	多模态交互	80
2.6.3	多模态情感计算	86

第3章	元宇宙技术平台	93
3.1	典型共性技术平台	93
3.1.1	超高清视音频技术	93
3.1.2	云渲染技术	95
3.2	典型元宇宙硬件平台	97
3.3	典型元宇宙内容构建平台	102
3.4	典型开源技术与工具	110
第4章	元宇宙典型应用	121
4.1	文旅元宇宙	121
4.2	教育元宇宙	124
4.3	工业元宇宙	127
4.4	办公元宇宙	131
4.5	娱乐元宇宙	134
4.6	体育元宇宙	138
4.7	金融元宇宙	140
4.8	供应链元宇宙	143
4.9	农业元宇宙	147
第5章	元宇宙安全治理：挑战与策略	150
5.1	元宇宙的安全风险	150
5.1.1	数据安全与治理	150
5.1.2	网络安全	150
5.1.3	社会安全	151
5.1.4	国家安全	151
5.2	元宇宙安全治理的现实挑战	152
5.2.1	数字平台的安全漏洞	152
5.2.2	数字资产的安全保障	153
5.2.3	国际合作与全球治理	153
5.3	元宇宙安全治理的策略与建议	154
5.3.1	技术层面的创新与应用	154
5.3.2	法律层面的完善与实施	155
第6章	总结与展望	157
6.1	元宇宙技术总结与展望	157
6.2	多方协同机制建设与元宇宙产业高质量发展	160
6.3	元宇宙应用规范制定和评测体系完善	162
参考文献	164
编写人员贡献	195

第1章 元宇宙技术概述

1992 年，美国科幻作家 Neal Stephenson 在《Snow Crash》中首次提出“元宇宙（Metaverse）”概念，他描绘了一种佩戴头显、连接终端即可通过虚拟化身进入的沉浸式数字空间。近年来，随着全球生产与生活方式持续加速数字化，社交、教育、办公等活动不断向线上延伸，社会对“持续在线、可交互的虚拟空间”的需求迅速增长，元宇宙的潜在价值也因此愈发凸显。

从技术视角来看，元宇宙由虚拟现实（VR）、增强现实（AR）、人工智能（AI）等多项技术共同支撑，是一种具备沉浸性、物理持久性、实时连接感知和多用户共享等特征的持续性虚拟环境。在这一空间中，用户可通过虚拟化身开展社交、协作、交易与创造活动，突破物理世界的时空边界，获得高度沉浸、无缝连接和强交互的数字体验。作为融合虚拟与现实的复杂数字生态系统，元宇宙构建了一个可持续演进的三维空间，支持用户在其中创造、探索、协作与娱乐，形成多元化的数字生活方式与数字经济形态。作为互联网形态的深度演进，元宇宙不仅承载社交、教育、商业、文旅、工业等多样化应用场景，还具备自组织、自管理与生态创新能力，为新型数字经济和数字社会的构建提供基础支撑。从科学视角来看，元宇宙的形成是多学科交叉融合的产物，涉及信息科学、数学、生命科学、人机交互、认知科学等多个领域。它在统一的技术体系下整合物理空间、社会空间、赛博空间与思维空间，通过软硬件设备、网络基础设施与智能体的协同运行，为人类打造一个持续演进的数字空间与复杂生态系统[1]。

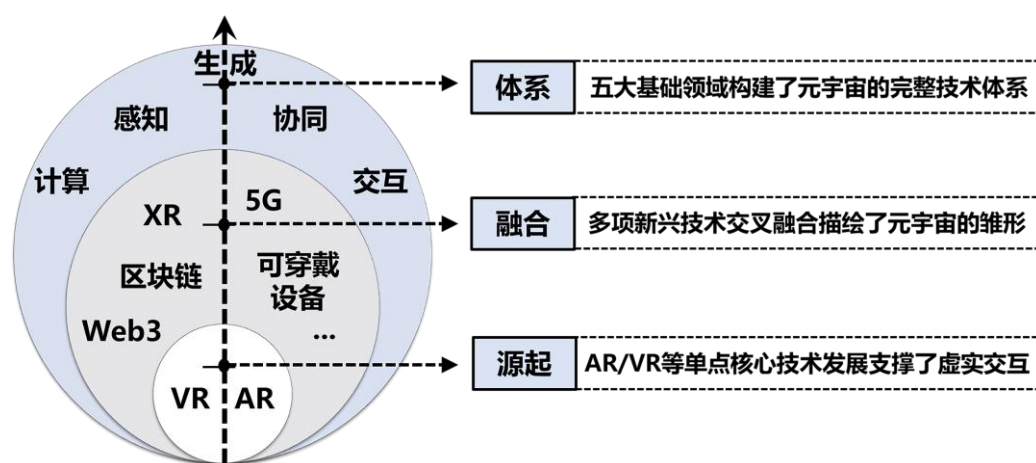


图 1-1 元宇宙的技术演进

1.1 元宇宙技术的发展历程

如图 1-1 所示，“元宇宙”的概念随着技术的演进而不断完善。元宇宙技术的萌芽始于互联网与 3D 图形技术的初步结合，源起于 VR/AR 局部体验的单点技术突破，其概念在科幻文学中孕育。这一时期，以《第二人生》为代表的大型多用户虚拟世界，首次在实践中构建了由用户生成内容并驱动经济的在线社会雏形，证明了持久化虚拟空间的技术可行性，但受限于当时的网络与算力，体验孤立且粗糙。进入 2010 年代，移动互联网、云计算和硬件技术迎来关键突破，为元宇宙积蓄了核心力量，形成由区块链、5G、Web3.0 和可穿戴设备等多项技术交叉融合的数字文明框架雏形。消费级 VR/AR 设备的复兴提供了沉浸式入口，4G/5G 网络保障了高速低延迟连接，而云计算则解决了海量数据运算与渲染的瓶颈。同时，人工智能与区块链技术的成熟，为未来世界的智能交互、内容生成及去中心化经济系统奠定了技术基石。当前，元宇宙发展进入生态整合与爆发期，科技巨头纷纷布局，技术焦点从构建单一虚拟世界，转向探索由 AIGC 驱动的内容生产、跨平台的互通性以及更逼真的沉浸体验。元宇宙正从一个分散的技术概念，演变为由计算、感知、生成、协同和交互等多五大基础领域共同构建的下一代智慧生态体系。

元宇宙的发展首先依赖于计算技术的跃迁，其演进正从“本地渲染”走向“云、边缘、AI 融合的空间计算”，从而支撑未来更真实、更智能、更大规模的虚拟世界。第一阶段是以 GPU 为核心的传统图形计算时代（1990s - 2010）。这一阶段中，实时 3D 渲染技术成熟，统一着色器架构普及，Unity 和 Unreal 等游戏引擎的出现奠定了虚拟世界构建的基础。但计算能力仍主要依赖本地设备，难以支撑大规模沉浸式体验。第二阶段进入云计算与分布式计算时代（2010 - 2020）。云计算平台迅速扩张，推动了云渲染、云游戏和边缘计算的发展，使得复杂场景可以在远端服务器实时计算，再以低延迟传输到终端。与此同时，大规模并行计算、物理仿真、分布式数据库等能力提升，使虚拟世界能够承载更多用户与更丰富的交互内容。第三阶段是 AI 驱动的智能计算时代（2020 至今）。大模型成为新的算力中心，推动了内容生成、场景理解和虚拟角色行为的智能化。NVIDIA Omniverse 等平台通过 GPU 加速计算、实时光线追踪、物理仿真和数字孪生技术，让工业级虚拟世界构建成为可能。同时，XR 设备出现专用芯片（如 XR2、R1），实现了高带宽传感数据处理与超低延迟渲染。

感知技术是元宇宙从“虚拟场景”迈向“空间计算时代”的核心基础。早期阶段主要聚焦在“采集”和“定位”，以单目/双目视觉感知（2000 - 2015）为主。Kinect 的推

出将深度感知带入消费级市场，同时 SLAM 技术在机器人和移动设备中广泛应用，为设备提供基本的空间定位与环境结构信息。多模态融合感知阶段（2016 - 2022），ARKit、ARCore 的发布让移动终端具备 IMU 与摄像头融合的即时定位与地图构建（VIO），显著提升了稳定性。消费级设备开始普及结构光、ToF、毫米波雷达等传感器，能够更准确地捕获深度、人体动作、手势轨迹和面部表情。眼动追踪和面部捕捉技术进一步推动了虚拟化身与用户状态之间的实时映射，使感知从几何理解发展到生理与行为理解。当前进入全场景理解与语义感知时代（2022 至今），以 Vision Pro 为代表的空间计算设备实现了高精度环境重建、物体识别、手势与视线交互融合处理，使设备能够真正“理解”用户及其所在空间。多模态感知开始与 AI 模型深度结合，具备场景语义解释、意图推断和复杂动态环境分析能力。随着传感器融合、空间神经场和大模型的发展，元宇宙中的感知正从“获取数据”走向“理解世界”，实现更自然、更沉浸的虚实融合体验。

生成技术是支撑元宇宙内容规模化构建的核心动力，其发展经历了从人工建模到大模型自动生成的跃迁。早期阶段主要依赖传统 3D 内容制作（2000 - 2015），虚拟场景、角色和动画大多通过 Maya、3ds Max 等软件由专业人员手工完成。这一模式尽管质量高，但成本昂贵、周期长，难以满足大规模虚拟世界的扩展需求。第二阶段出现 AI 辅助生成（2016 - 2021）。GAN 技术推动了虚拟人脸、纹理、物体外观等内容的自动生成，动作捕捉结合物理仿真为角色动画提供高真实度数据。基于程序化生成（Procedural Generation）的方法，也在游戏与虚拟场景中得到广泛应用，使内容生成能够部分自动化。然而，这一时期生成技术仍然碎片化，无法实现完整虚拟世界的自动构建。第三阶段伴随大模型的出现，生成技术进入多模态生成与自动构世界时代（2022 至今）。文生图、文生视频模型（如 Stable Diffusion、Sora）让虚拟素材生成速度成倍提升；文生 3D、NeRF 与 3D Gaussian Splatting 实现了虚拟物体与场景的快速重建与创作；AI 虚拟人可通过语言驱动表情、动作和语音的实时生成。Unity、Roblox 等平台也加入 AI 工具，实现从脚本、素材到场景的端到端生成。

协同技术是元宇宙能够承载多人活动、支持大规模创作与运营的关键能力，其发展经历了从在线共存到人机共创的系统性跃迁。初期阶段以在线协作与虚拟社区（2005 - 2015）为主，如 Second Life、Minecraft、VRChat 等平台，使用户能够在共享虚拟空间中社交、建造与互动。尽管功能有限，但奠定了“多人同步存在”的基础，标志着元宇宙协同的萌芽。第二阶段随着云计算、实时通信和创作工具的发展，进入平台化的实时协同时代（2018 - 2022）。Roblox、Fortnite Creative 等 UGC 生态提供脚本化创作、多

人编辑与虚拟活动承载能力。工业领域则出现 **BIM** 协同、云设计与数字孪生系统，使工程师能够在同一虚拟模型上同步查看、修改与仿真。低延迟网络、音视频实时传输与跨平台协作工具进一步推动了远程协同的普及。第三阶段是 **AI** 融合的智能协同时代（2022 至今）。**NVIDIA Omniverse** 等平台实现了高精度物理仿真、多软件互操作与稳定一致的实时协作，使虚拟世界能够支持工业级场景的多人设计与调试。同时，大模型与 **AI Agent** 开始进入协作流程，能够自动生成素材、优化模型、创建脚本并执行仿真任务，使人类与 **AI** 的协作成为新常态。元宇宙协同也扩展至虚拟会议、虚拟演唱会等大规模场景，支持百万级用户并发参与和互动。

交互技术是元宇宙体验从“看见虚拟”迈向“沉浸其中”的关键，它决定了人类如何进入虚拟世界、与数字对象沟通以及如何实现自然的人机融合。早期阶段主要依赖传统 **VR/AR** 交互方式（2014 - 2020）。**Oculus Rift**、**HTC Vive** 等设备引入 **6DoF** 位置追踪和控制器交互，使用户能够以手柄完成抓取、指向和移动。**HoloLens** 的空间投影与基础手势，也为早期混合现实体验奠定方向，但整体仍依赖“设备输入”。第二阶段随着传感技术与算法的突破，进入空间交互时代（2021 - 至今）。手部追踪、眼动追踪、面部捕捉等技术在 **Quest** 系列、**Pico** 设备和 **MR** 眼镜中成熟，用户不再需要控制器，即可用双手直接操控虚拟对象。苹果 **Vision Pro** 推动了全新的“眼动 + 手势 + 语音”三位一体交互范式，使交互更加自然、低学习成本。空间锚点、虚实融合渲染和深度分层让虚拟内容能够稳定地“生长”在真实空间中，使虚拟对象具备物理存在感。第三阶段正向全身交互与情感交互 发展。全身动作捕捉、躯体追踪与触觉反馈技术不断提升，用户的身体动作、姿态、表情与情绪可以完整映射到虚拟化身中；触觉手套、力反馈设备和可穿戴外骨骼正在实现真实触感的模拟。同时，大模型驱动的智能体（**AI Agent**）使数字角色能够理解意图、生成自然对话，提供高度沉浸的人机交互体验。

元宇宙是基于多种技术打造的虚实相生的数字世界，图 1-2 从技术的角度对元宇宙特点进行概括，其包括沉浸式、拓展性、多元化、永续性、去中心化、实时性等特点。元宇宙中用户不仅仅是观看内容，还可以利用多感官及认知体验，实现全面沉浸式的深度交互。元宇宙社会中，自然人、虚拟人、机器人三者共融共生，建设的环境可以永续运营，也可以发展为意识永生的载体和聚集地，跨越时间、空间乃至意识空间，达到多模互通。同时，元宇宙提供了跨虚实和跨生死的多维拓展，包括生存空间方面、视角维度方面、思想实践方面的多重延伸。元宇宙中用户可以一个身份或多个身份自由活动，内容生态是丰富多元的，同时还可以连接第三方的开放接口共享、修改、建设新内容。



图 1-2 元宇宙的技术特点

元宇宙和虚拟现实各自拥有其独特的内涵与应用场景，在定义、界面、持久性、内容、价值、支持技术、应用领域和未来发展等方面存在差异。

定义：虚拟现实是一种通过计算机科技生成的仿真三维环境，用户通过穿戴特制的头戴式显示器、VR 眼镜等设备与虚拟环境进行互动，具有类似真实环境的沉浸式体验。元宇宙是一个基于区块链、数字身份认证和智能合约等技术的虚拟空间，目的是在数字世界构建一个具有经济、社交和文化价值的生态系统。元宇宙不仅仅是一个虚拟环境，更是一个包含了现实世界和虚拟世界相互交织的综合生态系统。

界面：虚拟现实的界面由穿戴设备（如头戴式显示器和手柄控制器）、摄像头、耳机等关键设备组成，与用户的感官进行连接，让用户的感官能够沉浸在一个虚拟的或现实的环境中，虚拟现实通常是独立的虚拟体验，不一定与其他平台或服务互通。元宇宙的交互方式更加丰富多样，通常使用游戏手柄、键盘、鼠标等传统设备，还支持语音、触摸和动作识别等方式，让用户可以更加灵活地自由互动，并且强调不同虚拟平台之间的互通和互操作，形成一个统一的生态系统。

持久性：虚拟现实通常是短暂的体验，用户退出后虚拟环境可能就不再存在。元宇宙是一个持续存在的虚拟世界，即使用户退出，元宇宙仍然运行并发展。

内容：虚拟现实通常由独立团队开发设计，它的环境、角色和游戏物品都是由这些团队独立开发设计，用户通常是虚拟内容的消费者，而不是创造者。元宇宙的核心概念是用户可自由交换，这就意味着数据和资源的所有者是社区中的所有人，它的环境也可以由社区成员自己构建设计，成为一个更加开放和互动的环境。元宇宙鼓励用户生成内容，用户可以在元宇宙中创造和交易虚拟物品，甚至构建自己的虚拟空间。

价值：虚拟现实通常被视为游戏或娱乐的平台，元宇宙更多地重视于价值的保护和创造。元宇宙是一个真正的数字生态系统，由用户构建，不受中央机构控制，用户在这个环境中购买、出售、拥有、授权和交换任何在元宇宙中出现的数字资产。这提示着创造优质的虚拟资产以及给予创造者正确回报的机会。正因如此，元宇宙拥有一个商业生态系统，社区成员通过参与元宇宙的建设，为其他成员创造价值，通过买卖、投资、竞争和合作等方式实现利润，并将此用于推动元宇宙社区的发展。

支持技术：虚拟现实主要借助于计算机图形学、计算机视觉和人机交互等技术实现沉浸式体验。元宇宙则需借助于区块链技术和智能合约来平衡权益、管理数字资产和验证数字身份。

应用领域：虚拟现实目前主要应用于娱乐、游戏和培训等领域。例如，虚拟现实游戏可以提供给玩家身临其境的游戏体验，虚拟现实培训可以为学生或职业人士提供模拟的实战训练环境。元宇宙的应用领域则更加广泛，除了虚拟现实的应用外，元宇宙还可以支持在线社交、虚拟商店、虚拟会议等各种虚拟场景。在元宇宙中，用户可以与朋友进行聊天、购物、参加活动，还可以参观博物馆展览等场所，甚至可以创造自己的虚拟世界。

未来发展：虚拟现实和元宇宙是两个快速发展的领域，它们的应用场景各不相同。虚拟现实更加偏向于单纯地追求沉浸式的视听感知体验，它在游戏、娱乐、技术培训等领域有着广泛的应用。元宇宙以建立经济生态系统和数字资产管理为目标，涉及的行业更加广泛，包括金融、房地产、娱乐、艺术、文化等领域。

综上，元宇宙与虚拟现实之间存在紧密的关系。虚拟现实技术为元宇宙提供了技术基础和支持，元宇宙则为虚拟现实技术提供了广阔的应用场景和发展空间。元宇宙是一个更广泛、更综合的虚拟世界概念，虚拟现实则是一种实现这种世界的技术手段之一。随着技术的不断进步和应用领域的不断拓展，两者之间的融合将会更加深入，为用户带来更加沉浸式和多元化的体验。

元宇宙作为新一代信息技术发展的集大成者，带动了工业、商业、服务、娱乐等面向消费领域的创新和应用热潮，使得元宇宙技术在不同行业典型场景中深入应用并持续创新和快速发展。元宇宙技术虽然仍处于早期，但一旦落地也可能对诸多行业带来深远影响。元宇宙在以下六个领域的应用场景尤其值得期待，文化娱乐、虚拟人、教育培训、工业制造、医疗健康和远程办公等。

1.2 元宇宙产业发展现状

1.2.1 海外发展格局

元宇宙成为全球 ICT 领域热点议题。自 2021 年起，元宇宙概念快速升温，国内外政府给予高度关注和政策部署，Meta、谷歌、脸书、英伟达等全球市值前十科技企业在内的国内外 ICT 巨头与初创企业纷纷发声，主要从三个方面展开布局：一是依托消费电子领域发展大众创新应用，如 Meta、苹果等企业重点布局新型智能终端与下一代音视频。Meta Connect 2025 以智能眼镜为重点，提出“卓越眼镜为先（Great glasses first）+ 技术无感融入（Technology gets out of the way）+ 严肃对待超级智能（Take superintelligence seriously）”的头显终端设计理念，初步形成了“入门级 Ray-Ban Meta+进阶级 Hypernova+旗舰级 Orion/Artemis”轻量化智能终端生态。2025 年 10 月，苹果发布搭载 M5 芯片的新版 Apple Vision Pro，空间计算及 AI 处理能力大幅增强，带来更流畅、更真实、更具空间感的沉浸交互体验，同时 Vision Air 2027 年量产计划曝光，“高性能 pro + 消费级 Air + 智能交互”的元宇宙终端生态布局初步显现。二是依托工业互联网与云服务领域发展工业、企业元宇宙创新应用，如英伟达、微软等重点布局新型工业仿真、企业办公云服务及配套 3D 图形算力基础设施。三是依托区块链领域拓展互联网业务新的交易模式与经济活动，如 OpenSea、MetaMask 等初创企业积极布局元宇宙背后的新型数字资产与身份标识。

全球各国企业根据各自的产品与技术特点逐步推动向元宇宙探索发展，进而影响了各国政府的决策态度和政策导向。企业层面，海外以美国企业最为积极，美国在基础研发与核心技术优势明显，日本、韩国均基于已有的技术生态进行延展或迭代升级，巨头企业布局元宇宙及相关领域的动作频频，主要集中于 VR、人工智能、游戏等领域的元宇宙应用。新加坡、印度、欧洲等地也有众多初创企业得到了资本市场的关注。政府层面，尚未有国家将元宇宙作为国家层面的战略重点，而多以产业培育的方式引导行业发展。日本、韩国两国政府通过鼓励政企合作方式（业界团体、组织联盟等）以支持元宇宙相关产业的发展，欧洲、印度则偏重于出台个人隐私、数据安全等加强监管的要求，新加坡政府监管态度较为宽松。

全球元宇宙相关政策布局呈现三种策略。

一是，以美国为代表，政策焦点并未冠以“元宇宙”概念本身，而是集中在既有相

关重大领域的未来演进上，即沉浸式技术、基于区块链的分布式技术作为元宇宙概念技术面、经济面两大主攻方向。在重大技术研判上，2022 年 7 月，美国国会通过《2022 年芯片与科学法案》，授权拨款约 2000 亿美元用于促进美国未来十年的科研创新，并将先进的通信和 XR 为代表的沉浸式技术列入十项重大技术领域之一。同期，美国国会研究服务局（CRS）发布研究报告《元宇宙：概念内涵及国会考虑的问题》[1]，总结支撑元宇宙发展的三大核心技术一虚拟现实、5G/6G、区块链。在经济体系框架构建上，为填补加密货币技术的“政策空白”，2022 年拜登总统签署《关于确保负责任地发展数字资产的行政命令》，这是美国政府首次采取全面措施来应对数字资产及其基础技术风险的文件，表明美国正式推动构建基于美元的数字资产体系，强化数字资产风险防范，维持和保护美国在数字资产领域创新的领先地位。此外，该政策传递了美国政府对涵盖 Web3 概念在内的数字资产发展总体上持正面态度，并对美国商品期货交易委员会（CFTC）、证券交易委员会（SEC）等机构在加密资产监管职权分工上提供指导。

二是，以韩国为代表，高举元宇宙新产业的发展旗帜，从技术创新、经济发展、社会民生全局开展元宇宙顶层设计明确提出以“元宇宙”为政策落点的规划举措。2022 年韩国政府公布《元宇宙新产业领先战略——培育数字新政 2.0 引领新产业发展》以“数字新大陆，迈向元宇宙的韩国”为愿景口号，提出到 2026 年，元宇宙产业规模全球前五，专业人才规模不少于四万，50 亿韩元以上元宇宙企业数量不低于 220 家，并围绕生产生活诸多领域挖掘落地 50 个“元宇宙+”创新应用场景。

三是，以欧盟为代表，执行严格行业监管保护市场。近年来欧盟不断加强在数字领域的监管力度，建立《人工智能法案》《数字服务法案》《数字市场法案》等立法保障体系，通过严格监管手段防范国际巨头市场垄断。欧盟委员会考虑于 2023 年将元宇宙纳入立法，倡导构建一个安全和公平的元宇宙发展环境。2023 年 6 月，欧洲议会下辖公民权利和宪法事务政策部发布元宇宙研究报告，指出元宇宙可能对公民生活健康、工作 and 安全带来潜在挑战，建议构建更适应元宇宙的监管模式，鼓励创新合作的同时保障公民、企业的基本权利。

自 2021 年起，元宇宙概念快速升温，国内外政府给予高度关注和政策部署，Meta、谷歌、苹果、微软、英伟达等全球科技企业巨头与初创企业纷纷发声，如表 1-1 所示，主要从几个方面展开布局：

（a）基础设施方面：谷歌通过 AR Core 和操作系统、全息视频通话、沉浸式 3D 实景地图，积极布局元宇宙核心底座引擎。英伟达基于元宇宙图形渲染和 AI 学习等海

量计算问题，推出了专为 TB 级加速运算而设计的 CPUGrace 以及配套计算解决方案，同时针对元宇宙的创作需要，专门推出了 3D 设计协作平台 Omniverse。同时，在连接基础方面，Intel、高通等厂家均推出基于 RISC-V 架构的通讯芯片，旨在解决元宇宙的设备连接问题。

(b) 终端方面：通过跨时代的极致硬件体验提供消费级产品服务，发展大众创新应用。2024 年 Meta 头显（主要为 Quest 系列）全球出货量约 560 万台，VR 头显市场份额超过 70%。2025 年 9 月，Meta 发布带屏 AI 眼镜新产品 Meta Ray-Ban Display，支持在眼镜上显示导航、通知等功能，结合全新肌电手环可实现手势交互。2025 年 10 月，Apple Vision Pro（M5）例行更新，首次支持最高 120Hz 刷新率，同时 CPU 与 GPU 的性能显著增强，然而近 3 万元的起售价依旧限制该产品的受众面。

(c) 产品及内容方面：Roblox 通过游戏化创作打造开放消费者元宇宙，《Second life》以其逼真的建模与独立的经济体系在一众游戏中脱颖而出，《堡垒之夜》通过与明星合作举办线上演唱会开辟元宇宙演唱会道路。OpenSea、MetaMask 等初创企业依托区块链等核心技术构建全新元宇宙金融体系，积极布局元宇宙背后的新型数字资产与身份标识，持续拓展互联网业务交易模式与经济活动。

(d) 平台及生态方面：社交巨头 Meta 全面布局元宇宙领域，通过收购 Oculus 并持续加码 VR 研发，打造 HorizonHome、HorizonWorkrooms、HorizonWorld 等面向居家、办公和社交的平台生态，并通过底层技术研发开发工具及平台，以 SparkAR、PresencePlatform、Pytorch 等赋能内容创作，联手创作者共同实现元宇宙的宏大愿景。

表 1-1 全球主要科技企业在元宇宙布局范围及方式

排名	公司名称	2025 年 1-9 月营收（亿美元）	2025 年 9 月市值（亿美元）	布局范围及方式
1	APPLE	3136	37000+	研究 MR 设备；并购 25 家 AI 公司。
2	Alphabet（Google）	2888	30000+	研究 AR 设备和操作系统、全息视频通话、沉浸式 3D 实景地图。

3	Microsoft	2052	37000+	收购游戏开发商暴雪；发展游戏应用市场；研究 MR 设备。
4	Samsung	1695	4000+	推出数实融合的演唱会和商店；研究 VR 设备。
5	Meta	1410	18000+	综合布局 VR 和 AR 硬件、底层架构、人工智能、平台与内容。
6	Intel	392	1400+	研究高性能算力芯片，积极布局 GPU、IPU、RISC-V 芯片产品线。
7	NVIDIA	1185	40000+	从芯片技术优势延展，打造数字孪生引擎、数字人引擎和相关开发平台。

1.2.2 国内发展情况

随着数据、算力、算法的持续进步，国内元宇宙产业链已经初步形成，并且元宇宙市场规模在不断扩大，如图 1-3 所示。据中商产业研究院发布的《2024-2029 全球与中国元宇宙市场现状及未来发展趋势》[3]显示，2024 年中国元宇宙产业规模首次突破千亿元。展望未来，分析师预测，到 2026 年国内元宇宙市场规模至少将突破 4000 亿元大关。这一预测表明，未来几年元宇宙市场将继续保持高速增长，成为国内数字经济的重要增长点。

中国元宇宙行业市场也处于快速发展阶段，各级政府都对其发展给予了重要关注和支持，以百度、网易、字节跳动、华为、中国移动为代表的科技领军企业积极参与元宇宙领域的技术研发和创新应用。

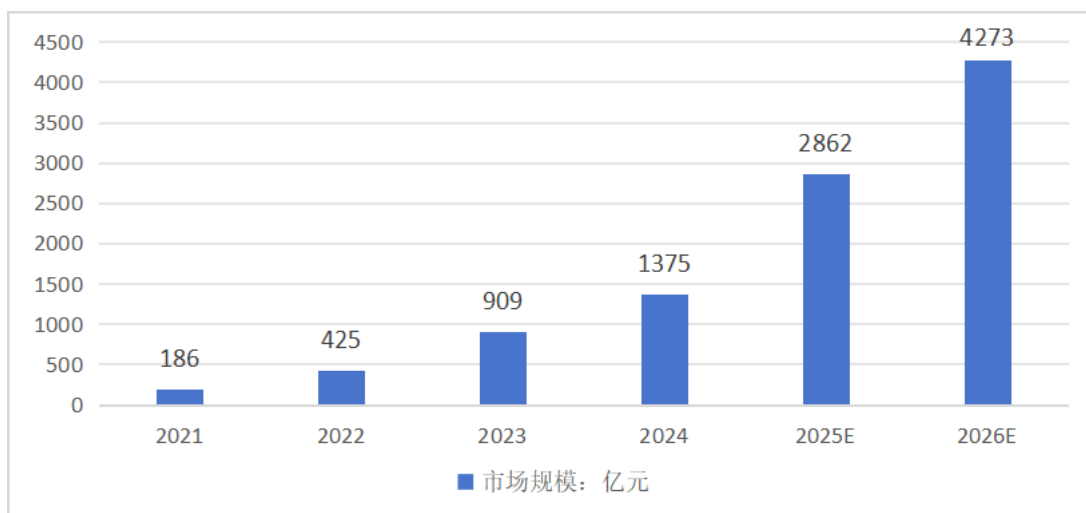


图 1-3 中国元宇宙市场规模（2021-2026E） 数据来源：中商产业研究院

（a）基础设施方面：华为从数字基础设施发力，研究操作系统、XR 引擎和开发平台等元宇宙底层技术，提供河图、AREngine、Meta Studio 等元宇宙产品及平台，融合华为云服务推进行业应用赋能。同时，国内基于 RISC-V 的芯片研究方兴未艾，正在逐步成为连接基础能力的主流方案。

（b）终端方面：字节跳动抓住 VR 终端硬件风口，收购 PICO 拿到元宇宙入场券，却终因面临硬件成本高、用户体验差、内容匮乏、市场规模小等诸多问题，发展不及预期进行战略性业务收缩。

（c）产品及内容方面：百度作为最早布局元宇宙的先行者之一，通过低门槛能力融合和场景快速复制推出希壤产品，致力于打造一个元宇宙虚拟创作体验和社交空间，并进一步升级产业营销规划。网易也推出瑶台，致力于打造线上社交和商务空间，通过结合数字分身、数字化场景提供了各种互动娱乐、商务活动模式，同时充分利用云网算力，依托于云游戏技术将渲染放在云端实现高精度场景体验，获得了不错的市场反馈和份额。

（d）平台及生态方面：咪咕围绕元宇宙产业开展体系化的业务技术布局，打造元宇宙领域综合型企业。技术上以算力网络为数字基座、以游戏化引擎为驱动、通过沉浸式媒介打造虚实融合的元宇宙平台能力。在业务上创新打造了冬奥会、世界杯和亚运会等多个大型赛事元宇宙数字化盛典，形成业内标杆示范应用，塑造了良好行业口碑。在生态上，通过元宇宙产业联盟，聚集产业生态伙伴，共建元宇宙产业链，进而不断扩大行业生态影响力。

1.2.3 国内外发展路径对比

元宇宙的发展离不开基础通讯、算力网络的支撑，从技术角度来看，5G 网络更多的是解决了面向用户的“最后一公里”问题，而元宇宙世界的应用则对全程全网的数据传输能力及算力调度能力提出更高的要求，这不但需要单点技术、单点能力的持续提升，更需要通信、存储、计算等技术的融合创新。网络快、算力高是元宇宙的基础条件，但只有二者的有机融合，统筹协作，才能发挥最大效益，否则将面临“木桶原理”的窘境，造成资源浪费。而二者的融合不仅仅是技术的融合，更需要标准化的打通、产业生态的重构，是一项需要提前布局，统筹协调，长期规划的大事，而这正是中国运营商的优势。

目前在全球范围内，元宇宙的发展尚未形成明显的竞争格局，但国内外企业在元宇宙的发展路径有所不同。

国外的元宇宙发展较早，受限于基础连接和算力网络发展不足，因此主要通过 AR/VR 硬件设施及 3D 渲染引擎等方面的技术、系统与装备进行布局，从终端设备布局“以实向虚”发展元宇宙产业。

我国元宇宙企业起步稍晚，并受国外尖端终端设备及芯片制约。为在元宇宙领域突破，我国扬长避短，选择了一条和国外元宇宙发展的不同路径，主要依托通信基础设施的雄厚实力，充分利用算力网络、基础通讯、大数据基础能力等方面的综合优势，从新型网络基础设施出发“以算向虚”结合企业自身元宇宙生产能力，打造具有中国特色的数实融合的新型算网应用。

1.3 元宇宙未来发展趋势

元宇宙概念的兴起既反映出当前互联网发展的痛点局限，又承载了人们对未来生活的目标愿景。放眼未来，全球元宇宙产业将形成以下多维度的发展趋势。

1.3.1 新信息形态

元宇宙中数字信息形态将由虚实分离向虚实融合演进。当前大众对美好生活的进阶需求对数字内容的体验方式提出了新要求，分辨率、帧率等视听质量维度的常规迭代难以带来用户体验的增量跃升。在元宇宙概念下，现实世界与虚拟世界不再彼此分割，而是逐渐融合，进入元宇宙这一虚实融合的沉浸空间有两条技术路径。其中，虚拟型元宇

宙是纯粹建立在虚拟世界里的数字空间，可独立于现实世界自主运行，上课、健身、娱乐、办公与购物等多元体验发生在虚拟世界，即虚拟世界真实化。增强型元宇宙须依附于现实世界存在，可基于用户在真实世界中实时位置、周围环境提供个性化的数字信息助手服务，即物理世界数字化。目前，苹果、脸书、微软、华为等 ICT 巨头对虚拟型、增强型元宇宙均有布局。

元宇宙沉浸体验的构建开放有赖于 2D 弱交互生态向 3D 强交互升级。传统互联网业务在实现用户虚实融合、身临其境的数字体验方面主要存在两大局限，一是 2D 化的虚拟世界与 3D 化物理世界难以直接融合，3D 技术产业链条的垂直突破与水平贯通将驱动元宇宙虚实融合数字空间成型，在这一背景下，当前涉足元宇宙的 ICT 巨头纷纷积极布局 3D 数据格式、传输、存储、渲染、建模、仿真与交互等细分领域。二是虚拟世界中“浏览旁观”的交互方式难以承载大众在现实世界中的“自主体验”，元宇宙 3D 虚实融合的数字空间将解锁“加强版”交互体验的新形态，除既有交互精度、频度与广度外，业界正对交互深度这一新方向大力投入。交互深度将推动人与虚实融合世界的自然交互，实现所见即所知，以视觉自然交互为基础的人物交互机制 HTI（Human Things Interaction）有望成为元宇宙中革命性的人机交互方式。

1.3.2 新数字器官

元宇宙新型智能终端将向适人体验方向发展。随着终端技术在感知、计算、传输、显示等方面的发展迭代，一方面，手机对移动互联网大众日常业务存在一定程度上性能过剩的迹象。另一方面，以手机为代表的传统智能终端难以承载新型人机交互的需求，桌面计算针对信息处理，智能手机聚焦沟通互动，元宇宙概念下终端入口或将是“体验式的”，适人体验成为终端用户的新需求与技术供给的新航道，沉浸体验、自然交互与使用舒适成为终端适人化的主要特点。沉浸体验契合元宇宙中数字信息虚实融合的体验形态，助力推动 3D 音视频处理、渲染与空间计算能力发展。自然交互旨在让用户聚焦交互活动本身，忘记交互界面（承载终端）的存在，人机交互界面愈发“透明”呈现自然化、情景化与智能化的发展趋势。使用舒适会结合人因工程社交通识等因素，重点关注聚焦元宇宙终端的佩戴舒适性与大众接受度。

虚拟现实头显将成为元宇宙重要的终端入口。个人电脑、XR 眼镜、脑机有望成为元宇宙终端入口的代表形态。手机等传统终端在虚实沉浸、自然交互、使用舒适方面存

在适人体验的固有限制，脑机接口准确率与效率距商用有较大差距。相比之下，国内外 ICT 巨头聚力抢位的 XR 终端有望成为元宇宙新兴终端入口的发展焦点。以 XR 为代表的适人化新终端存在多重技术挑战，一是功能性能与舒适性的再平衡，当前终端沉浸体验与重量体积、眩晕控制等部分舒适性指标间存在潜在冲突。二是日常使用与隐私保护的再平衡，元宇宙中用户环境隐私、行为隐私在种类、数量上呈现新的发展特点。三是终端交互效率与用户注意力分散的消长联动，信息交互效率的持续跃升将致使终端用户信息过载及注意力分散。

1.3.3 新经济模式

基于数字内容通证化探索元宇宙场景中的互联互通体系。元宇宙场景下，将依托于数字内容通证化探索建设互联互通的新型互联网价值体系，用户数字资产或将不再由具备“超级权限”的业务平台掌控即实现用户及其资产在元宇宙多元应用间的“星际穿梭”。在用户账户信息跨平台方面，借助新型数字钱包、DID 实现个人信息跨平台验证，“钱包”中存放的不再是货币，而是由个人而非平台掌握的用于验证元宇宙中数字身份的私钥，用于存储用户资产在区块链上对应的地址信息，DID 结合区块链技术实现用户身份自主可控和数据可信交换；在数字资产跨平台方面，即借助 NFT 等新型数字资产实现数字物品的资产所有权与数字身份绑定，与特定应用平台则可实现解绑、同时，虚拟经济活动会带来新的监管挑战，如假借元宇宙名义进行恶意炒作、非法集资、金融诈骗等。

创作者经济有望推动平台经济的价值重构。当前，任何拥有手机的人都已成为潜在的内容创作者，然而除了一些很有影响力的人外，大多数创作者并未得到相关经济回报。在元宇宙概念下，各大平台为获得创作者开展军备竞赛，创作者—消费者间的“付费直通”将为海量“长尾化”普通创作者提供内容变现的新机会，创作者经济有望兴起，将为众多互联网内容平台的“无薪实习生”提供新的就业生计同时，创作者经济或将削弱互联网平台的价值中介能力与编辑议价权：由于订阅付费使得创作者与消费者直接绑定，为创作者提供了一条直接与用户交流的途径，便于自身订阅粉丝跨平台转移，极大提升内容议价能力。当前互联网平台被元宇宙用户“分权”，掌控力下降，功能定位也在变化，即除内容推荐者与聚合外，在内容委托与策划方面发挥积极作用。

1.3.4 新社会图景

元宇宙有望创造由数字“比特”与人类“原子”深度融合的新型社会景观。作为一种社会媒介化技术，当前移动互联网推动了群体互动，日渐嵌入社会生活，人际沟通的物理距离溶解在双方所构建的虚拟空间中，并开始以一种持续加速的状态跳跃和流动。作为物理世界社会生活时空深度延伸的新型网络社会，元宇宙将加速人与数字信息的融合进程，大众通过技术对自身能力的延伸不再局限于物理空间，每一代人都比之前一代更加数字化。例如，在社会生活上，娱乐、社交、工作、上学、购物等体验也将发生在新型数字世界。在社会分工上，元宇宙或将总体增加就业岗位与工作类型如聚焦沉浸式技术工具的元宇宙世界的“建材商”、提供元宇宙人物环境海量场景内容的“开发商”、在元宇宙中组织经营多元生产生活体验活动的“运营商”。在社会传播上，元宇宙中网络效应的隐性强制程度与社会节奏或将进一步提升，呈现处处是中心、无处是边缘的发展态势，大众在信息过载的同时，难以在个体层面“下线”。

1.3.5 新沉浸业态

随着社会经济的发展和消费升级，消费者对“情绪价值”的需求日益增长，加速元宇宙沉浸体验业态创新裂变。幻境国际沉浸产业平台数据显示，当前中国已经出现了 171 种成熟沉浸业态类型，涵盖沉浸文化旅游、沉浸主题娱乐、沉浸实境娱乐、沉浸演艺演出、沉浸艺术展览、沉浸扩展现实、沉浸商业空间、沉浸美食餐饮、沉浸节庆活动、沉浸夜游夜娱、沉浸健康疗愈、沉浸教育成长、沉浸互动影游、沉浸酒店民宿、沉浸公共空间和其他沉浸小众领域等 16 大类。

1.4 元宇宙面临的机遇和挑战

1.4.1 元宇宙发展机遇

元宇宙是数字经济发展的下一波浪潮，可赋能实体经济的发展，是实现自身破局和提升数字经济实力的关键。此外，元宇宙作为前沿科学技术的集大成者，融合了多项高新技术产业链，有利于推动元宇宙相关的前沿科技繁荣发展。同时，未来随着相关技术

的不断演进和基础设施进一步完善，元宇宙应用场景将会不断丰富，可以融入到工业、医疗、商业、金融等各个实体经济领域，为产业带来新的形态、新的商业模式和新的增长空间，元宇宙通过供给新场景、塑造新赛道、培育新物种、加强新治理赋能传统产业，成为经济增长的重要引擎。

首先，居民消费需求升级刺激体验经济崛起。人们迫切需要一种数字经济背景下的沉浸式新型体验模式以满足多维互动式消费体验的需求。由此可见，居民消费需求和疫情的刺激触发了人们对元宇宙的期待，加速了元宇宙行业发展。

其次，数字经济亟须纾解内卷式发展瓶颈。数字经济亟须新的经济增长点作为行业发展与变革的新动力，为服务业与制造业数字化提供广阔的增量市场与技术协同发展的应用场景，以有效突破数字经济面临的存量市场“内卷化”困境。

最后，推进元宇宙发展与创新驱动发展战略高度契合。元宇宙是具备较长产业链条的高技术性新型经济形态，与各产业间的技术渗透性与互联性不断加强，并以创新型技术与平台型网络对各产业进行颠覆性改造，实现产品功能与结构升级换代、商业模式创新与企业组织管理模式扁平化。元宇宙相关数字技术将引发产业市场价值的伸展与整个产业网络的重塑，并演化发展出更多新的产业形态。元宇宙崛起必将成为引领经济发展的新引擎与新势能，是实施创新驱动发展战略的重要方向。

1.4.2 元宇宙发展面临的挑战

元宇宙整体发展虽有良好的发展基础，但在以下方面仍然存在巨大挑战：

一、关键底层技术发展不足成为制约元宇宙纵深发展的最大障碍。元宇宙本质上是数字技术的融合集成与创新应用，但目前元宇宙各技术分支的发展水平仍有待突破，底层支撑技术间互融与质变有待加深。另一方面，垄断加剧了元宇宙发展的困难。创新型企业行业进入门槛越来越高，造成新兴技术难以突破发展瓶颈。

二、传统监管模式与元宇宙运行逻辑存在冲突。元宇宙内逐渐形成开放式、扁平化、平等性的系统结构，从而逐渐弱化了强制性中心化机构的控制能力，导致传统监管手段难以介入。元宇宙下监管机构对虚拟经济运行的掌控力度被大幅削弱，传统中心化治理模式难以融入元宇宙去中心化系统，元宇宙发展面临虚拟经济运行逻辑与现实监管模式间协调问题。因此，在安全性方面，元宇宙需建立完善的数字备案和追踪体系，构建可信任的身份验证和设备保护措施，实现公平公正公开的信用评估和资产保护；在治理方

案上，一方面要设计以元宇宙平台为核心的算法治理以实现平台维护，另一方面要提供以用户为核心的社会治理以维持虚拟世界稳定发展。

三、元宇宙产业生态系统不完善制约各市场主体发挥溢出效应。受技术创新瓶颈难突破、元宇宙市场机制不健全等因素影响，元宇宙产业生态系统在基础设施层与核心层较为薄弱，大量创新主体通过引进国外先进技术涌入应用服务层，造成系统内要素与主体单一化、同质化严重，带来不同价值主体间要素连接不畅通、主体间融合创新能力差等问题，甚至在同一产业内不同企业间产生价格战等恶性竞争行为，导致元宇宙产业生态系统功能下降甚至衰退。

四、场景创新能力有待提升。游戏、社交等 C 端只是元宇宙早期阶段呈现的一种形态，B 端的行业一旦与元宇宙结合，会产生更大的商业价值。当前，元宇宙实时永续、数字孪生、融合现实等特性在生产制造上的深度应用前景远未被挖掘，需进一步创新场景应用，发挥元宇宙对实体经济虚实融合、以虚强实的赋能带动作用。

五、专项扶持政策力度不足。目前，对元宇宙产业的扶持政策较少，仅有对元宇宙产业的人才引进和服务支持设立相应的扶持政策，而对元宇宙产业的项目引进扶持（重大项目引进、产业化扶持）、创新扶持（研发成果奖励、技术改造奖励）和金融扶持（贷款贴息贴息、产业专项基金支持）等扶持型政策的建立略显不足。产业发展在初期都是相当脆弱，一个庞大的产业集群的发展离不开政府的支持和扶持。

此外，元宇宙的发展可能带来一系列新的问题，如隐私保护、伦理规则、舆论传播、金融安全等。

第2章 元宇宙关键技术

2.1 元宇宙技术架构

元宇宙（Metaverse）近年来在技术界与产业界持续受到高度关注，成为数字时代的重要前沿议题。元宇宙代表着一个由数字世界和物理世界深度融合而成的虚拟空间[4]。在这一空间中蕴含着广阔的创新潜力，使用户能够进行创造、社交、学习与娱乐等多样化活动。元宇宙的构建依赖多项关键技术，包括交互传感、区块链、人工智能、机器视觉、虚拟现实（VR）和增强现实（AR）等。本章将重点介绍支撑元宇宙发展的核心技术体系。如图 2-1 所示，元宇宙的五大支撑技术包括：计算、感知、生成、协同和交互[5]。这些技术共同构成元宇宙生态系统的核心功能，为元宇宙虚拟环境的基础架构提供支撑，使用户能够获得沉浸式体验并开展高效的协同活动。图 2-2 展示了上述五项支撑技术之间的关系及其协同作用。



图 2-1 元宇宙的五大支撑技术

计算技术是实现元宇宙愿景的基础支柱，并为其他技术提供核心支撑。元宇宙作为一个庞大的、持续运行的虚拟世界，需要处理海量实时渲染、行为反馈和状态更新等数据，这对计算性能和数据处理能力提出了极高要求。分布式并行计算是支撑元宇宙复杂计算需求的核心技术之一，可实现计算资源的高效调度与协同，并显著缩短计算时间，支撑虚拟世界中大量信息的实时处理和复杂场景的模拟。元宇宙高效构建、仿真演化和自然交互均依赖强大的计算能力。智能化云计算、广域计算、边缘计算和光计算等新型计算范式，使物理世界与数字世界得以实现低延迟、高可靠的深度融合。在强大算力的

支撑下，人工智能和大模型技术能够开展智能内容生成、场景自动构建和用户行为预测，从而显著增强用户的沉浸感与交互体验。随着计算技术持续演进与创新，其对元宇宙发展的推动作用将进一步增强，并将为元宇宙朝未来互联网的主流形态演进提供关键动力。

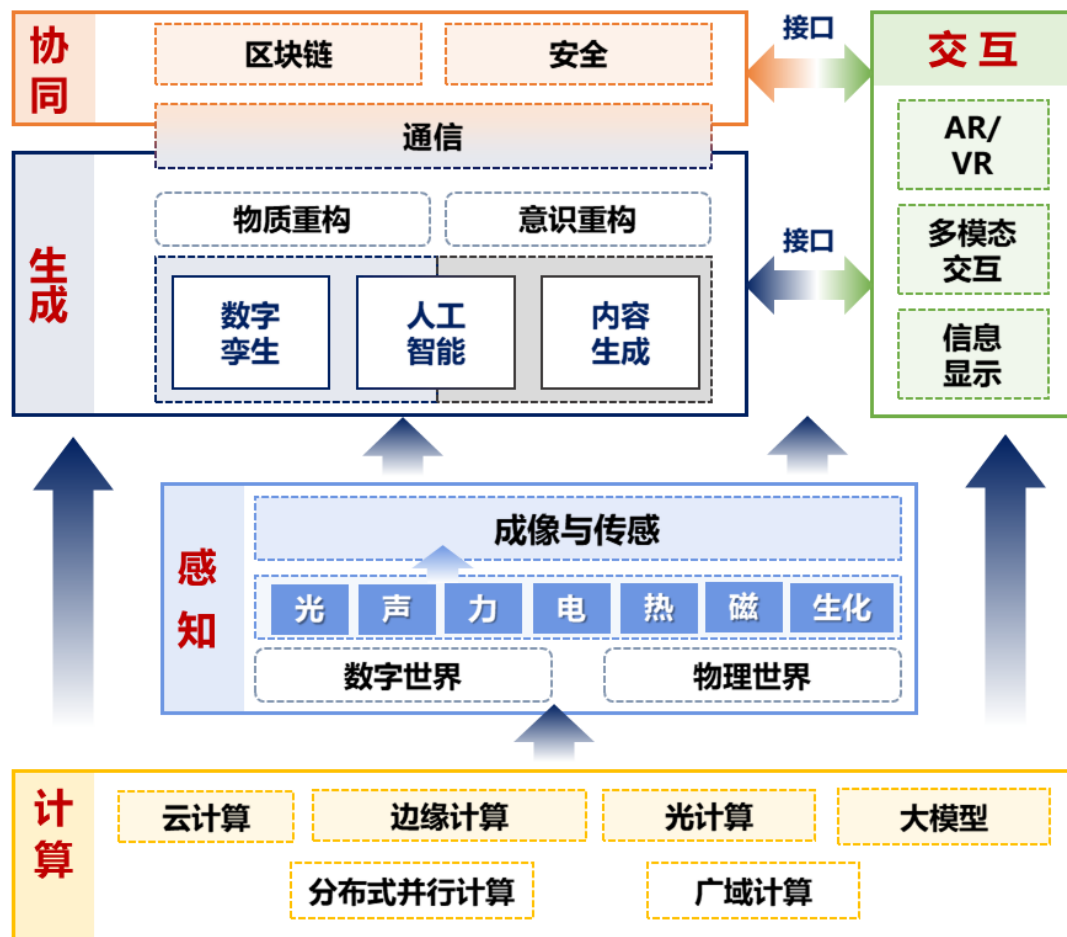


图 2-2 元宇宙理论与关键技术

感知技术在元宇宙构建中发挥着关键作用，可根据应用阶段分为内容生成感知和交互感知两大类。元宇宙感知技术通过将成像与多模态传感技术相结合，建立了物理世界与数字世界之间的桥梁，使元宇宙能够准确捕捉、解读并再现现实环境中的多维信息。元宇宙所依赖的感知体系涵盖光、声、力、电、热、磁及生化等多类型传感器。例如，光传感通过高精度摄像头与视觉雷达获取环境的视觉信息，并生成高分辨率三维图像；声传感捕捉并重建环境音频，增强空间真实感；力学与触觉传感模拟触感反馈，提升交互自然度；电与磁传感用于捕获电磁变化，为设备监测与控制提供精确信息；热与生化传感采集温度变化与化学成分，为虚拟场景构建提供更丰富的环境细节。通过多种感知技术的协同应用，元宇宙可实现物理世界和数字世界的信息无缝连接，为用户提供更高度保真度、更具沉浸感的虚拟体验。

生成技术在元宇宙发展中发挥着核心作用，能够实现虚拟环境的高效构建、个性化定制与动态更新，从而提供更加丰富、智能和沉浸的虚拟体验。基于多模态数据（图像、视频、音频、文本等）的生成技术，通过对数据的分析与建模，可实现虚拟内容的自动生成、场景重构与环境再现，使虚拟世界具备高度的真实性和可扩展性。数字孪生技术是生成技术的重要应用方向之一。通过对现实世界中的物体、系统与过程进行建模和仿真，数字孪生能够实时预测其行为与性能，并在人工智能的辅助下实现持续更新与优化，从而支持虚拟空间的动态调整与智能演化。此外，随着生成式人工智能的发展，虚拟世界的构建能力进一步延伸至认知层面。通过对人类意识、情感与行为模式的建模，生成技术能够模拟更高层次的认知过程，使元宇宙中的虚拟角色与交互体验更加生动、智能与具备情感表达能力。生成技术正在重塑元宇宙的内容生产方式，使虚拟环境具备逼真性、可塑性、自主演化能力，成为推动元宇宙生态持续扩展与创新的重要动力。

协同技术是支撑元宇宙中实时多人协作与互动的核心能力，构成元宇宙的关键连接与协作模式。作为未来互联网的演化方向，元宇宙将承载虚拟商品与服务、虚拟办公与生活场景，以及相应的社会与经济系统，因此，不同用户、不同场景乃至不同虚拟世界之间的信息协同与安全保障至关重要。在协同技术体系中，区块链发挥着重要作用。其去中心化特性为跨平台协作提供透明性与安全性，构建了物理世界与虚拟世界、以及不同元宇宙环境之间的可信连接基础。借助区块链，可构建可验证、难篡改的数字身份体系，实现用户身份的可信管理与隐私保护；同时，智能合约能够自动化执行虚拟资产交易，确保所有权、使用权和流转的准确性与可追溯性。非同质化代币（NFT）为虚拟资产提供唯一标识，使虚拟经济得以在安全和可控的条件下运行。区块链驱动的协同技术不仅增强了元宇宙的交互性与可信性，也为其安全性、可持续性和虚拟经济生态的健康发展奠定了基础。

交互技术对元宇宙至关重要，直接决定用户与虚拟环境之间的沟通方式和互动质量，是实现虚实融合的关键接口。多模态人机交互技术通过整合视觉、听觉、触觉等多种感官，使用户能够以语音、手势、触控等更自然、直观的方式与虚拟环境进行交流和操作，从而显著提升交互的沉浸感与真实度。因此，如何实现有效的人机交互是元宇宙中重要的问题[4]。增强现实（AR）和虚拟现实（VR）等沉浸式交互技术，为用户提供跨越物理世界与数字世界的沉浸式体验。先进的3D显示、空间音频、触觉反馈等前沿技术，进一步增强了AR/VR环境中的感知细节，使用户能够以更高的存在感与虚拟世界进行深度互动。交互技术的持续发展，使用户能够更自然地探索、操作并参与元宇宙中的虚

拟环境，是推动元宇宙体验丰富化、真实化和智能化的关键驱动力之一。

2.2 元宇宙计算

元宇宙高效构建、仿真演化和自然交互均建立在计算能力的基础之上，通过智能化云计算服务和端侧感知将物理世界与数字世界无缝融合[6]。由于元宇宙需要原生支持大规模用户接入，其计算架构的基础特征是端云结合的分布式并行计算。云计算通过多计算节点并行、低延时网络传输处理元宇宙大规模数据和复杂演化计算。当前热点的大模型技术为元宇宙所提供的自然语言交互和生成式能力也通常通过云计算提供，以满足其算力需求并降低成本。端侧计算设备的形式多样，包括智能终端、VR（Virtual Reality）/AR（Augmented Reality）、穿戴式设备、脑机接口（Brain-Computer Interface， BCI）等用户界面与使用者交互，负责用户行为或者其所处环境的基础感知计算，回传计算结果到云平台得到元宇宙反馈。因此，元宇宙的计算闭环需在端云甚至边缘协同发生，并以不同计算侧重点共同构建元宇宙相关应用，其主要计算场景包括三维图形渲染（3D Rendering）、人工智能（Artificial Intelligence， AI）、科学计算（Scientific Computing）、超高清流媒体（Ultra-High-Definition Streaming）等。以下是元宇宙计算的关键性技术[7]。

2.2.1 分布式并行计算

分布式并行计算（Distributed Parallel Computing）通过将计算任务分布到多个节点，实现并行处理和协同工作，以提高计算效率、扩展性和处理能力，广泛应用于需要处理大量数据和复杂计算的场景。通过分布式并行计算，可以实现计算资源的高效利用，并显著缩短计算时间[8]，是支撑元宇宙复杂计算需求的核心技术之一。分布式并行计算通常由计算节点、通讯网络、分布式资源管理系统等几个部分组成。在此之上，又有调度系统、负载管理系统、集群管理系统、运维监控系统等对基础资源进行统一的管理调度运维。通过统一管理系统，分布式并行计算可以支撑包括三维图形渲染、人工智能、科学计算、超高清流媒体等元宇宙核心计算需求。

元宇宙计算节点通常是由 GPU（Graphics Processing Unit）主导，负责实际的计算任务。每个节点通常配备高性能的处理器和存储设备。计算节点可以是物理服务器、虚拟机（Virtual Machine）或容器（Container）。支持虚拟化技术，可以实现资源的灵活分配和隔离，提高资源利用率。目前 GPU 仍然由 NVIDIA 公司产品主导，其 GPU CU

DA 开发生态在人工智能模型训练、推理应用广泛。国产人工智能计算硬件生态也正在快速发展，国内摩尔线程 GPU 卡、华为昇腾 NPU、寒武纪的 Cambricon-1M 芯片、阿里平头哥和沐曦科技系列训推 GPU 芯片得到了国家的大力支持,已经在市场取得了一席之地，但在软件开发生态上还需进一步完善。

计算节点之间通过高速网络互连，如 InfiniBand、高速以太网等，以实现低延迟、高带宽的数据传输。分布式资源管理系统负责管理和协调计算节点之间的资源分配和任务调度。它接收来自上层应用的计算任务，将任务分解为多个子任务，并将子任务分配给合适的计算节点执行。资源管理系统需要考虑负载均衡（Load Balancing）、容错（Fault Tolerance）、数据本地性（Data Locality）等因素，以实现高效的任务调度和执行。常见的资源管理多使用容器，常见的容器管理系统包括 Kubernetes、Docker Swarm 等。在分布式并行计算的基础上，元宇宙计算还需要一系列的优化技术，以进一步提高性能和可扩展性。这些优化技术包括：

(1) 异构计算优化（Heterogeneous Computing Optimization）：元宇宙计算通常涉及不同类型的计算任务，如图形渲染、人工智能、科学计算等。这些任务对计算资源的需求和特性不尽相同。异构计算优化旨在充分利用不同类型的计算资源，如 CPU、GPU、FPGA（Field-Programmable Gate Array）等，将任务映射到最适合的计算资源上，以获得最佳性能。这需要设计智能的任务调度算法和资源管理策略，根据任务特征和资源状态动态调整任务分配。

(2) 数据本地性优化（Data Locality Optimization）：元宇宙计算涉及大量的数据处理和传输。为了减少数据移动的开销，需要将计算任务尽可能地调度到离数据最近的计算节点上执行。这就需要数据本地性优化技术，如数据感知任务调度、数据预取（Data Prefetching）和缓存（Caching）、数据压缩（Data Compression）和编码（Encoding）等。通过将数据和计算任务合理地放置在集群中，可以显著减少数据传输的延迟和带宽消耗，提高计算效率。

(3) 通信优化（Communication Optimization）：元宇宙计算中的许多任务需要频繁的节点间通信，如数据同步、模型更新、中间结果交换等。通信优化技术旨在减小通信延迟和带宽消耗，提高通信效率。这包括网络拓扑优化、通信协议优化、集合通信原语（Collective Communication Primitives）优化等。例如，使用高速网络互连技术，设计智能路由算法，采用轻量级通信协议，利用硬件加速器等，都可以显著改善节点间的通信性能。

(4) 容错和故障恢复 (Fault Tolerance and Recovery)：在大规模分布式计算环境中，节点故障和网络中断是不可避免的。为了保证计算任务的可靠执行，需要采用容错和故障恢复技术。这包括检查点机制 (Checkpointing)、故障检测与隔离 (Fault Detection and Isolation)、任务重试与迁移 (Task Retry and Migration) 等。通过定期保存计算状态，快速检测和隔离故障节点，自动重试失败任务，将任务迁移到健康节点等手段，可以最大限度地减少故障对计算任务的影响，提高系统的可用性和鲁棒性。

(5) 自适应调度优化 (Adaptive Scheduling Optimization)：元宇宙计算的工作负载具有动态性和不确定性。用户请求的到达率、任务类型、数据规模等都可能随时间变化。因此，需要采用自适应调度优化技术，根据系统的实时状态和历史数据，动态调整任务调度策略、资源分配方案和参数配置等。通过实时监控系统指标，预测未来的负载趋势，利用机器学习算法不断优化调度决策，可以实现计算资源的自适应分配和动态伸缩，提高系统的灵活性和自适应能力。

(6) 低延时高速网络：元宇宙需要支持大规模用户的实时交互，这对网络的延迟和带宽提出了极高的要求。低延时高速网络技术，如 5G、Wi-Fi 6 等，可以提供高达 10Gbps 的峰值速率和毫秒级的端到端延迟，满足元宇宙的实时交互需求。同时，网络切片 (Network Slicing)、边缘计算 (Edge Computing) 等技术可以进一步优化网络性能，提供差异化的服务质量保证[9]。

元宇宙是一个全球规模的虚拟世界，需要跨地域、跨运营商的计算资源协同。最新提出的广域计算 (Wide-Area Computing) 技术可以认为是分布式并行计算的进一步拓展，其通过建立广域资源调度和任务编排机制，实现跨域计算资源的统一管理和调度。这需要解决资源发现、任务分解与映射、数据传输优化等挑战。区块链 (Blockchain)、多方安全计算 (Secure Multi-Party Computation) 等技术可以用于构建去中心化的广域计算平台，保证资源共享的安全性和隐私性。

2.2.2 元宇宙典型计算场景

在分布式并行计算的支持下，元宇宙计算架构正面临从“连接在线”向“智能共生”的代际跨越。当下的元宇宙不再是静态资源的简单堆砌，而是逐渐由世界模型 (World Models) 驱动的、遵循物理规律的动态演化系统。海量 AI 智能体 (Agents) 的引入，使得计算负载从传统的图形渲染扩展至高频的逻辑推理与认知规划，这对算力的实时性与

吞吐量提出了指数级的增长需求。分布式并行计算通过解耦算力供给与应用需求，利用云边端三级架构的异构优势，实现了从底层物理仿真到上层高斯泼溅（3DGS）渲染的全链路加速。它不仅要支撑视觉上的极致逼真，更要维持数字世界的逻辑自洽与持续运行，确保数以亿计的用户与智能体能够在同一时空下进行低延迟、高带宽的深度交互。在此基础上，元宇宙数字空间中的典型计算场景如下：

（1）云边端协同的三维图形渲染：在元宇宙场景下，计算任务在云端、边缘侧、终端设备之间动态分配和协同执行。云端提供强大的计算和存储能力，负责复杂的渲染、模拟、数据分析等任务；边缘侧提供低时延的本地计算和数据处理，负责实时交互、内容缓存、数据过滤等任务；终端设备提供个性化的交互和呈现，负责本地渲染、传感器数据采集、用户行为分析等任务[5, 6]。云边端协同计算（Cloud-Edge-End Collaborative Computing）需要统一的任务调度和资源管理机制，以及高效的数据同步和一致性维护技术。为了满足元宇宙中超大场景的实时绘制要求，云边端协同绘制技术将整体场景的绘制分解为三个部分：云侧的海量存储与复杂光栅化计算、边侧的低时延数据传输与指令协同、终端侧的轻量绘制与交互式显示。这种绘制框架高效地利用了不同侧的硬件特点，实现了以往在单一终端无法完成的超大场景绘制任务。以百亿级面片大规模场景为例：在绘制时，终端由显示或交互操作发出绘制指令；边侧负责将绘制指令传递到云侧并维持同步状态；云侧下发绘制指令到多张显卡，每张显卡负责部分场景的光栅化计算，并最终组合为整体场景的光栅化图形缓存数据（G-Buffers）；轻量化的 G-Buffers 数据经由边侧的图形数据传输协议低时延地传输到终端侧，并更新同步状态；最终由终端侧负责解码与绘制 G-Buffers 数据并显示绘制结果。云边端协同绘制（Cloud-Edge-End Collaborative Rendering）需要统一的任务调动和资源管理机制，以及定制的资源传输协议与多卡绘制技术。

近年来，尽管神经辐射场（Neural Radiance Fields, NeRF）技术为三维场景的表示和渲染奠定了基础，但在 2025 年，3D 高斯泼溅（3D Gaussian Splatting, 3DGS）技术已全面超越传统方法，成为元宇宙实时渲染的主流范式。不同于 NeRF 的隐式表达，3DGS 采用显式的点云表示，通过将每个点建模为具有各向异性协方差、位置、不透明度及球谐函数颜色的高斯椭球，结合高效的可微分光栅化管线，彻底解决了复杂场景下的实时推理难题。3DGS 技术在云边端架构中展现了极高的适应性：云端负责处理海量图像数据以训练高精度的高斯模型，并进行动态场景的 4D 高斯融合；边缘侧利用压缩算法对高斯流进行剪枝和量化，大幅降低传输带宽；而终端设备仅需极低的 GPU 算力

即可实现 100fps 以上的照片级渲染。相比传统的点云和网格渲染，3DGS 不仅消除了空洞和锯齿现象，生成更加平滑、连续的表面，还支持实时的场景编辑与物理碰撞模拟。此外，针对大场景的 LOD（Level of Detail）层级管理技术在 3DGS 中得到了进一步发展，系统可根据视点距离动态加载不同密度的高斯球集合，在保证视觉保真度的同时最大化渲染性能。

（2）元宇宙领域的人工智能应用：元宇宙将包含大量智能化的虚拟助手、非玩家角色（Non-Player Character, NPC）、推荐系统等。人工智能技术，如机器学习（Machine Learning）、知识图谱（Knowledge Graph）、自然语言处理（Natural Language Processing, NLP）等，可以赋予虚拟角色智能交互和决策能力，提供个性化的内容推荐和服务[11]。

随着大模型技术的演进，AI 智能体（AI Agents）的计算需求在 2025 年呈现爆发式增长，成为元宇宙算力消耗的核心来源。与传统的问答式 LLM 不同，Agent 具备感知、记忆、规划和工具使用的自主能力。在计算层面，这意味着单一任务的推理不再是线性的，而是包含了思维链（Chain of Thought）推理、多轮次自我反思以及复杂的任务拆解，导致单位交互的 Token 消耗量呈指数级上升。同时，为了维持 Agent 的长期记忆和个性化特征，系统需要频繁地对海量向量数据库（Vector Database）进行检索和更新，这对内存带宽和存储 I/O 提出了极高要求。此外，云端大模型与端侧小模型（SLM）的协同推理成为标配，端侧设备负责处理实时的意图识别和隐私敏感的轻量级任务，而复杂的逻辑规划则卸载至云端，这种混合计算架构要求极低延迟的通信链路。

与此同时，世界模型（World Models）技术的快速发展为元宇宙的构建带来了革命性变化。超越了传统的像素级生成，世界模型通过学习海量视频数据，掌握了物理世界的时空演化规律（如重力、碰撞、流体运动等）。结合 Sora 等视频生成技术，世界模型不仅能生成逼真的静态画面，还能预测环境对用户操作的动态反馈。这意味着元宇宙的物理引擎将逐渐从“基于规则的手工编码”转向“基于神经网络的物理仿真”，极大地降低了构建高保真虚拟世界的门槛，但也带来了巨大的模型训练和实时推理算力挑战。

多模态人工智能（Multimodal AI）是元宇宙中的另一个重要方向。多模态 AI 旨在处理和理解来自不同模态（如文本、图像、音频、视频等）的信息，并实现跨模态的信息融合和生成。视觉-语言预训练模型，如 CLIP（Contrastive Language-Image Pre-training）、DALL-E（A 12-Billion Parameter Model of Natural Language）等，可以根据文本描述生成或检索相关的图像，实现跨模态的理解和创作。多模态学习也已经扩展到视频和音频领域，支持更加丰富和逼真的虚拟场景生成。多模态交互技术，如视线追踪（Eye Tracking）、

手势识别（Gesture Recognition）、语音控制（Voice Control）等，可以为用户提供更加自然和沉浸式的交互体验[10]。

AIGC（AI-Generated Content）是元宇宙内容创作的新范式。扩散模型（Diffusion Models）作为一类生成式模型，通过迭代式的去噪过程，学习数据的分布，并生成高质量、多样化的内容。Stable Diffusion、Midjourney 等工具让用户能够根据文本提示自动生成图像，而 Runway、Stable Video Diffusion 等模型则支持生成连贯的高清视频。

为了支持元宇宙中海量数据和模型的训练和推理，还需要研究分布式机器学习算法（Distributed Machine Learning Algorithms）和联邦学习（Federated Learning）架构。隐私保护（Privacy Protection）、安全机制（Security Mechanism）也是不可或缺的，以确保用户数据的隐私和生成内容的合规性。人工智能技术，特别是 Agent 与世界模型的融合，在元宇宙中扮演着至关重要的角色，它不仅赋予虚拟世界以智能和自主性，还为用户提供个性化、沉浸式的交互体验，并催生了新的内容创作范式。

（3）科学计算：元宇宙的建模和模拟离不开大规模科学计算。传统的科学计算主要依赖物理规律和数学方程，通过数值模拟（Numerical Simulation）和仿真（Simulation）的方法，研究复杂系统的行为和演化。物理引擎（Physics Engine）、流体动力学（Computational Fluid Dynamics）、分子动力学（Molecular Dynamics）等模拟技术，可以生成逼真的物理环境和现象。高性能计算（High-Performance Computing, HPC）技术，如并行算法优化（Parallel Algorithm Optimization）、向量化（Vectorization）、高速互联（High-Speed Interconnect）等，可以加速科学计算的执行。

然而，随着问题复杂度的增加和对精度的更高要求，传统的科学计算方法面临着计算成本高、建模困难等挑战。近年来，人工智能技术为科学计算领域带来了新的突破（AI for Science）。通过将机器学习与物理模拟相结合，可以显著提高科学计算的效率和精度。一方面，机器学习可以用于加速传统的科学计算。例如，在流体动力学模拟中，可以使用机器学习模型预测湍流（Turbulence）、边界层（Boundary Layer）等复杂流动现象，从而减少对昂贵的直接数值模拟（Direct Numerical Simulation, DNS）的依赖。在分子动力学模拟中，机器学习力场（Machine Learning Force Fields）可以根据分子结构快速预测原子间的相互作用力，避免了繁琐的量子化学计算。图网络（Graph Networks）则可以学习粒子系统的相互作用关系，预测复杂材料和结构的力学性能。

此外，生成式模型如生成对抗网络（GAN）、变分自编码器（VAE）等，可以学习物理系统的概率分布，生成合理的状态样本，用于加速蒙特卡洛采样（Monte Carlo

Sampling) 和不确定性量化分析 (Uncertainty Quantification)。强化学习 (Reinforcement Learning) 在优化控制、实验设计等领域也得到了广泛应用。将人工智能与科学计算相融合, 不仅可以突破传统方法的瓶颈, 还能发掘新的建模范式和计算模式。在元宇宙中, AI 驱动的科学计算将支持更加高效、精准、智能的物理模拟和仿真, 为构建逼真的虚拟世界提供强大的技术支撑。

(4) 超高清流媒体: 元宇宙中将产生和消费海量的视频、音频等流媒体数据。超高清视频编解码 (Ultra-High-Definition Video Coding)、传输优化 (Transmission Optimization)、内容缓存 (Content Caching) 等技术, 可以提供身临其境的视听体验。同时, 需要研究面向元宇宙的新型流媒体架构, 如支持自由视点的沉浸式视频 (Free-Viewpoint Immersive Video)、面向 6DoF (6 Degrees of Freedom) 的空间音频 (Spatial Audio)、体积视频 (Volumetric Video) 等。基于 AI 的流媒体内容理解、推荐、审核等技术, 可以为用户提供个性化、安全、有序的内容服务。

未来, 随着技术的进一步发展和融合, 元宇宙计算平台将不断演进和完善, 有望为元宇宙注入更强大的算力; 元宇宙计算平台的建设将是一个长期、复杂的系统工程, 需要产学研各界的共同努力和协作。

(5) 端侧用户交互计算: 元宇宙的交互形态正经历从重型头显向轻量化 AI 眼镜 (AI Glasses) 的深刻变革。这种变革要求端侧设备在极低的功耗和重量限制下, 实现高密度的智能计算。AI 眼镜作为全天候佩戴 (All-day Wearable) 的计算终端, 集成了微型化的波导显示模组、麦克风阵列以及多光谱摄像头。这就对端侧计算提出了极高的能效比要求: 设备必须搭载专用的神经处理单元 (NPU), 以毫瓦级的功耗运行量化后的多模态大模型。这些端侧模型不仅需要实时处理语音指令, 更要具备看和问 (Look and Ask) 的能力, 即实时识别用户眼前的物体、文字及环境上下文, 并进行语义分割和意图理解。

与传统的 VR 渲染不同, AI 眼镜的计算重点从全沉浸式的三维重建转向了增强现实 (AR) 信息的精准叠加与情境感知。例如, 通过 SLAM (Simultaneous Localization and Mapping) 技术实现厘米级的空间定位, 结合端侧视觉模型, 将虚拟导航箭头精确锚定在现实道路上, 或在用户注视名画时自动弹出详细解说。为了保护用户隐私并降低延迟, 大部分感知计算需要在本地完成, 仅将复杂的逻辑推理请求发送至云端, 形成端云协同的混合 AI 架构。此外, AI 眼镜还将成为 Agent 的第一人称交互接口, 用户无需动手, 仅凭自然语言即可调度各类服务。这种无感、泛在的交互方式, 将混合现实技术无缝融入日常生活, 对端侧芯片的异构计算能力、传感器融合算法以及热管理设计提出了前所

未有的挑战

2.3 元宇宙感知

元宇宙感知按照应用的不同阶段分为内容生成感知和交互感知两大类，其中内容生成感知是指在沉浸式元宇宙内容生产过程中所使用的信息采集技术；交互感知是指在体验元宇宙内容过程中通过模拟人类的感知能力，使用户能够在虚拟世界中感受到真实的环境。当前业内研究的焦点主要集中在光场感知、音频感知、触觉感知等相关技术领域。

2.3.1 微透镜阵列光场感知

基于透镜阵列的光场采集[12]主要依靠在成像传感器与主镜头之间加入一片微透镜阵列，物体表面光线首先经过主镜头，然后经过微透镜，最后到达成像传感器（e.g. CCD/CMOS）。如图 2-3 所示，物体表面 A 点在 FOP 角度范围内发出的光线进入相机主镜头并聚焦于微透镜，微透镜将光线分成 4×4 束，并被成像传感器上对应的 16 个像素记录。类似的，空间中其它发光点，例如 B 点和 C 点，在其 FOP 角度范围内的光线都被分成 4×4 束并被分别记录。

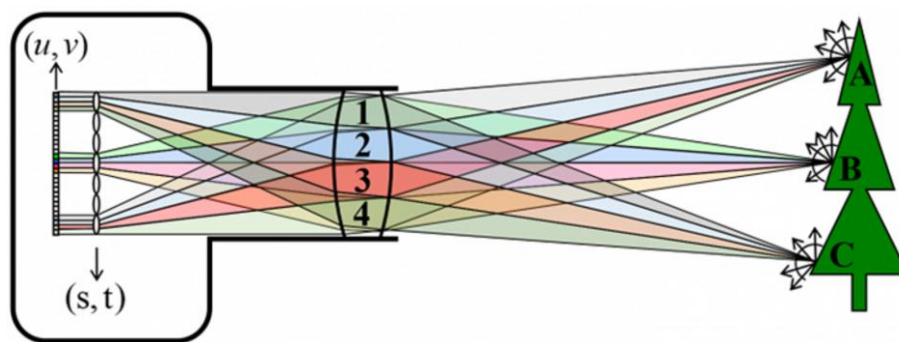


图 2-3 基于微透镜阵列的光场采集原理示意图[12]

微透镜阵列的主要作用是将物体表面同一点在一定角度范围内的各个方向发出光线进行离散化解耦，图 2-3 中的微透镜将光线离散化为 4×4 束，离散化程度由微透镜光学参数、主透镜光学参数及微透镜与成像传感器之间的距离等多方面因素决定。

光场相机[13]是基于微透镜阵列光场采集的原理，通过在主透镜与传感器元件间放置一块紧密排布的微透镜阵列，仅需单次曝光即可获得光线的空间、角度四维信息，经过后期算法处理可以得到场景处于不同视角和不同景深的图像以及深度信息，具有结构

紧凑，体积小，采集效率高等优点。光场相机在工业检测领域具备广泛的应用场景和潜力。其能够提供更加准确和全面的图像数据，有助于实现精确的工业检测和质量控制。

散焦型光场相机，微透镜阵列放置在主透镜像侧焦平面位置，因此微透镜阵列对主透镜所成的像直接成像。对于散焦型光场相机系统，微透镜阵列与传感器元件的间距为一个微透镜焦距，为微米尺度，封装要求极高，需要采用高精度芯片级封装工艺，从而得到精度更高的深度计算结果。

聚焦型光场相机，微透镜阵列并不是放置在主透镜像侧焦平面位置，因此微透镜阵列实际上是对主透镜所成的像进行二次成像。根据聚焦型光场相机的成像原理，微透镜阵列的位置可以进行调整，微透镜阵列与传感器元件的间距为毫米尺度，封装要求较低，微透镜阵列易于封装。

由于微透镜阵列光场感知技术受限于微透镜阵列基线的大小，光场相机通常只能对数米内的近景进行深度估计，在工业以外的应用场景有一定局限性。

2.3.2 计算光场成像

计算光场成像技术围绕光场及全光函数表示，旨在结合计算、数字传感器、光学系统和智能光照等技术，以及硬件设计、软件计算能力，突破经典成像模型和数字相机的局限性，建立光在空域、视角、光谱和时域等多个维度的关系，实现耦合感知、解耦重建与智能处理，具备面向大范围动态场景的多维多尺度成像能力。光场成像技术正逐渐被应用于生命科学、工业探测、无人系统和虚拟现实/增强现实等领域，具有重要的学术研究价值和广阔的产业应用前景。然而，伴随着高维数据的离散化采样，光场成像面临空间分辨率与视角分辨率的维度权衡挑战，如何对稀疏化的采样数据进行光场重建成为计算光场成像及其应用的基础难题。与此同时，受制于光场信号的高维数据感知量，光场处理面临有效数据感知与计算高效性的矛盾。如何用光场这一高维信息采集手段，取代传统二维成像视觉感知方法，并结合智能信息处理技术实现智能化高效感知，是实现光场成像技术产业化应用的巨大挑战。

杜克大学的 Brady 教授课题组[14]于 2012 年在《Nature》上发表了关于通过前端球透镜进行一次成像后，后端通过分视场相机阵列（约 98 个）进行采集的 AWARE-2 相机阵列的研究论文，如图 2-4 所示。AWARE-2 相机阵列采用的是多重光学镜头的设计，受益于球透镜的成像单一光心特性，该设计避免了分视场相机之间的视差，大大降低了图

像拼接的难度，可达到每秒几帧的 10 亿像素采集速度。尽管通过像传感器阵列的拼接可初步验证动态 10 亿像素成像的可行性，但类似 AWARE-2 这种结构化的相机阵列受到均匀化采样、信号密度同分布，相邻像传感器间遵从视场相互重叠的严格限制，相机阵列的装调和标定极其严格、复杂，其中大规模分视场像传感器的引入对实际安装、标定和拼接均带来了极大的技术挑战，实现成本高、难度大且灵活性低。传统的结构化像传感器阵列，其成像感知遵从均匀化采样机制，每个像传感器节点通过部分重叠区域连接到其相邻的像传感器节点，所感知的数据遵循“图”连接结构。在均匀化采样的假设下，将拼接得到的全景图像/视频投影为矩阵形式并进行数据的表征与存储。此外，各个像传感器均遵从双向图连接，单像传感器的调整势必会影响其他所有像传感器。图形结构的另一个问题是，由于其位姿估计严重依赖重叠区域的全局优化，随着像传感器数量增加，复杂度急剧增加，鲁棒性显著降低，可扩展性严重受限。

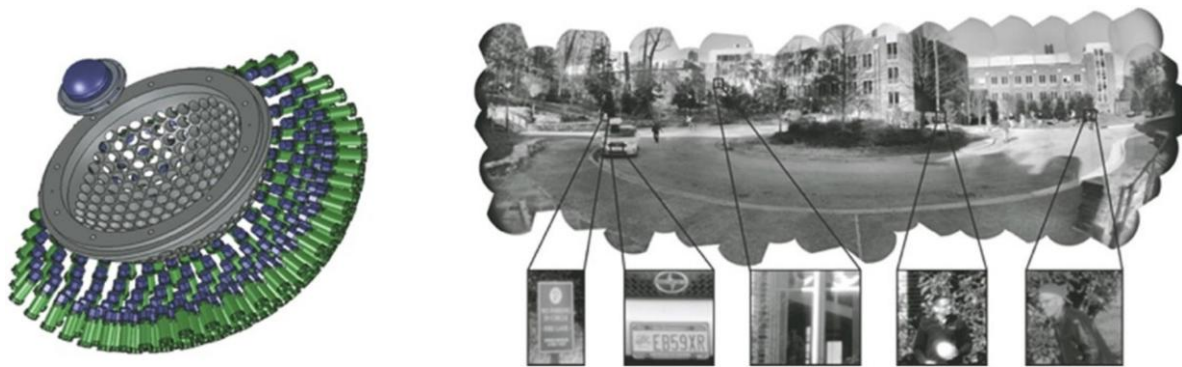


图 2-4 AWARE-2 相机阵列及成像结果[14]

Yuan 等[15]提出了非结构化光场感知算法，构建了“多层异构-同层自适应”的树形感知模型。其中，“多层异构”一改传统结构化阵列像传感器固有的图连接感知模型，解除了光学成像装置物理体积随数据通量线性增长（Scaling Law）的制约，突破了现有多维多尺度感知对庞大复杂硬件系统的依赖。与此同时，“同层非结构”突破了现有像传感器均匀化采样、信号同分布的假设，相邻像传感器间无需受视场相互重叠的严格限制，各像传感器面向场景内容独立感知、并行计算，突破了传统均匀化感知所面临的时空带宽积的瓶颈，实现了像传感器阵列硬件复杂度、数据感知量（信息容量）、维度扩展性的联合优化。该树形结构涉及分层架构，赋予层间异构互联、层内非结构化的特性。其中，分层架构的父层像传感器与子层像传感器之间自然重叠，同层内的各像传感器之间则完全非结构化，无需遵守任何重叠限制，从本质上赋予像传感器阵列极高的灵活性和可扩展性。在此基础上，各个局部像传感器的调整（移除/添加）不会影响其他像传感器，亦不会对多像传感器位姿的优

化带来复杂度剧增和鲁棒性降低的问题。

如图 2-5 所示的非结构化异构像传感器阵列本质上是将多个极低成本相机，以多尺度方式灵活地组合，基于先进的计算成像思想，将硬件的成本以算法替代，实现实用且性价比高的 10 亿像素视频相机。基于该原理的相机系统通过分视场采集突破了空间带宽积的限制，同时通过引入全局尺度相机克服了分视场相机的像差与畸变，并通过跨尺度视频融合算法规避了复杂的相机系统的装配和标定，同时回避了紧密光学部件的设计和制造需求，极大地降低了系统的成本。但是，目前成像性能远未发挥多尺度成像带来的益处，多尺度成像原理和跨尺度映射与融合机制仍未被揭示，系统设计（包括全局低分辨相机和局部高分辨相机各自的设计）的灵活性仍未被充分挖掘，全局低分辨相机与局部高分辨相机的互补性未被充分优化利用，新型、高效、高性能的非结构化像传感器阵列的研究具有广阔的空间。



图 2-5 非结构化异构像传感器阵列[15]

面向大场景多对象的十亿像素级光场智能感知，是智慧城市、无人系统、VR/AR 等应用的核心共性技术。光场感知通量的爆炸式增长，对智能计算的算力与功耗提出了全新的挑战，亟需探索光场智能感知计算的新架构、新理论与新技术。近年来，以衍射神经网络为代表的光学计算架构，虽具有突出的速度与能效优势，但普遍结构简单、规模受限且不易扩展，难以支撑复杂场景中多维光场的智能感知计算。

光场智能感知计算旨在实现大场景多对象“看得全”、“建得清”、“算得快”，是元宇宙智慧文旅、智慧城市建设等应用的核心共性技术。随着摩尔定律放缓，传统硅基电子计算芯片的算力和功耗严重受限，难以满足亿像素级高通量光场智能感知计算的急切需求。探索新的感知技术与计算范式是未来光场视觉计算发展的必然趋势。近年来，以光计算为基础、通过光电融合的方式构建光电神经网络成为国际前沿的热点研究。发挥光电计算高算力、低功耗的特有优势，有望突破传统电子计算长延时、高功耗等瓶颈，实现感算能效的颠覆性提升，引领新一代光场成像理论与技术的变革。

近年来以衍射神经网络为代表的各式光电神经网络不断涌现，然而现有光网络中的

基本光学计算单元受制于固化结构与低扩展性，大多仅能实现简单的单通道前馈网络，难以支撑实际应用中的复杂光场智能感知计算。具体来说，现有光电神经网络无法像电子神经网络一样轻易地部署几十到上百个“通道”以提取有效的特征，也缺少有效的光计算单元来融合多通道的信息，极大地限制了光网络的学习能力和推理性能。

清华大学电子工程系方璐课题组[16]提出了多通道光电智能计算新架构 **Monet**，如图 2-6 所示。其中，编码投影干涉计算单元 **IU**，通过相位和偏振的编码调制以及通道间的光学干涉，实现了特征匹配、加权求和等多通道光学基本计算。提出了干涉计算单元 **IU** 和衍射计算单元 **DU** 交替级联的光电网络模型，实现了光场信息的多通道高效处理，突破了现有光电神经网络结构简单、通道受限等瓶颈，为构建大规模光电神经网络实现复杂光场智能感算提供支撑。

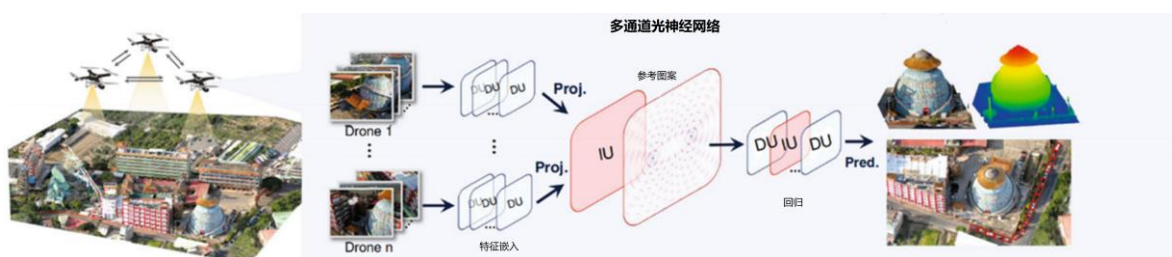


图 2-6 多通道光电智能计算新架构 Monet[16]

2.3.3 全景成像

随着高速通信和人工智能技术的快速发展，人类对现实世界场景的感知不再局限于使用小视场（FoV）和低维场景检测设备。全景成像作为下一代创新的环境感知和测量智能仪器应运而生。然而，在满足大视场摄影成像需求的同时，全景成像仪器有望实现高分辨率、无盲区、小型化、多维智能感知，并可与人工智能方法相结合，走向下一代智能仪器，使对 360° 真实世界周围环境的更深入理解和更全面的感知。幸运的是，自由曲面的最新进展，薄板光学和超表面提供了解决人类对环境感知的创新方法，提供了超越传统光学成像的有前途的想法。全景成像技术旨在捕捉 360° 全方位的场景，以实现对整个空间环境的全面感知。这一技术通过将多张图片或视频片段拼接成一个完整的 360° 视图来实现。全景成像技术广泛应用于虚拟现实（VR）、增强现实（AR）、360° 视频制作、全景地图制作等领域。这些应用不仅提供了沉浸式的视觉体验，还为环境感知和测量提供了强大的工具。

全景成像发展到现在，已成为了一项相对成熟的技术，就目前而言，全景成像方法

[17]主要有三种：拼接式全景成像、鱼眼全景成像以及折反射全景成像。

拼接式全景成像是全景成像技术刚提出时所采用的方法，通过用传统的相机拍摄得到场景的多幅图像，然后基于拼接、融合技术，合成得到场景的全景图像。拼接式全景成像的具体方法有两种，一种是采用一个相机，通过旋转相机对不同方向进行拍摄，从而获得多幅场景图像。这种方法简单易行，能够获得高分辨率的全景图像，但由于成像速度缓慢，不具有实时性；另一种是采用多个相机，同时对多个方向同时进行拍摄，从而获得多幅场景图像。这种方法能够满足实时性的要求，也能获得高分辨率的全景图像。两种方法原理都是采集得到包含整个场景信息的多幅图像，最后再合成得到全景图像。

鱼眼全景成像利用超广角物镜来实现全景成像，超广角物镜的焦距非常短，能够获得接近 180° 视场角的半球面场景。而且，由于它采用凝视方式进行拍摄，不需要旋转移动，因此具有较好的隐蔽性。然而，这种全景成像方法存在缺点，它拍摄获得的图像存在极大的桶形畸变，形成鱼眼效果。因此，想要对拍摄得到的鱼眼全景图像进行显示观察，需要事先进行畸变校正，将之校正为符合人眼观察习惯的全景图像。目前，鱼眼全景成像技术已经比较成熟，大量的鱼眼全景成像设备出现在市场。

折反射全景成像是将常规成像设备和反射光学元件相结合的成像技术，利用反射光学元件的反射特性来扩大成像设备的视场，从而能够获得水平方向 360° 和垂直方向一定角度的全景图像。这里的反射光学元件通常指的是各种类型的反射镜，如球面反射镜、圆锥面反射镜、椭球面反射镜、抛物面反射镜以及双曲面反射镜等。根据投影模型的不同，折反射全成像系统可分为基于透视投影成像模型的系统 and 基于平行投影成像模型的系统。根据是否满足单视点成像约束，又可分为单视点成像系统和非单视点成像系统由于折反射全景成像能够获得大于半球空间的大视场，同时系统结构较为简单，设计较为灵活，成本也比较低，近年来得到了越来越广泛的研究和应用，已经成为了当前全景成像技术研究的主流方向。

2.3.4 多视图立体成像

多视图立体视觉 (MVS) 一直是计算机视觉研究的一个热点。它的目的是从多个已知相机姿态的图像中建立密集的对对应关系，从而产生稠密的三维点云重建结果。

在过去的几年里，人们在提高稠密三维重建的质量上付出了很大的努力，一些传统的几何算法通过描述同一场景从不同视角下获取的多幅图像与物体之间投影关系的几

何模型进行三维重建，如 PMVS、GIPUMA 以及 COLMAP 等取得了令人印象深刻的效果。而近年来深度学习也在多视图重建中取得了非凡的性能，如 MVSNet、CasMVSNet 等。然而，在三维重建任务中，由于数据量大、弱纹理、遮挡、反射等问题，如何高效准确地实现多视图立体视觉仍然是一个具有挑战性的任务。

基于深度学习[18]的三维重建最近几年取得了非常大进展，是当前计算机视觉领域比较流行的方法之一。学习系统是由多层卷积层和多层全连接层组成的卷积神经网络（CNN）进行学习图像的层次化特征表示，这种方法能够精确地重建出物体的几何形状。与传统的三维重建相比较，基于深度学习的三维重建技术不需要进行特征提取与匹配以及复杂的几何运算，使用深度学习的方法更加方便，使得实时性和鲁棒性进一步提高。

基于语义的三维重建可以运用在移动的行人或车辆等大的场景，这种方法能够精确地对环境中的目标物体进行识别，而深度学习技术也是最近几年刚刚兴起的比较有优势的识别方法，因此，深度学习和语义相结合的三维重建是未来几年的研究趋势，也会受到该领域的研究者们广泛关注。Savinov 等[19]提出了一个稠密的语义三维重建方法。Blaha 等[20]提出了一个自适应多分辨率的语义三维重建方法。该方法的目的是在给定一组场景图像中进行 3D 场景稠密重建，并分割成语义目标类。Sunderhauf 等[21]提出了面向对象的语义建图方法，该方法的主要目的是构建环境的 3D 点云地图，然后再通过卷积神经网络对关键帧图像中的物体进行识别检测，最后对 3D 目标物体进行点云分割，从而在地图中更新或添加目标物体的 3D 点云等信息。

传统的基于明确表面的三维重建方法（MVS），尽管在获取大规模结构方面表现出色，但往往难以捕获场景的微观细节。这类方法在构建三维模型时依赖于高分辨率的离散化过程，这限制了它们在处理细微纹理和小尺度特征时的有效性。此外，这些技术通常使用 UV 映射来纹理化模型，但在复杂或广阔的场景中，UV 映射易受到畸变影响，这会在渲染过程中产生可见的伪影和质量退化。

隐式体积表示方法（如基于神经辐射场的 NeRF）提供了一种表达高空间分辨率的新途径，通过对整个体积场进行连续建模，能够从任意新视角渲染出细致的图像。然而，这些方法在实际操作中需要处理大量的体积数据，这不仅计算成本高，而且在大规模场景中效率低下。更重要的是，由于体积渲染依赖于沿视线对样本的积分，它在处理复杂场景中的视图不一致性时常常导致渲染质量下降，例如在边缘或细节处出现模糊和失真。

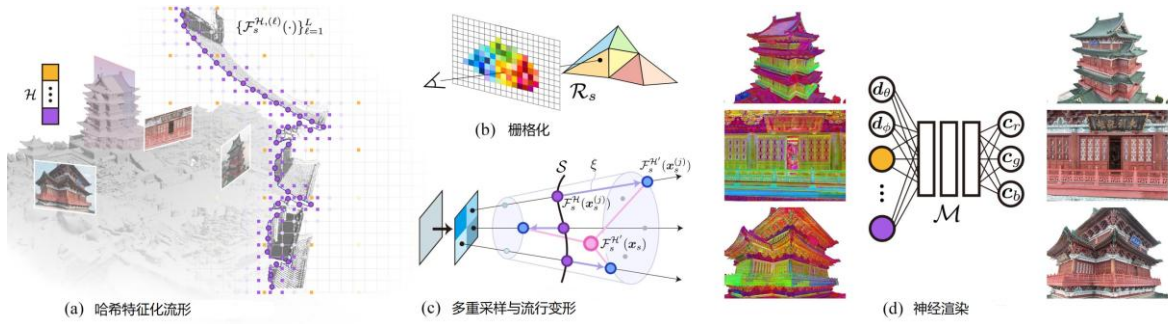


图 2-7 基于哈希特征流形表征的神经渲染框架[22]

为了解决以上挑战，XScale-NVS[22]研究提出了哈希特征化流形的表示方法，通过对表面流形进行栅格化，显式地优先考虑多视角一致性，充分释放了体积哈希编码的表达能力，如图 2-7 所示。同时设计了一个延迟神经渲染框架，以高效解码表示，并提出了两种定制设计来更好地描述跨尺度细节。另外还引入了 GigaNVS 数据集，用于评估实际大规模场景的跨尺度、高分辨率新视角合成，XScale-NVS 在 GigaNVS 上相对于 SOTA 平均 LPIPS 降低了 40%，将野生环境下跨尺度神经渲染的边界推向了前所未有的细节和真实感水平。

Meta AI[23]提出了一个统一的几何感知框架 VGGT (View Generalizable Geometry Transformer)，可以从任意数量的视图中直接推理出相机参数、点云、深度图、三维轨迹等全部关键 3D 属性，重建时间大幅降至 1 秒以内，并在多个任务中刷新了性能指标。VGGT 不仅可用于标准几何任务，在作为特征骨干网络时，同样能大幅提升点跟踪与新视角合成等下游任务表现，被认为是“通用 3D 视觉模型”迈出的关键一步。

在人体光场感知领域，Light Stage[24]是当前数字人高真实感重建的一个标杆技术。它通过构建相机阵列，使用多角度、高精度照片来重建三维空间中的 CG 模型。这种技术能够还原拍摄人物的三维结构，并获取面部的反射信息，从而在不同环境光下重构数字人模型光效。Light Stage 技术已经应用于好莱坞电影渲染中，从第一代系统 Light Stage 1 至今已经升级到 Light Stage 6，最新的一代系统命名为 Light Stage X。

Google 研发的 Lightstage 系统采用了 331 个可编程光源以及一个 90 台 12.4M 像素的工业相机组成的相机阵列。每个光源包含有 63 个高亮 LED 灯珠，分为不同的颜色和不同的发散角。这个系统能够结合主动与被动传感器，以取得最佳的建模效果。相机系统工作在 60Hz，相机系统与灯光系统时序匹配，以获取三维模型和相应的材质贴图。这样的系统，每秒钟的数据量达到惊人的 63.5GB。

元客视界研发了 LuStage 数字人元数据采集系统，如图 2-8 所示，系统包含 776 个

6 通道彩色 LED 光源阵列、16 台 RGB-D 深度传感器以及 150 个视点的相机阵列，相机分辨率最高达 8K，采集帧率最高达 450fps。其中深度传感器采用近红外结构光重建技术辅助实现至少 30fps 的动态稠密 3D 重建；多视点相机阵列用于采集各种模拟光照环境下人体的空间视频原始数据，以便采用 3D 高斯等技术进行空间视频重构；高帧率相机用于采集高速动态数据集，以采用 OLAT 技术对数据集进行重光照编辑，理论上支持数据集对环境光照变化条件的无限扩展。系统还创新性地提出了全局动态积分光场成像技术，计算中控系统按照场景内的实时照度需求计算每个曝光周期内各个 LED 光源的积分光强，换算成 LED 的光通量和频闪脉宽，每个 LED 光源支持实时变通量变脉宽控制，以此实现分布视点动态 HDR 成像，大幅提升了系统对于明暗细节高动态范围成像的能力以及对室内室外各种环境光照特别是日光环境的模拟能力。

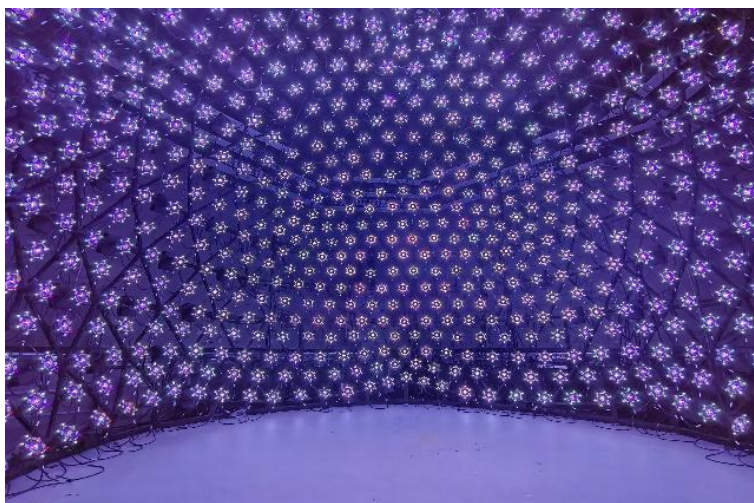


图 2-8 LuStage 数字人元数据采集系统

虚拟数字人技术的发展已经进入快速成长期，特别是 NeRF 和 3DGS 技术的出现，带动了数字人重构研究工作蓬勃发展，并将数字分身创作推向了应用的前沿。

2.3.5 音频感知

音频感知是元宇宙多模态感知的重要一环，音频感知主要分为前端信号处理、语音识别、声纹识别、情感识别、语音合成等多个模块，如图 2-9 所示。

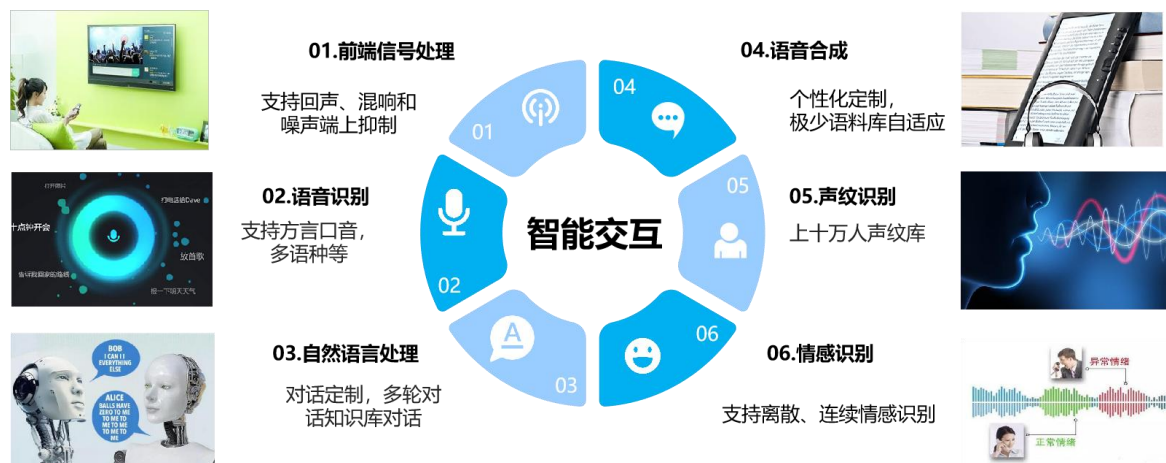


图 2-9 音频感知交互

语音前端信号处理对于提高音频感知的鲁棒性起到了非常重要的作用；通过语音增强模块抑制各种干扰，使待识别的语音更干净；语音增强模块通过回声消除、噪声抑制、去混响提高语音交互的鲁棒性；真实环境中包含着背景噪声、人声、混响、回声等多种干扰源，上述因素组合到一起，使得这一问题更具挑战性。语音前端信号处理模块包括多通道模块以及单通道模块。对于麦克风阵列采集，当前研究更倾向于将传统的时频波束形成（beamforming）与 DNN 驱动的估计器结合——例如用神经网络估计声源/干扰协方差或掩码，再通过 MVDR/GEV 等波束形成器实现稳健增强；此外“可微分波束形成器/神经波束器”允许将空间滤波器参数纳入端到端训练流程，从而显著改善在复杂声场（多声源、非平稳噪声、强混响）下的识别性能。同时，深度学习方法（如深滤波 deep-filtering、端到端时域网络）在抑制晚期混响与近端回声方面取得了显著进步；同时出现了更高效的实时/边缘实现方案以满足元宇宙场景对低延迟的要求。对于资源受限设备，可以采用轻量化网络或模型蒸馏/量化技术减少计算并保持音质与可懂度。需要强调的是可以先进行多通道混响消除再进行波束形成，也可以先进行波束形成再进行单通道混响消除。经过上述处理后的单路语音进行后置滤波消除残留的音乐噪声，然后通过自动增益算法调节各个频带的能量后作为前端处理的输出，将输出的音频传递给后端进行音频感知。

语音识别是音频内容感知的重要一环。语音识别系统主要包括四个部分：特征提取、声学模型、语言模型和解码搜索，具体如图所示。自动语音识别的过程是先对前端采集的语音信号进行处理，以提取声学特征，再将声学特征作为声学模型的输入，计算声学得分，然后结合语言模型的语言得分，在解码空间中搜索最佳路径，从而识别出语音信号对应的文本。经典的混合式（Hybrid）语音识别系统的构成，包括声学特征提取，声

学模型，语言模型，以及解码搜索算法。信号处理模块将语音波形提取成更具有代表性的声学特征序列，以便后续处理。经典的声学模型基于隐含马尔可夫模型构建，其中隐含马尔可夫模型的观测概率通过高斯混合模型或深度神经网络估计。语言模型估计给定前文的情况下，下一个词发生的概率。结合声学模型和语言模型给出的分数，解码器搜索出给定声学特征序列，概率最大的词序列。以发音词典为桥梁，声学模型和语言模型被结合起来形成解码搜索空间。加权有限状态转换器是构建解码搜索空间的重要工具。除了基于隐马尔可夫模型的声学模型，相关技术基于联结主义时序分类的端到端声学模型，这种声学模型不需要逐帧对应的标签，而是直接训练神经网络。随着深度学习技术的发展，语音识别系统正快速从传统 Hybrid 模型向端到端架构和“大规模语音基座模型（Speech Foundation Models）”演进。Conformer、Transducer 与 CTC/Attention 等端到端架构已成为工业主流，而 wav2vec 2.0、HuBERT、WavLM 与近期的 UniSpeech/Whisper 系列等自监督或多任务预训练模型，为噪声鲁棒性、跨语言迁移与低资源学习建立了新的基线。更重要的是，2024 - 2025 年出现大量语音大模型（Speech-LLM），如 SpeechGPT、Qwen-Audio、Whisper-large-v3、Meta Audio-Omni，它们可在统一框架下执行语音识别、音频问答、说话人属性分析、口语理解、语音翻译与语音合成等多任务能力，实现“听 - 理解 - 说”的端到端语音智能。这类模型通常采用音频编码器+大语言模型+语音解码器的结构，使音频理解与文本推理紧密耦合，大幅提升长语音、多说话人、跨语言、跨模态等复杂场景下的表现。随着这些大模型与工程化工具（如 WhisperX、NeMo ASR/TTS、实时流式推理框架）结合，长录音转写、智能对话、多模态内容理解与沉浸式交互正在逐渐实现实用化，大模型也使音频感知从“识别为中心”迈向“理解与生成统一建模”的新时代。

声纹识别是音频深度感知的重要研究方向，其主要目的是在语音中自动获取说话人身份。早期的声纹识别主要采用有效的声学特征参数和模式匹配的方法，匹配往往通过特征矢量之间的距离测度来实现。在说话人训练阶段，系统首先对训练语音进行静音剔除和降噪处理，尽可能得到纯净有效的语音片段，然后再提取语音对应的声学特征参数，根据系统建模算法，得到说话人的特征模型，每个说话人的训练语音经过训练阶段后得到一个说话人模型。随着深度学习的发展，基于深度网络的特征学习方法成为了声纹识别的主要方法，利用复杂非线性结构赋予的特征提取能力，能自动对输入的语音信号进行特征分析，提取出更高层、更抽象的说话人声纹表征，如 d-vector[25]、x-vector[26]。而后，谷歌公司的 Heigold 等人提出了端到端声纹识别系统，端到端的网络包含两部分：

预先训练好的特征提取网络和用于决策打分的判决网络，输入为不同说话人的语音信号，输出即为说话人识别结果，之后如注意力机制、自适应方法等在端到端系统中的应用进一步提高了系统的性能。近年来，自监督学习成为声纹识别领域的重要方向。自监督学习方法如 wav2vec 2.0 和 HuBERT 等模型通过大量无标注数据进行预训练，能够捕获语音的潜在特征并提升系统在少数据情况下的识别能力。这些模型不仅减少了对大规模有标签数据的依赖，还极大地增强了模型在多语言、多域环境下的泛化能力。

情感识别是实现深度音频感知不可或缺的关键一环，通过声音这一媒介，让机器深入感知并理解人的情感变化。语音情感识别系统主要由三部分组成：语音信号采集、语音情感特征提取和语音情感识别模型训练。这三部分相互协作，共同支撑从原始语音信号到情感状态描述的完整过程。语音信号采集模块通过语音传感器（例如麦克风等语音录制设备）来获取语音信号。这些设备能够捕捉到声音的细微变化，将人类的情感波动转化为可处理的电信号。采集到的语音信号随后被传递到语音情感特征表征模块。在这一模块中，系统会对语音信号中与情感关联紧密的声学线索进行抽取，这些线索中可能蕴含声音的音调、音量、节奏、音色等多个方面。通过分析这些线索，系统能够初步判断出语音中蕴含的情感倾向。所抽取到的情感特征被送入语音情感识别模块，该模块的核心任务是对被测语音的情感状态进行描述。为了更有效地完成这一任务，系统需要借助当前主流的音频预训练模型来进行语音情感特征抽取。这些预训练模型包括了 Wav2Vec、Hubert、Whisper 等，它们能够通过海量大数据中蕴含的情感信息，学习到数据的本质结构或分布，从复杂的语音信号中提取出最具代表性的情感特征。随着人工智能技术的不断发展，大语言模型在语音情感识别领域的应用也日益广泛。这些模型可以替代传统的统计机器学习模型和深度神经网络，不仅能够对情感状态进行分类和描述，还能够对情感产生的原因进行可信溯因。例如，经过微调后的 QWen 系列大模型，就具备强大的情感推理能力。它们能够结合上下文信息，分析出语音中情感变化的深层原因，从而实现语音情感分析从感知分类到认知推理的跨越。这种跨越意味着语音情感识别系统不再仅仅局限于对情感状态的简单分类，而是能够深入到情感的认知层面，理解情感背后的动机和原因。这对于提升人机交互的体验、推动情感计算技术的发展都具有重要意义。

语音合成是实现音频感知中的关键表达环节。随着深度学习的发展，目前语音合成技术已经发展相对成熟，且已取得了广泛应用，但仍然存在音色风格单一，数据成本，构建时间等诸多限制。一个好的语音合成系统建立在一个人工精心设计的语音数据库之

上,一个音素均衡、高音质的语音合成数据库制作成本较高,要在专业录音环境下录制,同时需要具有相应语音学背景知识的专业人员进行标注,整体设计制作时间较长。较高的语音合成数据库门槛限制了语音合成的输出风格。就语音合成系统的声学模型而言,微软、IBM、谷歌、百度和讯飞等采用了几十个小时以上的训练数据。然而大多数语言的资源是有限的,标注成本也极其昂贵。为了丰富语音合成系统的音色,为了每个音色都花费较长时间对数据和模型进行定制是不经济也是不现实的。为新的说话人快速搭建个性化语音合成系统时,首先面临的问题就是用来训练合成模型的数据较少,这导致不能用传统的方法去训练模型,如何用少量数据构建高性能的语音合成系统成为一个极富挑战的研究课题,同时也具有极为迫切的应用需求。2021年,扩散模型在语音合成任务中占据了重要地位。DiffWave 是扩散模型在语音合成中的典型应用,它使用扩散过程从随机噪声生成高保真音频,并展现出接近甚至超越 WaveNet 的音质表现。扩散模型的生成过程相比传统的序列模型具有更高的灵活性,且其并行化潜力使得其在大规模语音合成任务中具备更高的效率。2022年,Codec 技术不仅用于压缩和解压缩音频信号,还能够用于构建高效的音频生成系统。例如,EnCodec 是 Meta 推出的一种基于神经网络的高效音频压缩框架。该系统不仅能够在较低的比特率下保持高音质,还可以用于实时语音合成和音频生成。自 2023 年起,一系列基于语言模型的语音合成框架被提出,如 VALL-E 和 Cosyvoice 等,这些模型在减少语音失真、提高语音自然度和保持个性化声音特征方面都展现了强大的能力,特别是在少样本语音合成和多语言支持方面,表现十分突出。

2.3.6 触觉感知

触觉在具身智能发展的浪潮下,视触觉传感技术作为连接物理世界与数字感知的关键桥梁,正逐步改变人机交互的边界[27]。该技术通过触觉信息可视化方式,显著提升了触觉数据的可解释性与交互性,从而有效打破了虚拟与现实之间的感知壁垒[28]。借助高分辨率的视觉反馈机制,视触觉传感能够对压力、形变、纹理、滑动等多维触觉信息进行精准捕捉与数字化模拟,进而补全了元宇宙中“视-触”感官闭环,如图 2-10 所示。

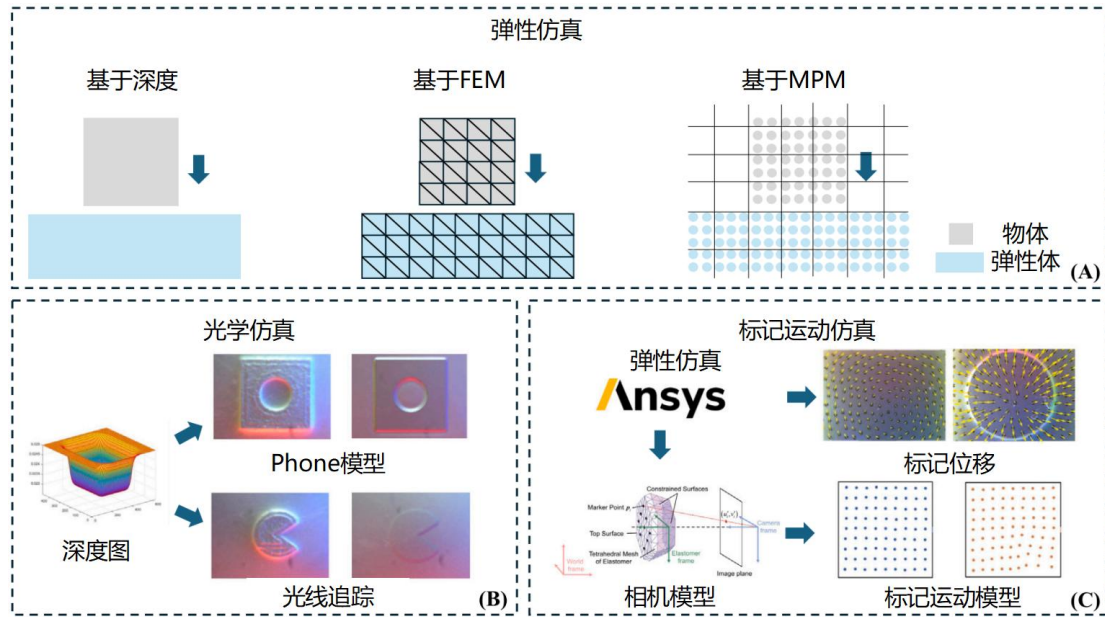


图 2-10 视触觉仿真概览

触觉仿真作为实现这一目标的核心手段，为虚拟场景注入了可感知的物理质感，显著增强了元宇宙的沉浸感与用户参与度。当前视触觉仿真技术主要分为物理建模与数据驱动两大类[29]。物理建模进一步细分为弹性体仿真（如基于深度法[30]、有限元法 FEM[31]、材料点法 MPM[32]）、光学仿真（如 Phong 模型[32]、路径追踪[33]）以及标记点运动仿真（如基于线性位移关系的方法[34]和基于相机模型的方法[35]）；数据驱动包括跨模态生成[36]与虚实迁移生成[37]等方法。

为了构建起贴合人类感知逻辑的触觉描述体系，触觉多模态表征通过整合触觉本征维度与视觉纹理、物体力学属性等跨模态关联信息，显著提升了触觉感知的语义丰富性与物理一致性。传统方法多依赖单一触觉信号，难以还原复杂交互过程中的材质细节、动态变化与上下文语义，而近年来提出的多模态学习方法，如 Touch100k[38]中引入的多粒度语言描述，将触觉感受与自然语言紧密结合，增强了触觉的语义表达能力。同时，UniTouch[39]和 AnyTouch[40]等模型通过跨模态对齐技术，将触觉嵌入与视觉、语言等模态进行联合表示，显著提升了模型在零样本理解、跨传感器泛化等方面的表现。此外，AnyTouch 进一步引入静态与动态触觉信息的统一表示，结合多传感器融合策略，有效建模了触觉在时间维度上的演变过程，从而更好地模拟人类在真实交互中的触觉认知机制。



图 2-11 基于触觉传感的人机交互

借助视触觉传感技术，研究界搭建起人机高效协同的交互桥梁。**OneTip**[41]在元宇宙环境中实现接触形变的连续重建，可输出稳定的六维触觉姿态，用于人机输入与元宇宙虚拟操控，如图 2-11 所示。结合神经场，进一步将高分辨率视-触觉图像与本体感知融合，在线构建物体的神经场形状与姿态，实现“操作即建模”的动态触觉感知模式[42]，如图 2-12 所示。基于 TIR 的视触觉盲文识别系统能够将现实世界中的触觉符号即时映射到虚拟空间，使触觉内容在数字环境中可读、可交流、可共享，强化了元宇宙的普适交互能力[43]。视触觉传感已逐步形成从“连续形变重建-动态交互建模-精细模式识别”的完整路线，可为元宇宙提供从触觉感知到语义理解的全流程支撑。

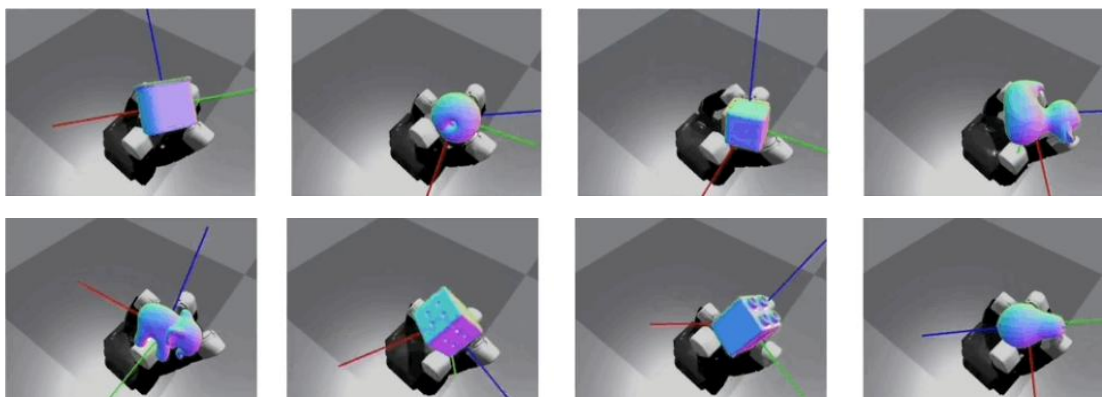


图 2-12 元宇宙中自主生成大规模触觉数据

2.4 元宇宙生成

数字化三维空间是元宇宙的基本形式之一，不仅可以三维重建现实三维世界并与其信息融合，服务于城市治理、工厂运维和智能交通等场景，而且可以通过人工智能技术生成和创造出新的数字化空间，为人类社会的未来形态提供想象空间。与此对应，元宇

宙生成技术也包括三维重建和智能生成两大类。三维重建是指通过各类光学传感器，包括图像、激光点云、深度相机或事件相机等，获取现实世界的视觉信息，通过传感器位姿计算注册融合传感器数据，构建现实世界的精准数字化映射，其典型技术包括传统的运动结构重建（Structure from Motion）、多视点立体视觉（Multiview Stereo Vision）、光度立体法（Photometric Stereo）等。近年来，随着可微渲染技术的发展，神经网络隐式表达成为了新的研究热点，该方法采用神经网络表达以三维位置为变量的三维颜色和密度场，集合可微体渲染和光度损失函数从拍摄图像中端到端的重建三维场景，极大地简化了三维重建的流程，并提高了渲染的真实感。研究人员进一步提出采用空间离散点采样表达三维光场，代表性工作为最新的三维高斯泼溅（3DGS）技术，极大提高了神经网络隐式表达的渲染速度。三维重建以及现有的三维制作软件积累了大量的三维数据，为生成式人工智能模型自动生成三维内容奠定了数据基础。生成对抗模型（Generative Adversarial Network）、扩散模型（Diffusion Model）是典型的生成式人工智能模型，在学习数据分布上展现了强劲能力，在三维内容生成上应用广泛。通过在 Stable Diffusion 开源的 Objaverse 数据集上的训练扩散模型，已经出现了 Craftsman3D、Rodin-Gen1 等较高质量的三维生成模型，为元宇宙生成提供了新的动力。与此同时，“世界模型（World Model）”正成为元宇宙生成技术的重要方向。其通过统一建模空间结构、物体关系与行为逻辑，使虚拟世界具备理解、推理和预测能力。借助大模型的跨模态泛化，世界模型将推动虚拟环境从静态呈现迈向可演化、可交互的智能系统。

2.4.1 三维重建

三维重建技术是利用二维投影或影像来恢复物体/场景的三维信息的数学过程和计算机技术，通过三维重建技术可以为元宇宙提供三维物体模型和孪生虚拟的三维空间，如图 2-13 所示。三维重建从技术上可以分为传统 SFM&MVS 构建和近年来被广泛研究的可微渲染引导（NeRF，3DGS）的三维重建两大类。

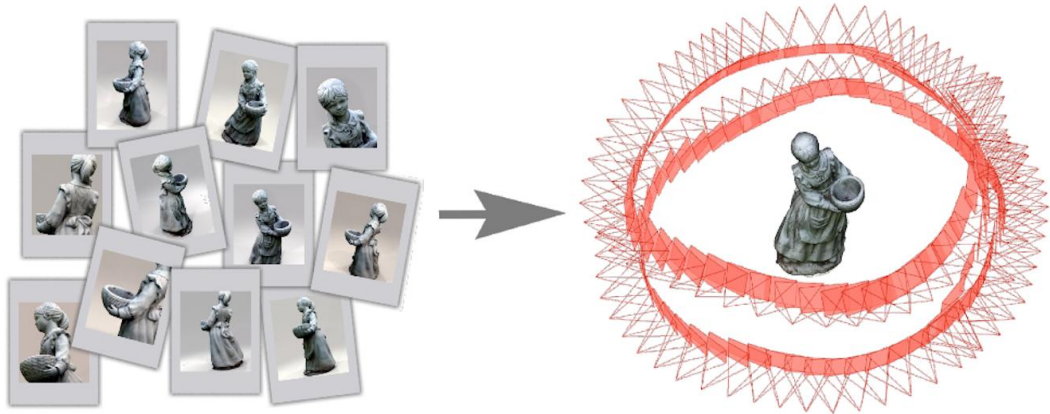


图 2-13 三维重建的概念

(1) 传统 SFM&MVS 重建

SFM 的基本原理是通过分析多张不同视角的二维图像，基于多视角几何的原理恢复出场景的三维稀疏点云及相机的运动轨迹，主要包括特征检测与匹配、相机姿态估计和三维点云重建这几个过程。MVS 是在 SFM 生成的稀疏点云基础上，通过密集匹配技术生成稠密点云，包括图像对的选择、视差估计和点云优化几个过程。完成稠密点云估计后，在元宇宙三维应用中，一般还需要通过 Mesh 提取和贴图技术实现应用生态中标准三维模型的输出，如图 2-14 所示。传统 SFM&MVS 重建技术经过多年的发展已经比较成熟，在三维构建软硬件系统中有着广泛的应用，例如手持扫描仪、无人机倾斜摄影建模等，也出现了采用光滑性等各种先验和分块等各种方法优化 SFM&MVS 来进一步提高其三维建模效果。

在 SFM&MVS 重建的过程中，最重要的环节是图像之间点对的匹配问题，一般容易受到光照变化、重复纹理、光滑或无纹理表面的影响导致重建效果不佳，近年来有大量的基于深度学习特征匹配度方法例如 SuperPoint、SuperGlue 等来提升这方面的性能，但面对光滑或无纹理等问题仍然存在应用问题。为了解决这些问题，业界采用了各种方案包括基于深度或激光点云融合的方案、基于辅助标记的方案等，这些方法可以有效减少上述算法由于原理上的限制导致的建模效果不佳的问题，在大量三维扫描重建设备上被应用。

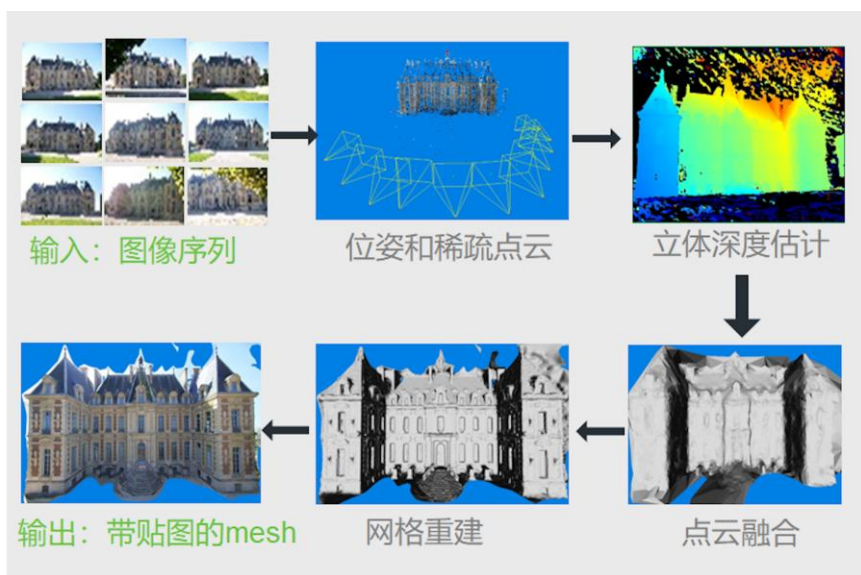


图 2-14 SFM&MVS 三维重建技术的流程

(2) 可微渲染引导的三维重建

可微渲染引导的三维重建（Differentiable Rendering based 3D Reconstruction）是一种近几年迅速发展的新兴的重建方法，它通过结合计算机图形学中的渲染技术和深度学习中的自动微分技术，来改进三维重建的效果和效率，其与传统方法的区别如图 2-15 所示。该方法利用可微渲染器计算图像生成过程的梯度信息，从而在三维重建中进行更精确的优化，将三维重建任务从几何计算问题转化为一个学习训练的问题，改变了三维建模的范式，显著提高了三维重建模型的拟真度（Fidelity）。目前最为代表性的方法包括神经渲染辐射场（NeRF）重建和 3D 高斯辐射场（3DGS）重建这两种。

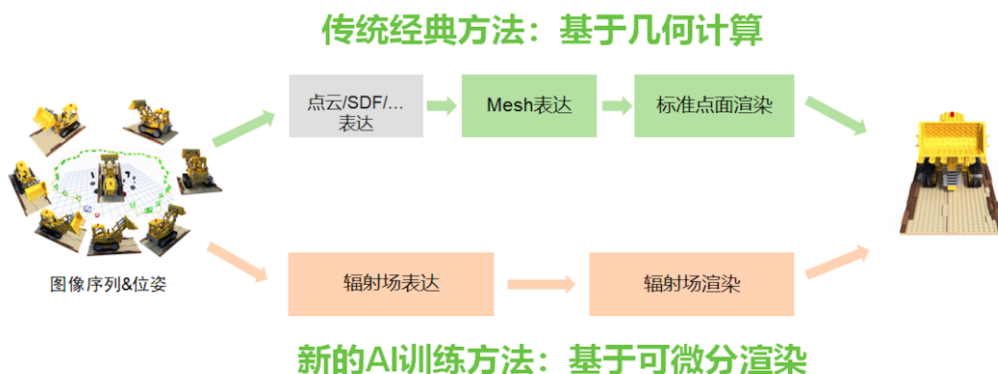


图 2-15 可微渲染引导的三维重建：将建模变成 AI 训练的过程

- 神经渲染辐射场（NeRF）重建

神经渲染辐射场（Neural Radiance Fields, NeRF）是一种基于神经网络的三维重建方法。它利用神经网络来隐式表示场景中的颜色和密度信息，并通过体渲染技术将这些信息转换为二维图像。其重建过程是通过优化神经网络参数，使得渲染生成的图像与输入图像匹配，从而重建出场景的三维结构，大致流程如图 2-16 所示。NeRF 算法过程主要包括坐标采样、网络推理、体渲染、误差计算、梯度优化这几个关键步骤。该过程中首先从三维空间中均匀采样点，这些点的坐标作为神经网络的输入，然后将采样点的坐标输入到神经网络中，网络输出该点的体素密度和颜色，再使用体渲染公式将密度和颜色沿射线进行积分，生成最终的像素颜色，将渲染生成的图像与输入图像进行对比，计算误差并通过反向传播算法，根据误差调整网络参数，逐步优化三维场景的表示。

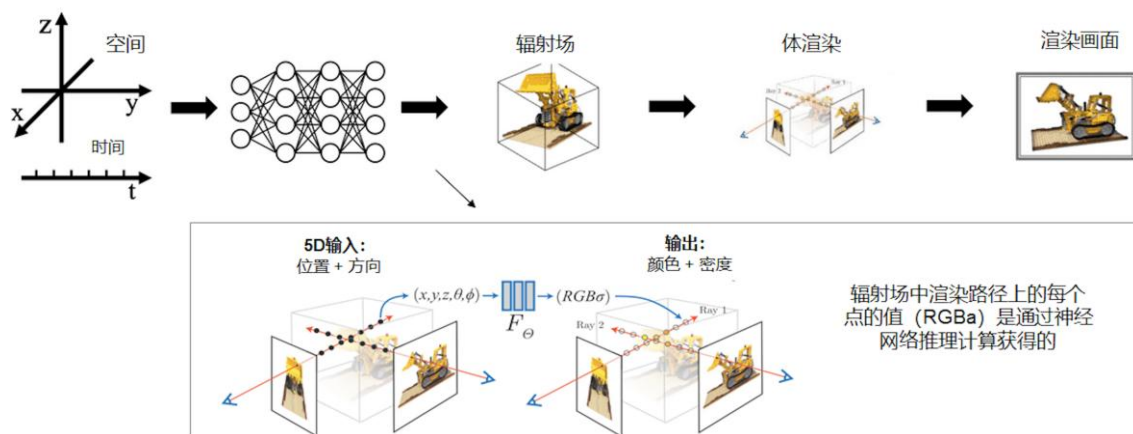


图 2-16 神经渲染辐射场（NeRF）重建

NeRF 能够生成高分辨率和细节丰富的高拟真三维重建结果，特别是对于复杂纹理和细微结构的场景，但是通过 NeRF 建模好的三维模型是一个神经网络模型，不同于传统的点云和 Mesh 表示，它是一种隐式表示，因此不能和现有渲染引擎和软件生态兼容，在当前阶段影响了其广泛的应用。随着基于 NeRF 的大量的相关算法的研究和成熟和其高质量的优点，在未来有可能变成工业三维模型软件生态的一种，目前业界也有开发一些基于 nReal 和 Unity 的软件插件来实现兼容性渲染。

• 3D 高斯泼建（3DGS）方法

3D 高斯泼建（3D Gaussian Splatting, 3DGS）是一种新兴的三维重建方法。与 NeRF 构建的神经网络整体表达不同，该方法将三维空间中的点拓展为三维高斯分布，并在每个点上附加球谐系数表达其随视点变化的颜色，并通过优化这些高斯分布和球谐系数来实现高质量的三维重建，如图 2-17 所示。3DGS 方法重建的过程主要包括初始点云生成、

高斯分布初始化、高斯渲染、误差计算和梯度优化这几个关键步骤。首先，通过使用传统三维重建方法（如 SFM 或 MVS）生成初始的三维点云，然后将初始点云中的每个点表示为一个三维高斯分布，初始参数可以包括位置、形状（方差）和颜色，使用高斯渲染技术，将三维高斯分布投影到图像平面生成合成图像，再将合成图像与输入图像进行对比来计算误差，最后使用反向传播算法，根据误差调整高斯分布的参数，最终使渲染图像逐步逼近真实图像。

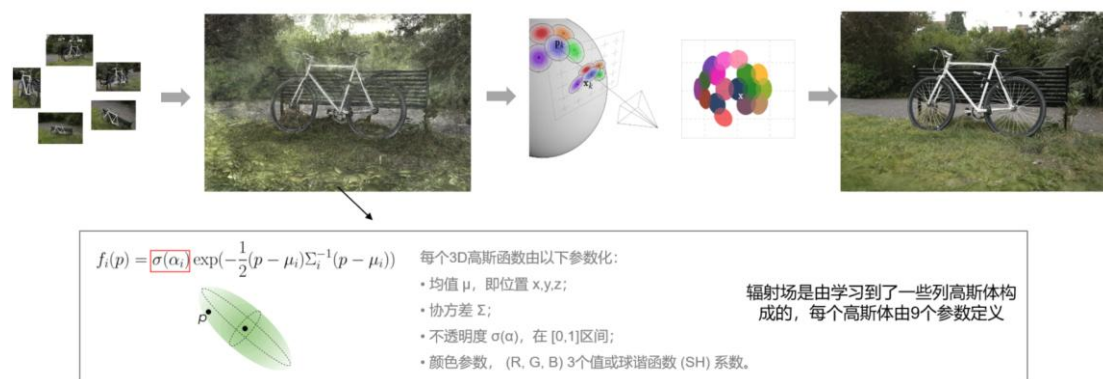


图 2-17 3D 高斯辐射场（3DGS）重建

3D 高斯方法兼顾了显式表达、高拟真重建、实时渲染的优势，受到了广泛的关注，在 2023 年提出之后得到了快速发展。但是，3D 高斯方法依赖于初始三维点云的构建，在实际应用中会面对传统三维重建方法的相同的问题例如弱纹理、重复纹理等问题导致无法实现初始点云和相机位姿的估计，因此有部分产品方案采用基于激光点云和深度点云初始化的方法来解决该问题。得益于其优秀的计算性能和渲染质量，业界部分元宇宙平台、三维应用中已经集成了 3D 高斯方法重建的模型。

（3）前馈式三维重建（Feed-Forward 3D Reconstruction）

前馈式三维重建利用深度学习模型，尤其是 Transformer 架构，直接从 2D 图像中端到端高效推断场景 3D 结构、深度图和相机姿态，能够鲁棒应对图像中光照变化以及图像间重叠度变化，显著提升了基于优化的 Structure from Motion (SfM) 或 Multi-View Stereo (MVS) 方法的速度和鲁棒性，节省了 NeRF 和 3DGS 需要位姿带来的额外开销。Dust3R[44]是这一方向的开创性工作，该方法采用 ViT 编码器将图像编码成 Token 序列，利用交叉注意力机制从输入的两幅图像中预测三维点云，并通过三维注册计算相对的相机位姿，对于大的场景可以依次输入图像进行重建。IEEE CVPR 2025 年的最佳论文 VGGT(Vision Geometry Guided Transformer)[45]进一步拓展了该方法，可以从无

序图像输入中推断三维结构。VGGT 设计了相机和注册的 Token 以取得相较于图像输入第一帧的相对相机位姿，并支持了下游点对应轨迹的提取等应用。

2.4.2 数字人分身重构

数字人是元宇宙数字内容中的必不可少的元素，是连接现实与虚拟的桥梁，可以让元宇宙变得更加多元化、更加生动。因此，数字分身重构是元宇宙生成中的技术研究热点。元宇宙内容对数字分身的需求通常有两种，一种是可驱动可交互的模型，另一种是可自由视点交互的空间视频。前者的实现通常是先 3D 重构数字分身模型，再利用动画驱动生成动态内容；后者的实现通常是基于多视点视频生成可自由视点交互的空间视频。

传统的人物 3D 重构是使用三维软件创建三维模型的过程。在建模过程中，艺术家使用各种工具和技术来创建模型，包括几何体、网格和曲面等。这些工具可以用来构建复杂的三维模型，并可以对其进行纹理贴图、添加材质和光照等操作，以便在最终作品中呈现出逼真的效果。传统的人物 3D 重构采用几何面片模型表征以适配传统的 3D 内容制作和渲染管线，对模型几何精度、材质贴图精度、光照匹配以及渲染引擎都有严格的要求以保证输出人物的逼真度，代价是高昂的制作成本。

由于 3DGS[46]方法能够做到实时高质量渲染，在数字分身重构上也得到了迅速应用。D3GA[47]首先使用可驱动的 3DGS 和四面体笼来创建可动画化的人类化身，这实现了有前景的几何和外观建模。为了捕捉更多的动态细节，SplatArmor[48]利用两个不同的 MLP 来预测建立在 SMPL 和规范空间上的大幅运动，并允许所提出的 SE(3) 场产生与姿势相关的效果，从而实现更详细的结果。HuGS[49]使用线性混合蒙皮和基于局部学习的细化，创建了一个从粗到细的变形模块，用于基于 3DGS 构建和动画化虚拟人化身，它能够实现 20 FPS 的人物神经渲染性能。由于 3DGS 学习了大量的高斯椭球，HiFi4G[50]将 3DGS 与其对偶图机制提供的非刚性跟踪相结合，用于高保真渲染，从而以更紧凑的方式成功地保持了时空一致性。为了在消费者级设备上实现更高的渲染速度和高分辨率，GPS Gaussian[51]在稀疏源视图上引入高斯参数图，以与深度估计模块联合回归高斯参数，而无需任何微调或优化。为了更容易控制头部表情，GaussianAvatars[52]在 3DGS 中引入了几何先验（Flame 参数化人脸模型），将 Gaussians 绑定到显式网格上，并优化高斯椭球的参数。HeadGas[53]为 3DGS 赋予了由 3DMM 的表达向量加权的潜在特征的基础，从而实现了实时可动画化的头部重建。FlashAvatar[54]进一步在参数人脸模型中嵌

入均匀的 3D 高斯场，并学习额外的空间偏移来捕捉人脸细节，成功地将渲染速度提高到 300 FPS。为了合成高分辨率的结果，高斯头像采用超分辨率网络来实现高保真度的头像学习。3D-PSHR[55]将手部几何先验（MANO）与 3DGS 相结合，首次实现了手部的实时重建。GaussianHair[56]首先将 Marschner 头发模型与 UE4 的实时头发渲染相结合，以创建高斯头发散射模型，它可以捕捉复杂的头发几何体和外观，以实现快速光栅化和体积渲染，从而实现包括编辑和重新照明在内的应用。虽然计算机视觉和人工智能技术的发展大大加速了数字分身重构的应用落地进程，3DGS 技术和视频人体姿态估计技术让高保真、高效率、低成本的数字分身的动静态重构成为可能，但是面向工程化应用的挑战主要在于如何建模毛发、衣物布料等运动变形、提高重建精度、支持重光照，以及如何建立完整的基于可微渲染的制作管线。

数字分身的动态重构依赖于动画文件来驱动模型变形，进而生成动态内容。最高效的动画生成方式是运用运动捕捉技术，将采集到的动作迁移至数字分身。传统的动作捕捉技术，如光学式和惯性式，虽已成熟并在市场广泛应用多年，生成动画的精度高，但它们对环境要求严苛且造价高昂。随着计算机视觉技术的进步，基于视频的人体姿态估计[57]为高效且低成本的动画生成提供了新的技术路径。Hossain 和 Little[58]提出了一种带有快捷连接的长短期记忆（LSTM）单元的循环神经网络，以利用人类姿势序列中的时间信息，预测帧间连续的 3D 姿态。SMPL[59]人体参数化模型在人体姿态估计中应用广泛。由于该模型预定义了人体 3D 形状及参数空间，神经网络只需确定 SMPL 模型中的形状参数和关节回归矩阵，即可重建人体 3D 姿态。Kocabas 等[60]等人利用大规模动作捕捉数据集 AMASS，对基于 SMPL 的方法进行对抗训练，该方法被命名为 VIBE（视频推断的身体姿态和形状估计）。VIBE 通过 AMASS 训练区分真实人体运动与姿势回归模块预测的姿势的网络，有效提高了所推断的人体姿态的真实度。此外，局部遮挡也是单视图环境下 3D 姿态估计面临的一个难题。解决这一问题的有效方法是从多个视图进行三维人体姿势估计，因为在一个视图中被遮挡的部分可能在其他视图中可见。然而，从多个视角重建三维姿态需要解决不同摄像机之间的位置关联问题。Rhodin 等[61]在网络中应用了多视图一致性约束，但这种方法需要大量的 3D 真实数据进行训练。为了突破这一限制，他们进一步提出了一个 encoder-decoder 框架[62]，以学习多视图图像的几何感知三维潜在表示。Pavlakos 等[63]人则基于标定的相机参数，将多视角图像的二维关键点热图整合成一个三维图形结构模型。Wang 等[64]人开发了一种基于 transformer 的模型，该模型能够直接从多视图图像中回归出 3D 姿态，而无需依赖任何中间任务。他

们提出了多视图位姿转换器（Multi-view Pose Transformer, MvP）用于表示多人关节的 query embedding，并采用了融合多视角信息的几何引导注意机制。

与传统的动作捕捉技术不同，生成式运动捕捉[65]不需要依赖传感器、摄像机或演员的实际动作记录，而是通过从数据学习到的模型来生成动作，提供了一种新的数字分身动画生成技术。所使用的运动数据集通常包含行走、跑步、跳跃等各种动作，以用于训练模型提取动作中的特征和模式。模型训练完成后，生成式运动捕捉就能够通过输入控制信号（例如 text to motion 文本到动作）或关键帧来生成连续的动作序列，并将其应用于虚拟角色或动画中。此外，生成式运动捕捉还能够拓展动作的创造性，通过修改和控制模型，创造出一些在现实中难以实现的动作效果。

来自复旦大学、腾讯和上海科技大学等机构的团队开发了 MotionGPT 模型[66]。研究团队发现人体运动与语言存在一定的相似性，因此将人体运动视为一种特殊的语言形式。他们将语言数据与大规模运动模型结合起来，实现了对运动相关任务性能的提升。该模型使用离散向量量化来处理人体运动，并将 3D 运动转换为运动标记，类似于生成单词标记的过程。MotionGPT 可以通过文本生成逼真的运动序列、为运动添加字幕、预测运动和生成中间动作等。它不仅能提供高质量的运动生成，还能处理多种与运动相关的任务，为运动捕捉领域带来了重要的进展。南洋理工大学团队开发了 MotionDiffuse 模型。这是一种新的框架，可以通过文本生成多样且逼真的人体动作。它采用概率映射而非确定性的方式生成运动，可以通过细节指令和文本提示进行灵活控制。针对特定的动作需求，网易互娱 AI LAB 与清华大学合作推出了舞蹈动画合成系统 ChoreoMaster。该系统可以根据音乐风格生成不同类型的舞蹈动画，如爵士舞、二次元舞和街舞等。

无论是基于视频的人体 3D 姿态估计还是生成式动画技术，普遍存在动作精度、鲁棒性不足的问题，也是有待进一步研究攻克的重点问题。

2.4.3 三维内容生成

3D 内容生成技术泛指从文本、图像等多模态输入中自动生成三维内容，是元宇宙构建的热点研究方向。生成对抗网络（GAN）[67]和扩散模型（diffusion models）[68]作为最常用的生成式人工智能模型，在 3D 内容生成上展现出了显著的潜力。

生成对抗网络（GAN）[67]由 Ian Goodfellow 等人在 2014 年提出，它由两个主要部分组成：生成器（Generator）和判别器（Discriminator）。生成器负责生成逼真的数据，

而判别器则尝试区分生成的数据和真实数据。两者之间的博弈过程使得生成器不断提高其生成数据的质量，其结构如图 2-18 所示。GAN 在图像生成、视频合成和 3D 模型创建等方面取得了显著进展。例如，StyleGAN[69]系列模型通过引入风格控制机制，实现了高分辨率和高质量的图像生成。最新的三维人脸生成的工作，如 EG3D[70]、Portrait3D[71]等，很多都采用了 StyleGAN2 作为基本的网络架构。

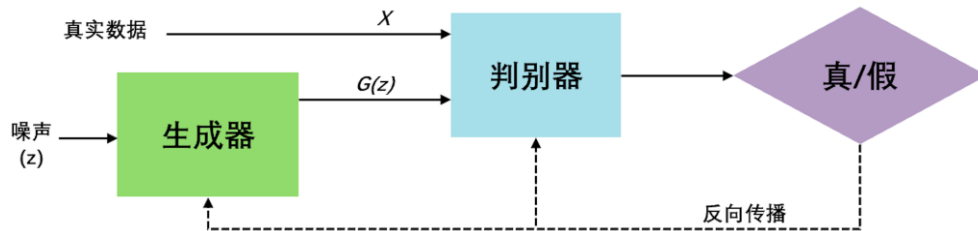


图 2-18 生成对抗模型结构示意图

扩散模型[68]是一类基于去噪扩散概率模型（DDPM）的生成技术。其核心思想是通过迭代去噪过程，从随机噪声生成目标数据。扩散模型的训练过程包括两个步骤：前向噪声注入和反向去噪预测。前向噪声注入是指在数据中逐步引入高斯噪声，而反向去噪预测则是通过学习去噪函数来恢复原始数据，如图 2-19 所示。

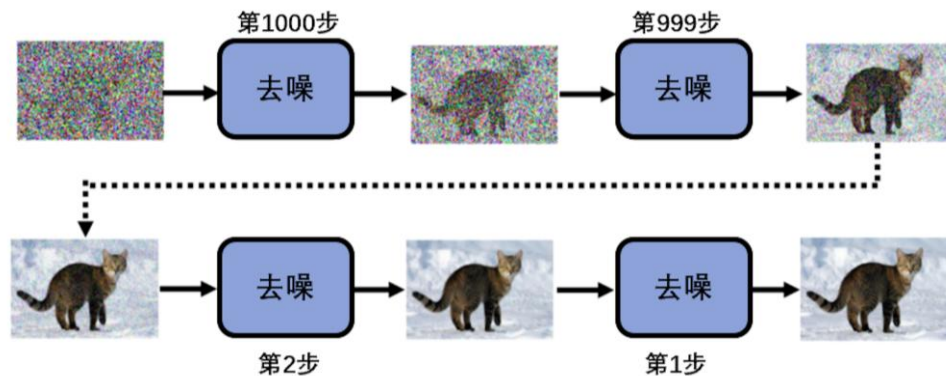


图 2-19 扩散模型的前向加噪与反向去噪过程

近年来，基于 CLIP 和扩散模型的方法大幅提升了文本到 3D 生成的能力。例如，Point-E[72]是一种图像到点云扩散模型，利用 CLIP 视觉潜在编码作为输入，生成 3D 点云。Shap-E[73]通过引入隐空间投影，实现了对 SDF 表示的重建，支持大规模词汇 3D 生成。Gaussian-Dreamer[74]结合 3D 和 2D 扩散模型，通过高斯分裂确保 3D 一致性和细节生成，从文本描述生成 3D 实例。

通过图像生成 3D 内容的方法同样取得了显著进展。基于图像的 3D 生成利用扩散模型从单视点图像生成多视点图像，利用多视点的一致性，生成一致的 3D 模型。例如，

NPGs (Neural Parametric Gaussians) [75]通过局部表示和时间共享锚定的 3D 高斯体实现高质量的非刚性对象重建。Kratimenos 等人提出的 DynMF[76]方法,通过绑定运动系数实现动态场景的优化,有效分离场景运动,生成高质量的动态场景。多模态 3D 生成是当前研究的热点之一,探索如何整合来自不同模态的数据(如文本、图像和声音)来创建更加丰富和互动的 3D 内容。例如,Clip-Nerf[77]结合 CLIP 模型,实现了高质量的 3D 场景生成。CG3D[78]框架利用显式辐射场,实现了多物体场景的生成和快速编辑。GaussianDiffusion[79]框架结合高斯斑点和朗之万动力学扩散模型,加速渲染过程。这些技术的发展不仅能够提升内容的生成效率和质量,还有助于创造更加个性化和互动的用户体验。

3D 内容生成也可面向可自由视点交互的数字分身空间视频生成问题,基于 3DGS 技术进一步对 4D 生成进行了初步探索。AYG 为 3DGS 赋予了动力学,并为文本到 4D 的生成提供了变形网络。它分为两个阶段,基于稳定扩散和 MVDream 的具有 SDS 损失的静态 3DGS 生成,以及基于文本到视频扩散模型的具有视频 SDS 损失的动态生成。在动态生成阶段,只对变形场网络进行优化,并随机选择一些帧来添加基于图像的 SDS,以确保生成质量。DreamGaussian4D 在给定参考图像的情况下实现 4D 生成。首先使用 DreamGaussian 的改进版本生成静态 3DGS。现成的稳定扩散视频用于从给定的图像生成视频。然后,还通过优化添加到静态 3DGS 的变形网络来实现动态生成,并且生成的视频被用作监督,以及基于来自采样视图的 Zero-1-to-3 XL 的 3D SDS 损失。最后,该方法还提取了一个网格序列,并利用图像到视频的扩散模型对纹理进行了优化。最后,对于视频到 4D 的生成,4DGen 和 Efficient4D 都提出利用 SyncDreamer 从输入帧生成多视图图像作为伪地面实况来训练动态 3DGS。前者引入 HexPlane 作为动态表示,并使用生成的多视图图像作为三维变形伪地面实况来构建点云。后者直接将 3D 高斯转换为 4D 高斯,并通过融合相邻时间戳的空间体积来增强 SyncDreamer 的时间连续性,实现时间同步,以生成更好的跨时间多视图图像进行监督。

大规模 3D 模型数据集对三维内容生成的推动效果十分明显。从在 Stable Diffusion 发布的 Objaverse 三维模型数据集上训练的 Tripo3D[80]和 Rodin Gen_1[81]等具体方法可以看出,3D 内容生成技术将在大规模数据集和高效生成算法的支持下,已经可以支持文本、图像等条件下的三维内容生成。Trellis 通过创新三维表达以及美学标准驱动的三维数据训练集的构建大幅提升了生成的三维模型的质量和细节,Spac3D 通过提升体表达分辨率,进一步提升了所生成三维模型的质量。随着研究的不断深入,这些技术将

在元宇宙中的应用前景将更加广阔，为虚拟世界的构建带来更多可能性。

2.4.4 世界模型

在人工智能从“感知智能”向“认知智能”跨越的进程中，世界模型（World Model）成为关键突破口。这种被称为 AI “内在心智”的技术，通过对现实世界的抽象建模与动态仿真，让机器具备了类似人类的预测、推理与决策能力，正在重塑智能系统与物理世界的交互方式。

世界模型的核心定义，是通过整合文本、图像、传感器等多模态数据，构建对环境结构化表征，进而模拟现象变化、预测未来状态的生成式 AI 模型。简单来说，它并非直接“复制”世界，而是在 AI 系统中建立一个可运行的“虚拟沙盘”，实现对世界规律的隐式理解。这一概念源于认知科学中的“心理模型”理论，2018 年 Jurgen 在 NeurIPS 上发表的论文[82]首次将其系统化，提出通过循环神经网络构建具备预测能力的智能体框架，奠定了技术基础。

从发展历程看，世界模型已从理论构想走向技术落地。2019 年 DeepMind 的 MuZero 算法突破传统限制，无需预先知晓规则即可通过自我模拟掌握围棋等复杂游戏；2022 年 LeCun 提出的 JEPA[83]模型强化了因果表征学习；2024 年 OpenAI 的 Sora[84]和城市仿真模型 UrbanWorld[85]，以及 Google 的 Genie[86]系列则将其推向高潮，实现了对物理世界的高精度动态复刻。这些进展背后，是深度学习技术与计算资源的双重突破，让高维数据的建模与预测成为可能。

世界模型的技术架构由三大核心模块构成，形成“感知-预测-决策”的闭环。首先是状态表征模型，通过变分自动编码器（VAE）等技术将图像、声音等原始数据压缩为低维潜在状态，过滤冗余信息并保留核心特征，这是建模的基础。其次是动态模型，作为核心组件，它通过循环神经网络（RNN）或状态空间模型（SSM）学习状态转移规律，比如预测物体在重力作用下的运动轨迹，为 AI 提供“推演能力”。最后是决策模型，利用强化学习或模型预测控制（MPC）等方法，基于预测结果规划最优动作序列，完成从模拟到执行的转化。典型的世界模型框架图如下图所示：

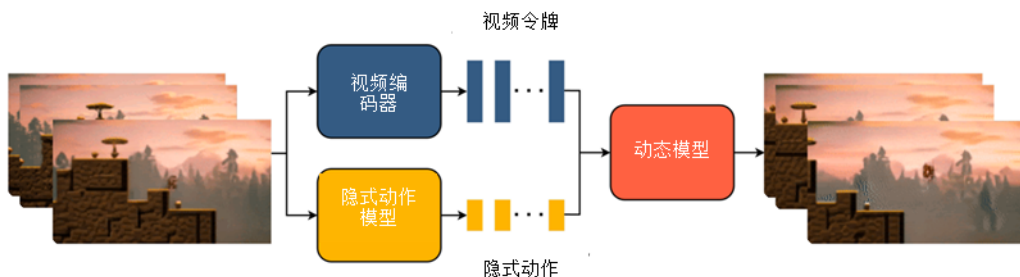


图 2-20 生成对抗模型结构示意图

与传统 AI 相比，世界模型的独特优势体现在三方面。一是强大的预测能力，能基于当前状态推演未来多种可能性，如自动驾驶系统预测行人的行走路径。二是物理认知与因果推理，理解“因为推了杯子，所以杯子会倒”这类底层逻辑，而非单纯依赖数据关联。三是反事实推理能力，可回答“如果改变某个条件，结果会怎样”的假设性问题，为复杂决策提供支撑。这些能力让 AI 从“被动响应”升级为“主动规划”。

在实际应用中，世界模型已渗透多个领域[87]-[96]。自动驾驶领域，特斯拉等企业利用它整合多摄像头数据，构建 3D 场景向量空间，实时预测车辆与行人的运动轨迹，毫秒级响应突发状况。机器人领域，工业机器人通过建模零件物理特性与操作规律，精准完成装配任务，降低试错成本。游戏与 VR 领域，它让 NPC 能根据玩家行为动态调整策略，提升交互真实性。在更宏观的场景中，气候模型、疫情传播 SEIR 模型也可看作世界模型的延伸，助力复杂系统预测与调控。

尽管发展迅速，世界模型仍面临诸多挑战。维度灾难导致高维场景（如城市交通）建模需巨大计算资源；开放世界中的未知事件（如突发交通事故）难以精准预测；深度学习的“黑箱特性”使其决策过程缺乏可解释性，带来安全隐患。此外，多模态数据的有效对齐、伦理偏见的规避等问题，仍需技术突破。

未来，随着多模态融合、神经符号系统等技术的发展，世界模型将更贴近人类认知模式。当 AI 能在“脑海”中完成充分模拟与推演后，其在复杂环境中的适应能力将实现质的飞跃。从工业生产到日常生活，这种让机器“读懂世界、预测未来”的技术，正推动人工智能进入更具想象力的新阶段，成为连接数字世界与物理世界的核心桥梁。

2.5 元宇宙协同

元宇宙常常被视为未来互联网的演化，将包括多样化的虚拟商品和服务、完全虚拟的工作和生活空间以及相应的社会和经济系统。元宇宙协同的概念涉及元宇宙环境中不

同用户、系统和平台之间的合作与互动，具体是指如何利用各种技术、工具以及跨平台的协议来促进在虚拟空间内的协作、共享和通信。元宇宙协同能够极大地提升元宇宙的功能性和用户体验，使得元宇宙不仅是一个娱乐和交流的平台，也成为可以进行教育、商务和创新的多功能空间。

从技术角度，实现元宇宙协同的核心要素是保证跨平台互操作性，一方面是确保不同的技术平台可以无缝连接，使得用户可以在不同的虚拟环境之间流畅地移动和交互；另一方面是实现不同系统或平台间的数据共享，使得一个平台上虚拟事物的变化可以在另一个平台上体现出来，由此实现了穿透现实和不同虚拟世界的经济体系。此外，从社会的角度，元宇宙协同还包含在元宇宙中形成有组织的社区以及相应的社区自治，以及实现不同元宇宙之间的文化交流。元宇宙协同极大地增强了元宇宙的功能性和吸引力，让元宇宙不仅成为一个生动真实的虚拟环境，还支持了其作为未来数字生活和工作空间的核心地位，为持续的技术创新和社会变革创造了重要条件。

在元宇宙中，用户不断地分享和交互个人信息、参与经济活动并构建社交网络，不同的元宇宙之间也在共享数据和资源。若安全措施不足，用户将面临数据泄露、金融诈骗和身份盗窃等风险。有效的协同安全机制能够确保跨平台互操作性不会成为潜在的安全漏洞，保障用户操作的可靠性，从而促进元宇宙的健康发展和用户信任，这对于元宇宙长期繁荣至关重要。

随着元宇宙的迅速发展，协同的需求愈发迫切，基于区块链技术的去中心化的数据服务为此提供了重要的技术基础，这种服务不仅被视为互联网的下一时期的形态——Web3.0[97]，也成为当前元宇宙概念的核心要素[98]。

2.5.1 区块链技术原理

（1）基本原理

区块链技术是一种分布式数据库或账本技术，通过网络中多个节点以去中心化的方式维护和验证数据。这种去中心化的特性为元宇宙构建了一个无需许可、开放透明的协作基础。

在元宇宙的场景下，区块链的核心技术原理[99]主要体现在以下几个关键方面：

- 不可篡改的加密结构

区块链的基础是“区块”和“链”。如图 2-21 所示，每个区块包含前一个区块的哈希值，形成链式结构。这种密码学设计确保了数据一旦上链就无法被非法修改。对于元宇宙而言，这意味着用户的虚拟资产（如土地、装备、数字艺术品）拥有了永久、安全的产权证明，不会因单一节点的关闭或数据故障而消失。

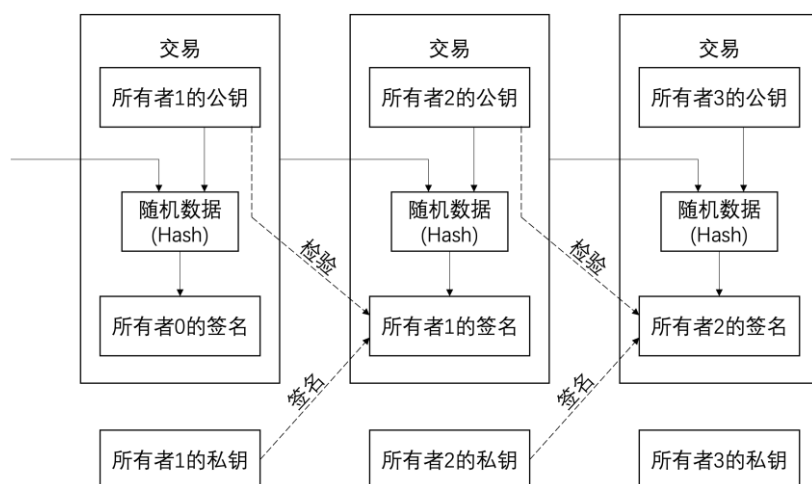


图 2-21 “区块”与“链”的结构图¹

• 分布式存储与共识机制

区块链将数据存储在全网节点上，并通过共识机制（如 PoS、PoW 等）确保所有节点账本状态的一致性。这种机制保证了元宇宙中虚拟资产的所有权记录、交易历史和身份信息是真实可信且全网同步的，解决了虚拟世界中的信任难题。目前以太坊已完成从 PoW 到 PoS 的转型，并在 Dencun 升级中进一步优化网络性能[100]。Solana 采用了历史证明（PoH）与 PoS 相结合的创新共识模型，通过加密时间戳序列实现高吞吐量，这些都有力支持了元宇宙的实时交互场景。

• 智能合约[101][102]

区块链的应用已经超越了简单的数据库账本功能。以太坊[103]引入的图灵完备的链上虚拟机，使得在区块链上部署和执行复杂的业务逻辑成为可能。智能合约是运行在区块链上的自动执行程序，充当了元宇宙的“底层操作系统”和“商业逻辑层”。它允许在没有中介的情况下自动执行复杂的交互逻辑（如自动化交易、版税分配、门禁控制等），是构建元宇宙去中心化经济体系和治理规则的关键。

（2）优势和不足

¹ Nakamoto S, Bitcoin A. A peer-to-peer electronic cash system[J]. Bitcoin.—URL: <https://bitcoin.org/bitcoin.pdf>, 2008

将区块链技术应用元宇宙后，能为元宇宙的协同带来大量显著的优势：

- **提供去中心化的信任框架**

区块链的去中心化特性意味着没有单一的控制点或故障点，这提高了元宇宙系统的整体可靠性和抗攻击性。在元宇宙协同中，这种去中心化的信任机制能够确保各种跨平台交易和互动无需通过第三方中介即可直接进行，从而降低协作成本、提高效率[104]。

- **促进跨平台资产和数据的流动性**

区块链可以使元宇宙中的资产（如虚拟土地、物品或货币）和数据在不同的平台间流动。通过智能合约和标准化的接口，用户可以将他们的资产从一个虚拟环境无缝迁移到另一个，实现更广泛的互操作性和用户参与。

- **加强用户协同和社区治理**

去中心化自治组织（DAOs）区块链技术赋予元宇宙独特的组织形式，这种结构使得元宇宙的用户能够从单纯的消费者转变为共同的所有者和建设者。通过 DAOs，用户可以投票决定元宇宙的更新、规则变更和资金分配等事务，这种民主化的治理方式增强了社区的凝聚力和透明度。

- **增加交易和互动的安全性**

区块链的加密技术确保所有交易都是安全的，不可篡改的记录和智能合约的自动执行减少了人为错误和欺诈的可能。在元宇宙中，这意味着用户在购买、销售或交换虚拟资产时可以得到代码层面的安全保障。

- **创建持久的、可验证的历史记录**

每个交易和互动都会被永久记录在区块链上，为元宇宙中的每一次活动提供了时间戳和不可更改的证据。这种透明度和可审计性对于构建这种透明度和可审计性对于构建元宇宙的信用体系、追溯虚拟商品来源（Provenance）以及确立数字版权至关重要。

- **实现经济模型的创新**

区块链的智能合约允许元宇宙内创建复杂的经济模型，包括创造和交易虚拟商品（Token 化）、实施自动化的数字版税系统（Creator Economy）以及通过代币经济激励用户参与。这种新型经济模型为元宇宙的内容创造者提供了直接的变现渠道，为元宇宙的可持续发展和扩展提供了源源不断的动力。

然而，区块链技术在元宇宙的大规模应用中仍面临一些挑战[105]：

- **可扩展性问题[106]**

元宇宙的高并发交互对底层基础设施的吞吐量提出了极高要求。许多公链在处理海

量交易时面临速度和容量限制，容易导致网络拥堵和高昂的费用。为适应元宇宙的需求，业界正在积极推进 Layer 2 扩展方案（如 Optimism、Arbitrum、zkSync 等）[107][108]和高性能公链（如 Solana 的 Firedancer 升级）[109]，以及模块化区块链架构，旨在为元宇宙提供低成本、高响应的底层网络。

- **能源消耗**

传统（PoW）机制的区块链需要庞大的计算资源来维护网络，导致巨大的能源消耗和环境问题[125]。引发了大量社会关注。但随着行业向 PoS 等低能耗共识机制转型（如以太坊能源消耗降低 99.95%[111]），能耗和环境问题已大幅改善，这对元宇宙的可持续发展至关重要，有效降低了大规模虚拟活动的环境成本。

- **用户友好度**

区块链技术的复杂性（如私钥管理、Gas 费等）构成了普通用户进入元宇宙的门槛。为提升元宇宙的用户体验，账户抽象（Account Abstraction，如 ERC-4337）技术正在推动智能合约钱包的普及，允许通过社交登录、无脑记词恢复等方式降低使用难度[112]，使元宇宙的 Web3 底层对用户更加友好。

- **数据隐私**

公链的透明性虽然带来了信任，但也暴露了元宇宙用户的链上行为轨迹。为保护元宇宙中的用户隐私，零知识证明（Zero-Knowledge Proof，如 zkSNARKs）等隐私增强技术正得到广泛应用[113]。这些技术允许用户在不泄露具体信息的前提下证明身份或资产，为构建既开放又保护隐私的元宇宙环境提供了解决方案。

2.5.2 区块链关键技术

前文中我们探讨了区块链技术的基本原理以及其优缺点，为了克服这些挑战并扩展区块链的功能，开发者们引入了一系列高级技术解决方案。接下来，我们将逐一介绍其中几种关键技术，这些技术增强了区块链的功能和性能，大大拓展了其在元宇宙中的应用场景。

（1）智能合约（Smart Contract）

早在 20 世纪 90 年代，就有学者提出智能合约的概念[114]，以太坊虚拟机的出现使其成为可能。智能合约本质上是一组运行在区块链虚拟机上的程序代码，这些代码具备自动执行预设任务或合约条款的能力，通常在满足特定条件时触发。它们允许在没有第

三方中介的情况下进行可信交易，这些交易是透明的、不可逆的，并且可以追踪。智能合约是区块链技术的核心特征之一，也被用于许多其他区块链系统中。

智能合约具备以下特点：

- **自动化**：智能合约的一大特点是一旦预设的条件被满足，相关的合约条款将自动执行。这种自动化减少了人工干预的需要，提高了处理速度和效率。
- **不可更改性**：一旦智能合约被部署到区块链上，它的代码就不能被修改。这保证了程序的稳定性和可靠性。
- **分布式**：智能合约在区块链网络上的多个节点中部署和执行，增加了其抗攻击能力和故障容错性。
- **透明性**：合约的条款对所有相关方都是可见的，这增加了操作的透明度。

一组具有特定功能的智能合约组成了 DApp（去中心化应用）。这些合约共同定义了 DApp 的业务逻辑和功能，确保了应用的去中心化和自动化执行。智能合约的组合方式及其与用户界面的集成决定了 DApp 的效能和用户体验。

随着技术发展，智能合约的安全性和可用性得到持续改进。形式化验证（Formal Verification）技术通过数学方法证明合约代码的正确性，减少漏洞风险[115]。可升级合约模式（如代理模式 Proxy Pattern）允许在保持合约地址不变的情况下更新业务逻辑[116]。此外，账户抽象（Account Abstraction）技术如 ERC-4337 使智能合约钱包成为“一等公民”，实现了 gas 费代付、批量交易等高级功能，大幅提升了用户体验[112]。

（2）二层网络（Layer 2 Solutions） [117]

区块链的性能和可扩展性问题一直是困扰其发展的重大问题，也是其应用于元宇宙的最重要的阻碍之一。二层网络（Layer 2）技术是解决这个问题的重要手段之一，它通过在主链之上构建额外的网络层来处理交易，显著提高了处理速度和降低成本，使得大规模应用成为可能。

	状态通道	侧链[0]	Plasma	Optimistic Rollups	Validium	zkRollup
安全性						
在线假设（例如，瞭望塔）	是	使用保证金机制来绕过	是	使用保证金机制来绕过	否	否
大规模退出假设	否	否	是	否	否	否
达到法定人数的验证者可冻结资金	否	是	否	否	是	否
达到法定人数的验证者可没收资金	否	是	否	否	是	否
对热钱包密钥攻击的脆弱性	高	高	中等	中等	高[1]	免疫
对加密经济攻击的脆弱性	中等	高	中等	中等	中等	免疫
加密原语	标准	标准	标准	标准	新	新
性能/经济性						
以太坊1.0上的最大吞吐量	1..∞ TPS[2]	10k+ TPS	1k..9k TPS[2]	2k TPS[3]	20k+ TPS	2k TPS
以太坊2.0上的最大吞吐量	1..∞ TPS[2]	10k+ TPS	1k..9k TPS[2]	20k+ TPS	20k+ TPS	20k+ TPS
资本效率	否	是	是	是	是	是
开设新账户需要额外的链上交易	是	否	否	否	否	否[5]
交易费用	非常低	低	非常低	低	低	低
可用性						
提现时间	1确认	1确认	1周[4][7]	1周[4][7]	1..10分钟[7]	1..10分钟[7]
达到主观确定性的时间	即时	不可用	1确认	1确认	1..10分钟	1..10分钟
主观确定性的客户端可验证性	是	不可用	否	否	是	是
即时交易确认	完全	担保	担保	担保	担保	担保
其他角度						
智能合约	有限	灵活	有限	灵活	灵活	灵活
EVM字节码可移植性	否	是	否	是	是	是
原生隐私选项	有限	否	否	否	是	是
注： [0] 某些研究者完全不认为侧链应被归入 L2 范畴，可见： https://twitter.com/gakonst/status/1146793685545304064 [1] 要看相关升级机制的实现，不过一般来说都可以 [2] 有非常复杂的限制 [3] 为保证与 EVM 的兼容性，吞吐量的上限是 300 TPS [4] 这个参数实际上是可调的，但大部分研究者都觉得 1 到 2 周时间比较安全 [5] 要看相关的实现。zkSync 是不需要的，但 Loopring 需要 [7] 理论上来说，可以通过流动性提供商缓解这个问题，但是会让整个方案的资金利用效率性变差						

图 2-22 不同二层网络技术特性对比²

主要类型的二层网络技术方案包括：

- **状态通道（State Channels）** [118]：状态通道允许用户在区块链之外进行多个交易，并仅在开始和结束时与主链交互。这些交易在私下完成，最终状态会被提交到主链。状态通道的一个典型例子是比特币的闪电网络。
- **侧链（Sidechains）** [119]：侧链是独立的区块链，与主链并行运行，并通过双向锚定（两边的资产可以互相转移）与主链交互。侧链可以有自己的共识机制和区块生成算法，允许进行实验或特定用途的优化。
- **Plasma** [120]：Plasma 是一种特殊的框架，创建了一个由主链托管的子链树。每个子链可以处理大量交易，并定期将压缩后的交易数据提交回主链。这种结构旨在大幅提高交易吞吐量。
- **Rollups**：Rollups 是当前最受关注的 Layer 2 解决方案之一，它们将多个交易组

² <https://blog.matter-labs.io/evaluating-ethereum-l2-scaling-solutions-a-comparison-framework-b6b2f410f955>

合在一起，在链外处理交易数据，然后将交易数据的一个压缩版本（roll up）提交到主链。根据数据处理方式的不同，Rollups 可以分为 Optimistic Rollups[121] 和 ZK Rollups[122][123]。

- **Validium**[124]: Validium 是一种扩容解决方案，使用零知识卷叠等有效性证明来执行交易的完整性，但它不在主网上存储交易数据。

上述几种二层网络技术的特性对比如图 2-22 所示，这些二层网络技术的出现保证了不同的元宇宙可以根据其自身的需求建立其独有的二级网络，从而使其既能获得经特制化处理的高效率，又可以享受底层区块链所带来的安全性。

以太坊 Dencun 升级后，Layer 2 生态获得显著发展。基于 OP Stack 的 Rollup 网络（如 Optimism、Base、Blast 等）和基于 zkEVM 的 Rollup（如 zkSync、Scroll、Linea 等）在 2024 年实现了交易费用降低 90%以上的突破[108]。模块化区块链架构进一步推动了这一趋势：数据可用性层（如 Celestia、EigenDA）将数据存储与执行分离，使 Layer 2 可以选择更经济的数据发布方案[125]；共享排序器（Shared Sequencer）和基于再质押的安全共享（如 EigenLayer 提供的 AVS 服务）降低了独立 Layer 2 的启动成本和安全风险[126]。这些创新为元宇宙应用提供了更灵活、更经济的基础设施选择。

（3）跨链桥（Cross-Chain Bridges）

跨链桥（Cross-Chain Bridge）[119]是一种技术，允许在不同的区块链网络之间转移资产和/或数据。这种技术使得原本在单一区块链上独立运行的资产和信息可以在多个不同的区块链平台之间进行互操作和交换。跨链桥解决了区块链生态系统中的孤岛问题，提高了各个区块链平台的连通性和功能性[127][128]。如图 2-23 所示，目前已有数十种不同的跨链桥用来实现各类区块链资产间的联通。

跨链桥的工作原理通常涉及以下几个步骤：

- **锁定源链资产**：用户首先在原始区块链（源链）上锁定希望转移的资产。这通常通过发送资产到一个特定的智能合约地址来完成，该地址由跨链桥的运营者控制。
- **发行代表资产的代币**：在目标区块链（目标链）上，跨链桥会发行等量的代币，这些代币代表被锁定的原始资产。例如，将比特币转移到以太坊上时，可能会发行等额的 Wrapped Bitcoin（wBTC）。
- **赎回和解锁**：当用户希望将资产转回原始区块链时，他们需要在目标链上将代

币发送回跨链桥的合约,合约随后会销毁这些代币,并在源链上解锁原始资产,返回给用户。

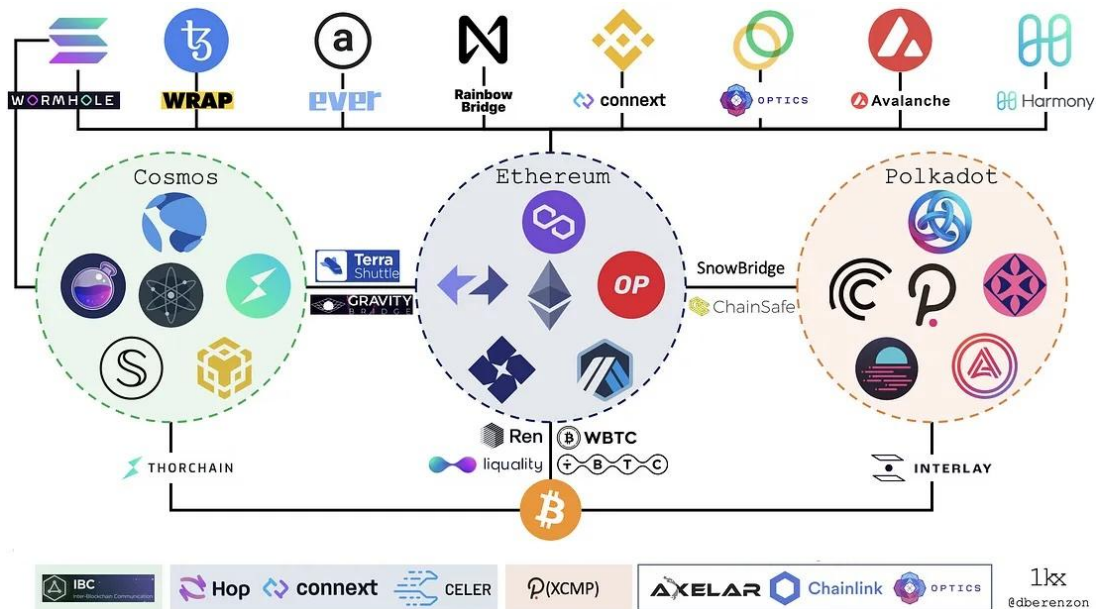


图 2-23 目前已经有几十种不同的跨链桥³

从跨链桥的可信任程度上,我们一般对其分成两大类[127]:

- **需信任 (Trusted) 链桥。**它们使用一组外部验证者(具有多重签名的联盟、多方计算系统、预言机网络)跨链发送数据。因此,它们可以提供出色的连通性,并完全支持跨链通用信息传递。在速度和成本效益方面它们通常也表现良好。
- **去信任 (Trust-less) 链桥。**这类链桥依靠它们连接的区块链及其验证者来传输信息和代币。它们是“去信任”的,因为它们没有增加新的信任假设(区块链除外)。

跨链桥在元宇宙协同中扮演了重要的角色,因为它们允许不同区块链网络之间的资产和数据互操作[129]。这是实现元宇宙不同平台和应用之间无缝整合的基础,有助于打造一个连贯、统一的虚拟体验,实现不同元宇宙中资产的互通和流动,也便于数据及服务共享和合作。

然而,跨链桥也面临显著的安全挑战。2022 年以来,Ronin、Nomad、Wormhole、Multichain 等跨链桥遭受重大安全攻击,累计损失超过 20 亿美元[130]。为提升安全性,业界正在探索多种新型解决方案:轻客户端桥(Light Client Bridge)通过在目标链上验

³ <https://medium.com/1kxnetwork/blockchain-bridges-5db6afac44f8>

证源链的区块头实现去信任跨链[131]；共享验证/再质押安全模型（如 EigenLayer）允许多个跨链桥共享以太坊验证者的安全性[125]；此外，跨链桥也开始区分资产桥（Asset Bridge）和消息桥（Message Bridge），后者专注于跨链消息传递，为更复杂的跨链应用提供基础[132]。

（4）预言机（Oracles）

预言机（Oracle）是区块链技术中的一个重要组成部分，它的主要功能是作为链外数据源和区块链（链内）之间的桥梁。区块链是封闭的系统[133]，无法直接访问外部的数据，预言机解决了这一问题，它允许智能合约根据现实世界的数据事件自动执行操作，这些数据包括但不限于股票价格、天气信息、体育比赛结果等[134]。随着技术的发展，一些新的预言机为区块链提供了更加复杂和可靠的服务，比如引入去中心化网络的 Chainlink[135][136][137]。

预言机的工作原理通常涉及以下几个步骤：

- **数据收集：**预言机从各种外部数据源收集信息。这些数据源可能是数据库、网站、API 接口或其他实时数据提供者。
- **数据验证：**为保证数据的准确性和安全性，预言机可能会从多个源收集同一数据，通过某种形式的共识机制验证数据的正确性。
- **数据传输：**验证后的数据被传输到区块链网络，智能合约根据这些数据执行预设的逻辑。例如，根据最新的股票价格自动执行交易。

此外，2020 年自动做市商（AMM）[138]技术的出现，使得虚拟世界能够通过简单的数学公式和人的套利行为相结合，将现实世界的价格信息反映到区块链上。

预言机加强了虚拟和现实世界之间的互动，使元宇宙能够更为真实地反映并响应现实世界的变化，同时也增强了元宇宙平台的功能性和互动性。这些特性对于构建一个多元化且互动性强的元宇宙生态系统至关重要。

预言机生态在近年来呈现多元化发展。除 Chainlink 外，Pyth Network 专注于高频金融数据，通过与交易所和做市商直接合作提供亚秒级价格更新[139]；Band Protocol 采用委托权益证明（DPoS）机制，为多链环境提供定制化数据服务[140]；API3 通过 Airnode 技术使 API 提供商能够直接成为预言机节点，减少中间环节[141]。此外，去中心化物理基础设施网络（DePIN）的兴起为预言机带来了新的数据源，传感器网络、物联网设备可以为元宇宙提供实时的物理世界数据[<https://messari.io/report/depin>]。预言机与 AI 的结

合也在探索中，使智能合约能够访问 AI 模型推理结果，为元宇宙中的智能体提供决策支持。

（5）非同质化代币（NFTs）

NFT 是“非同质化代币”（Non-Fungible Token）[142][143]的缩写，代表一种独特的、不可互换的数字资产，它使用区块链技术来证明所有权和真实性。与传统的数字货币或加密货币不同，如比特币或以太坊这些是可互换的（即一个单位可以无差别地与另一个同样单位交换），NFT 则每一个都具有独一无二的属性和价值。

在元宇宙中，非同质化代币（NFTs）主要用于确立和转让虚拟资产的所有权，如虚拟艺术品、服装和地块，支持艺术家和创作者通过确保版权来保护和货币化他们的作品[144]，也让虚拟财产所有者（如游戏玩家）能够更有效的保证自己的虚拟资产。

以太坊上最早的 NFT 协议标准是 ERC-721[142]提出的协议，这也是最为广泛接受的 NFT 项目标准。在此之后，大量支持更多功能或特质的 NFT 标准被提出，比如支持批量转账的 ERC-1155[143]、可存储动态数据的 ERC-1948[145]、支持版税机制的 ERC-2981[146]、适用于 NFT 保单的 ERC-1523[147]等。配合智能合约，这些新的 NFT 标准为元宇宙提供了复杂且多样化的功能，比如复杂资产所有权、交易功能和经济激励模型等，极大丰富了元宇宙的交互性和商业潜力。

NFT 应用正在从单纯的数字收藏品向更广泛的场景拓展。动态 NFT（dNFT）可根据外部条件或链上事件自动更新元数据，为游戏装备升级、会员权益变更等场景提供技术支持[148]。现实世界资产代币化（RWA）将房地产、艺术品等传统资产以 NFT 形式上链，促进了传统金融与区块链的融合[149]。灵魂绑定代币（Soulbound Token, SBT）作为不可转让的 NFT，用于表示身份、资质、声誉等，为元宇宙中的去中心化身份系统提供基础[150]。这些创新应用展示了 NFT 技术在元宇宙生态中的多样化潜力。

（6）去中心化自治组织（DAOs）

去中心化自治组织（DAOs）[151]是一种基于区块链技术的组织形式，通过智能合约自动执行组织的决策和运营，无需传统的管理层或中心化的控制结构。在 DAO 中，所有的决策都是通过社区成员的投票来进行，确保每个成员都能直接参与到组织的治理中[152]。

在元宇宙中，DAOs 可以发挥构建和管理虚拟社区或项目的关键角色。它们使用户能够共同管理和决定元宇宙空间内的规则、资产分配及发展方向，从而促进了一种公平、

透明且去中心化的社交和经济结构。DAOs 通过提供一种民主化的治理机制，强化了用户的参与感和归属感，有助于推动元宇宙平台和项目的长期繁荣和创新。

虽然目前 DAOs 的实践仍然存在因财富/代币集中导致的问题，但是因为开放的标准和基础设施允许人们在必要的时候用脚投票甚至进行分叉，使得基于区块链的治理系统和所有权层构成了中立元宇宙基础设施的必要条件[153]。

DAOs 治理机制在不断创新和完善。模块化治理框架（如 Aragon OSx、Tally）提供了可定制的治理模块，使 DAO 可以灵活选择投票机制、提案流程和执行方式[154]。二次投票（Quadratic Voting）通过非线性投票权重分配，缓解了“一币一票”带来的 plutocracy 问题[155]。追溯性公共物品融资（Retroactive Public Goods Funding, RPGF）由 Optimism 等项目实践，根据项目实际产生的价值进行事后奖励，激励长期建设者[156]。此外，DAO 与 AI 代理（AI Agent）的结合也在探索中，AI 可以辅助提案分析、投票建议和自动化执行，提升治理效率；最新研究表明，AI 代理的投票决策已能与人类投票结果保持高度一致，并可通过解释提案内容和检索历史数据来增强治理透明度[157][158]。这些创新为元宇宙中的社区治理和价值分配提供了更多可能性。

此外，作为一个蓬勃发展的技术领域，大量区块链相关的新技术和新理念不断涌现出来，譬如历史证明（PoH）[159]、再质押（Restaking）和主动验证服务（AVS）[160]、意图驱动架构（Intent-Centric Architecture）、可验证计算（Verifiable Computation）、去中心化存储与检索等，由于篇幅限制，本文难以一一展开详述。这些技术不断拓展区块链的功能和性能的边界，为其在元宇宙中的应用打下了坚实的基础，为元宇宙间的协同、现实与元宇宙之间信息交互提供了重要的技术方案。

2.5.3 元宇宙协同安全

由于各个元宇宙在架构、管理和用户群体上的差异性，传统的中心化身份管理难以适用于去中心化的元宇宙；通信安全需要确保数据在跨元宇宙传输过程中的完整性和机密性；数据安全需要保护在采集、共享、存储和交易过程中的隐私和完整性；计算安全则必须保障协同计算过程中的数据不被泄露和滥用。此外，监管审计在保障数据安全和合规性上也扮演着重要角色。

本节将详细探讨这些方面的安全威胁及其应对技术，以期为实现安全、高效的元宇宙协同提供指导和参考。

(1) 用户身份安全

元宇宙是一个多用户的虚拟环境，用户来自不同的元宇宙平台、应用等。用户的身份安全是元宇宙协同的基础，即不同平台和应用之间用户能够轻松迁移并保持身份的安全性。然而，传统的中心化身份管理技术难以适用于去中心化的元宇宙环境。如图 2-24 所示，为了保障元宇宙协同下的身份安全，需要从以下四方面实现去中心化的身份管理。

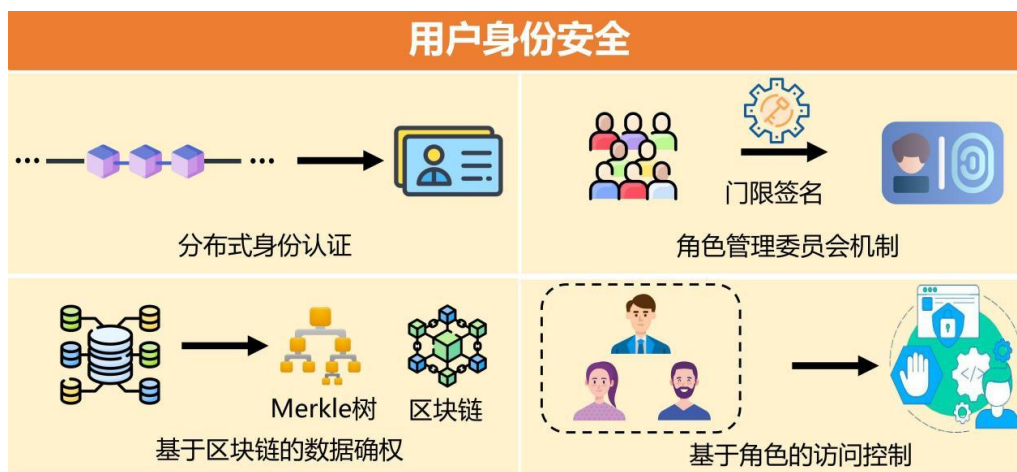


图 2-24 元宇宙协同中的用户身份安全

- **身份认证：**在元宇宙协同中，用户可能需要在不同的虚拟世界和平台之间无缝移动他们的身份。分布式身份技术（例如 W3C 的 DID 标准[161]）具有去中心化和用户自主控制的特点，使用户能够携带自己的身份认证信息跨越不同的应用场景。这可以促进整个元宇宙生态系统的开放性、互操作性和增长潜力。元宇宙安全实施需要遵循结构化顺序,首先建立基础安全架构,随后处理身份管理、隐私控制和信任构建[162]。区块链透明且不可篡改的记录保存,赋予了用户对个人数据访问的控制权，有效增强了物联网驱动的元宇宙系统的信任度[163]。
- **角色管理：**通过门限密码[164]、分布式密钥生成技术[165]等构建元宇宙用户角色管理委员会，确保任何角色管理操作都需要达到预设的管理者数量的同意，从而实现去中心化的授权与决策。此外，每个角色管理委员会成员仅掌握密钥片段，即使部分信息泄露，也能保持元宇宙系统的安全性。基于格密码的去中心化分布式密钥生成(DKG)协议通过 SIS 承诺验证、Ring-LWE 秘密共享和安全多方密钥派生,在不重构私钥的情况下维持后量子安全性[166]。
- **数据确权：**在元宇宙中，数字资产如虚拟土地、艺术品（NFT）、游戏道具等具有真实经济价值。利用区块链的分布式账本特性，可以创建不可篡改的所有

权记录,确保每个数据资产的唯一性、可追溯性和所有权透明度。区块链智能合约[167]能够自动执行合同条款,无需第三方介入,可以用来规定数字资产的使用、转让、收益分配等规则,从而实现自动化的确权和执行,减少纠纷。基于区块链的自主身份钱包作为元宇宙身份验证和资产管理的基础设施,将访问权限、凭证和声誉以可携带方式进行连接,成为数字经济标准制定的关键要素[168]。通过 NFT 实现现实世界资产(RWA)代币化,支持房地产、大宗商品等资产的分数化所有权,提供了新的投资机会和市场效率[169]。

- **访问控制:** 根据用户在元宇宙中的角色(如游客、居民、管理员等),分配相应的访问权限。基于角色的访问控制[170]减少了权限管理的复杂性,并确保用户只能访问与其角色相关的资源。云端元宇宙的访问控制模型需强调最小权限原则、职责分离、基于角色的访问控制(RBAC)、纵深防御以及可审计性和可问责性等核心原则[171]。上下文访问控制机制根据虚拟环境风险动态调整安全要求,结合去中心化身份验证支持基于区块链的凭证,以保护虚拟资产[172]。

(2) 通信安全

在元宇宙协同场景中,通信安全扮演着至关重要的角色。不同元宇宙之间的数据交换和交互操作需要通过安全的通信渠道进行,以防止数据在传输过程中被窃取、篡改或泄露。在元宇宙协同场景下,确保不同元宇宙能够安全通信的关键技术可分为以下三个方面。

- **安全通讯协议:** 安全通讯协议是构建数据传输防护网的关键。如图 2-25 所示,安全通讯协议确保了跨元宇宙数据交换的完整性、认证性和机密性。例如,使用 TLS/SSL (Transport Layer Security/Secure Sockets Layer) [173]协议可以为通信双方提供端到端的加密,防止中间人攻击和数据被窃听。更进一步,采用量子安全加密技术,如量子密钥分发(Quantum Key Distribution, QKD) [174],可以在未来抵御潜在的量子计算机攻击,为元宇宙间的超安全通信奠定基础。基于格密码的后量子属性加密(ABAC)方案与多项式函数技术相结合,为元宇宙数据管理提供了灵活且安全的数据共享框架,同时具备抵御量子攻击的能力[175]。

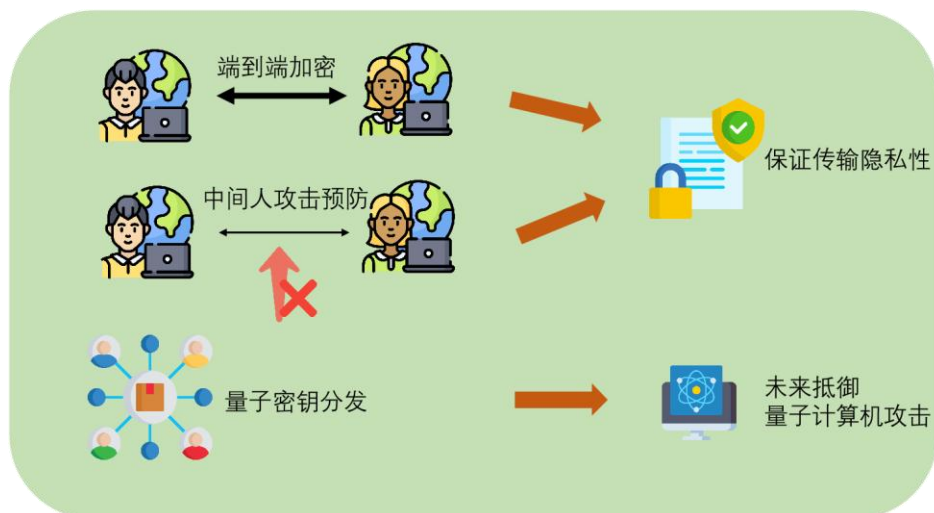


图 2-25 安全通讯协议

- 匿名通信：**匿名性是在元宇宙交互中保护用户隐私的重要方面[176][177]。利用诸如 Tor 网络（如图 2-26 所示）或零知识证明（Zero-Knowledge Proof）[178]（如图 2-27 所示）的匿名通信技术，用户能够在不透露自己真实身份或敏感信息的前提下参与跨元宇宙活动。零知识证明技术尤其强大，它允许一方证明其拥有某些信息、是某种状态或能执行某项计算，而无需揭露除断言真实性以外的任何信息。这在验证用户权限或资产所有权时尤为关键，同时保持用户隐私完整无损。元宇宙无线通信系统中零知识证明等技术的集成,通过理论洞察与实证证据的结合,增强了隐私保护、安全性和操作效率[179]。



图 2-26 匿名通信技术中的 Tor 网络

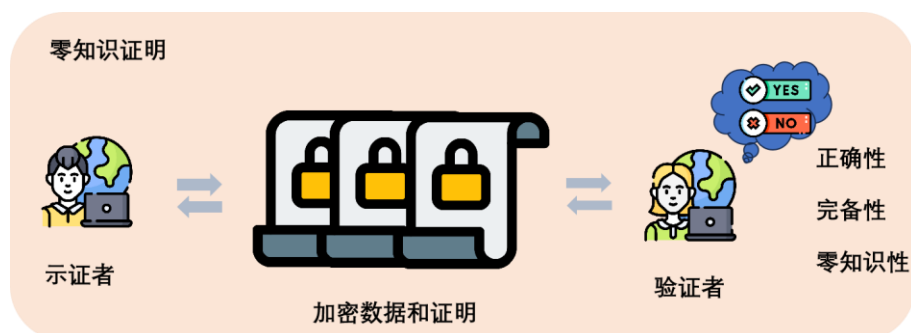


图 2-27 匿名通信技术中的零知识证明

- 流量安全：** 为了维护元宇宙间数据流动的安全性，实施高效的流量安全机制至关重要。如图 2-28 所示，流量安全机制包括但不限于入侵检测系统（Intrusion Detection Systems, IDS）和入侵防御系统（Intrusion Prevention Systems, IPS）[180][181][182]，它们能够监控网络流量，识别并阻止恶意行为。此外，采用分布式账本技术（如区块链）可以为交易和信息交换提供不可篡改的记录，确保数据源头的可追溯性和交易的透明度，同时利用智能合约自动执行规则，增加安全性。



图 2-28 流量安全机制

(3) 数据安全

在元宇宙协同场景中，根据不同元宇宙的安全需求，协同过程中需确保隐私数据不泄露，同时保证数据的真实性与完整性，从而维护元宇宙交互的安全与公平性。如图 2-29 所示，本文总结了元宇宙协同中数据采集、数据共享、数据存储和数据交易四类场景的安全技术。

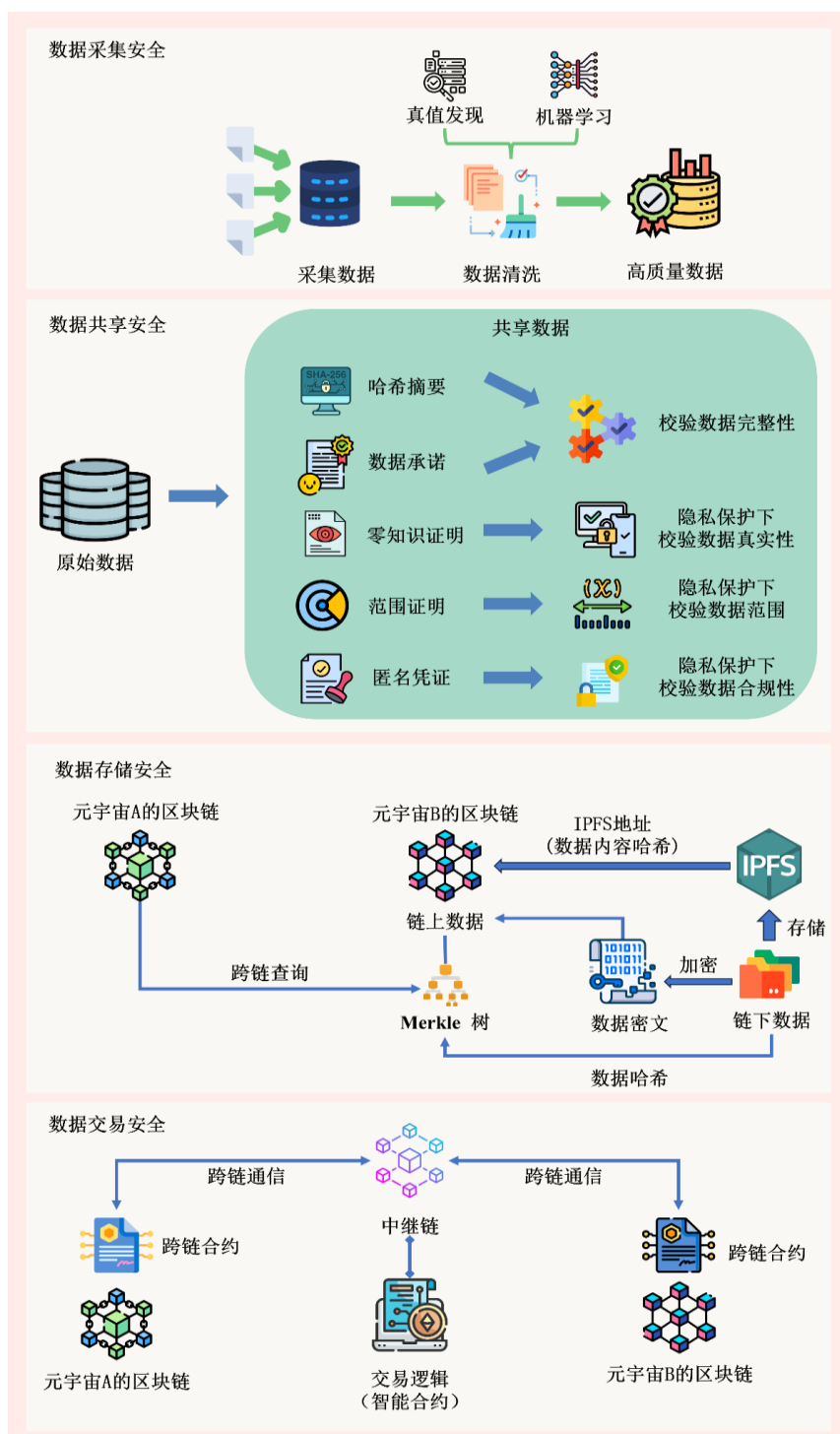


图 2-29 元宇宙协同中的数据安全

- **数据采集安全:** 确保源数据真实可靠，采集端需预处理数据，利用真值发现算法[183]和机器学习进行数据清洗[184]、异常检测，剔除噪声或异常数据，并防御数据投毒攻击[185]，提高数据质量。
- **数据共享安全:** 确保共享数据的完整性和真实性，以保障数据共享的正确性。通过哈希数字摘要[186]和承诺技术[187]，结合零知识证明[188]（如 zk-

SNARKs[189])，在隐私保护的同时保障数据完整性和真实性。引入代理重加密技术[190]，确保数据在传输过程中的机密性。使用范围证明技术[191]，分享者可以在不暴露原数据的情况下共享统计结果和验证结果。结合同态加密或安全多方计算[192]，在加密数据上直接进行可验证计算，避免数据解密过程中的泄露风险。结合匿名凭证技术，验证数据来源的真实可靠。

- **数据存储安全：**为保障数据存储安全，使用区块链技术[193]和 IPFS 存储数据[194]，分布式存储消除单点故障风险，提高抗攻击性。IPFS 通过内容寻址减少链上存储开销。在节点故障时，通过共识机制恢复数据存储。区块链不可篡改的特性保障数据真实性、完整性及可追溯性。结合公开可验证的可检索性证明[195]，提供高效的完整性审计，防止数据在未被察觉的情况下遭破坏或丢失。结合抗量子的加密技术[196]保护数据隐私。结合 Merkle 树技术[197]轻量化链上存储，减少存储开销，提高查询效率。通过跨链查询技术[198]，不同元宇宙间在保持数据隐私和安全前提下进行数据访问和验证，进一步提高协同效率。
- **数据交易安全：**数据资产交易是元宇宙协同的重要目的之一。结合智能合约、区块链和跨链技术，使交易逻辑透明、支付追责自动化，保障数据资产交易的公平性和记录可追溯性。通过细粒度的并发控制与确定性回滚机制[199]，确保高并发、复杂场景下数据处理状态的正确性和一致性。结合匿名凭证、零知识证明和范围证明技术，在保护交易数据与交易者隐私的同时完成资金、数据验证和身份认证，保证数据资产交易的安全可靠。

(4) 计算安全

在元宇宙协同中，涉及大量个人和敏感数据，如用户身份、行为记录和虚拟资产。隐私计算技术对数据进行处理和分析，防止数据泄露和滥用，保护用户的隐私权益保证协同过程的计算安全。这不仅能够提高数据处理的安全性和可靠性，还能增强用户对元宇宙平台的信任，促进元宇宙的健康和可持续发展。如图 2-30 所示，计算安全的主要技术包括同态加密（Homomorphic Encryption）、安全多方计算（Secure Multi-Party Computation, MPC）、可信执行环境（Trusted Execution Environment, TEE）、联邦学习（Federated Learning）和差分隐私（Differential Privacy）等。

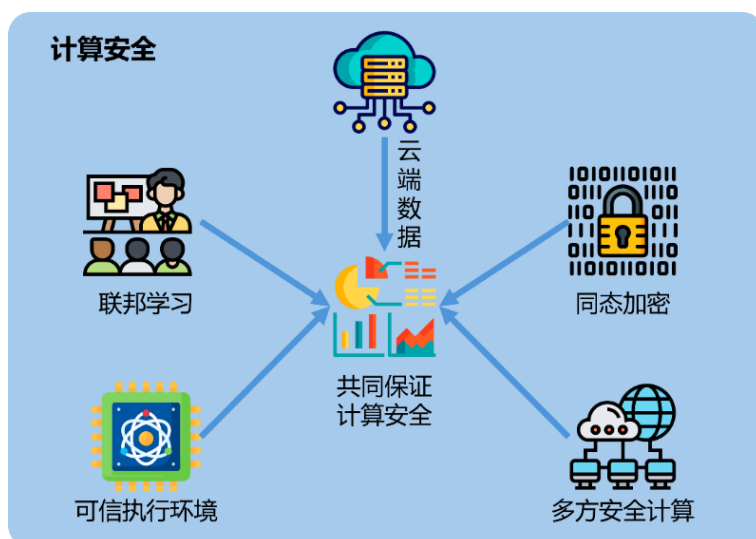


图 2-30 元宇宙协同中的计算安全

- **同态加密**：同态加密允许在不解密数据的情况下直接对加密数据进行计算[200]，从而在云端或其他不可信环境中执行数据处理任务时确保数据的隐私和安全。这种技术通过对加密数据进行数学运算，使计算结果在解密后与对未加密数据进行相同运算的结果一致，从而保护数据的机密性[201]。
- **安全多方计算**：安全多方计算[202]让多个参与方在不泄露各自输入数据的情况下共同完成计算任务。通过复杂的加密协议和算法，确保计算过程中的每一步都保证数据的保密性和正确性，从而实现协同计算的安全。引入神经网络分类器[203]实现对数据泄露漏洞的自动检测，进一步增强了 MPC 系统的实际安全性。
- **可信执行环境**：可信执行环境[204]提供了一个安全区域，允许在主操作系统之外执行敏感代码和处理敏感数据。TEE 依赖硬件支持，防止未经授权的访问和修改，确保计算过程的完整性和机密性。典型的 TEE 如 ARM TrustZone[205]，通过隔离敏感操作和数据，实现更高的安全性；SGX[206]则通过构建的“飞地”抵御包括操作系统在内的特权软件攻击，从而确保关键计算逻辑的安全执行。
- **联邦学习**：联邦学习是一种分布式机器学习技术[207]，允许各方在不共享原始数据的情况下共同训练模型。通过仅交换模型参数而非数据本身，保护数据隐私。MOON (Model-Contrastive Federated Learning) 通过对比模型表示来校正各参与方的本地训练，提高了非独立同分布数据 (non-IID data) 下联邦学习的性能。
- **差分隐私**：差分隐私[208]是一种通过向数据中添加精心设计的数学噪声，来实

现隐私保护的技术。它能够在保证数据可用性的前提下，确保任何单个数据项的存留都不会对统计分析结果产生显著影响，从而防止从计算结果中反推并识别出特定个体的信息。

（5）监管审计

在元宇宙协同场景中，数据的监管与审计至关重要。随着元宇宙协同的进行，用户在虚拟环境中生成的大量数据会一并进行数据流通。由于各个元宇宙之间信任程度不同，数据跨域流通需要得到严格保护，防止数据泄露、滥用和垄断，以确保元宇宙协同过程中的数据安全，推动元宇宙协同健康、有序地发展。

数据监管是保障用户隐私和数据安全的核心。如图 2-31 所示，在元宇宙协同场景下，数据生命周期管理（Data Lifecycle Management, DLM）[209]可以提供全局性的管理框架，从数据生成到销毁的每一个环节进行系统性控制，确保数据在各个元宇宙场景下始终安全、合规且高效。随着元宇宙数据形态的多样化，多模态数据分类技术[210]被引入到监管体系中，通过对行为数据、传感器数据、语音与文本信息以及三维内容进行统一识别与分类，为 DLM 提供更准确的数据类型管理与风险分级依据。与此同时，数据水印（Data Watermarking）[211]技术在微观层面对具体数据进行精准的监控和追踪，增强数据的安全性和可追溯性，实现跨元宇宙数据的有效监管。

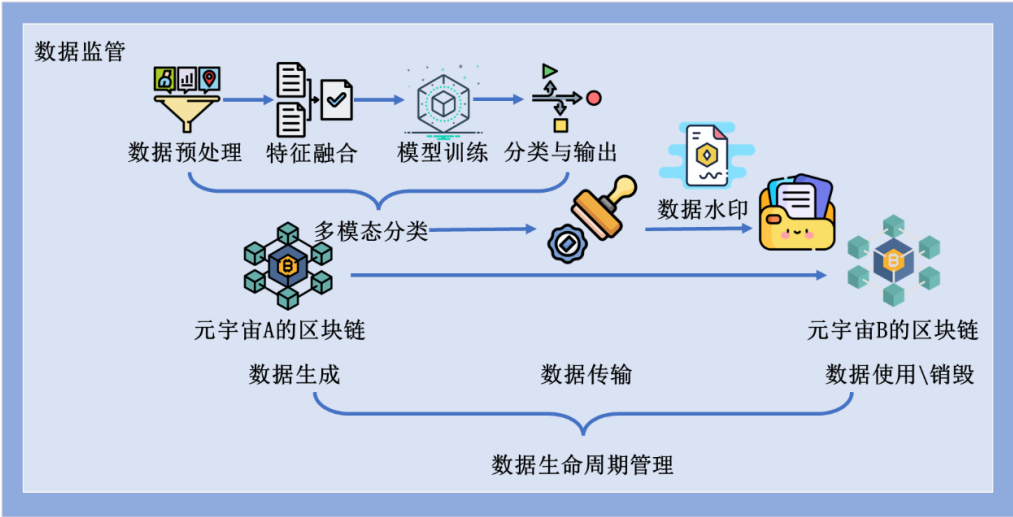


图 2-31 元宇宙协同中的数据监管

数据审计是确保元宇宙企业行为合规的关键手段。如图 2-32 所示，元宇宙协同中的数据审计方法主要包括，通过定期审计，评估和验证元宇宙协同场景下跨域数据管理实践的有效性和合规性，确保用户数据的安全和隐私保护。区块链技术[212]的引入，可以

提升数据审计的透明度和可信度。由于区块链的不可篡改性和去中心化特点，将审计记录写入区块链可以确保所有操作和更改被记录在案，实现审计过程的透明和可追溯。此外，零知识证明(Zero-Knowledge Proof, ZKP)[213]技术也在数据审计中发挥重要作用。零知识证明技术允许在不泄露数据内容的前提下验证数据的真实性和完整性，从而在进行数据审计的过程中保护用户隐私。

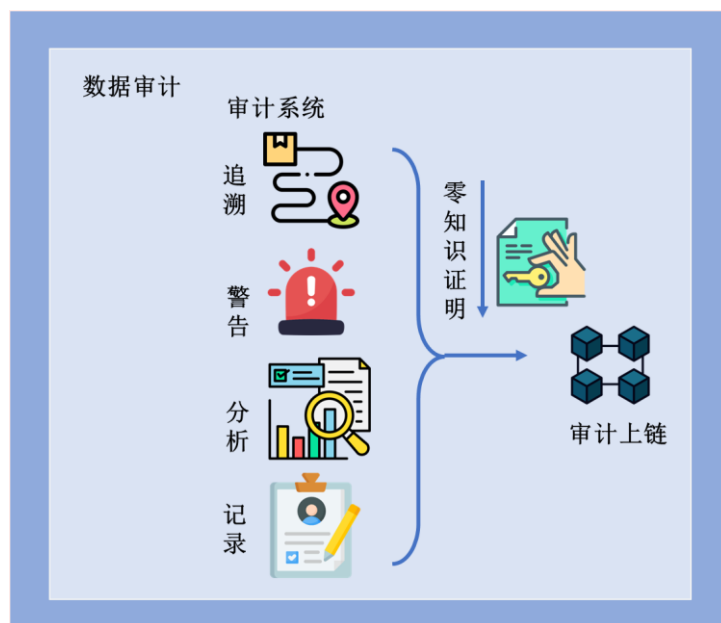


图 2-32 元宇宙协同中的数据审计

综上所述，元宇宙协同中的安全问题主要包括用户身份、通信、数据、计算和监管审计等多个方面。用户身份安全是元宇宙协同的基础，通过去中心化的身份管理技术实现用户身份的安全迁移。通信安全保障元宇宙间的数据交换和交互操作的完整性和机密性，采用安全通讯协议、匿名通信和流量安全技术。数据安全通过保护数据的采集、共享、存储和交易过程中的隐私和完整性来维护元宇宙的交互安全。计算安全则通过隐私计算技术保护用户数据，确保计算过程的安全性和可靠性。监管审计通过数据生命周期管理和数据水印技术，确保数据的安全和合规，增强用户对元宇宙平台的信任。综合运用这些技术，元宇宙协同可以在不同平台和应用之间实现安全、高效的互动，促进元宇宙生态系统的健康发展。

2.6 元宇宙交互

在元宇宙的世界中，交互技术是连接虚拟和现实的重要桥梁。随着科技的发展，交互方式变得越来越多样和复杂。显示技术决定了用户的视觉体验和沉浸感，多模态交互

通过结合多种感官输入提升了用户的参与度，而多模态情感计算则通过分析用户的情感信号增强了情感互动体验。这些技术共同推动了元宇宙交互的进步。

2.6.1 信息显示

人对信息的接受 80%来自于视觉，视觉交互是人类接受外界信息的重要渠道之一。目前的视觉交互都是基于显示技术而实现的，下面对元宇宙中的主要显示技术分别进行介绍。

显示技术的总体分类如图 2-33 所示，其中 2D 显示技术是目前最常用的显示技术，通过在平面或曲面屏幕上显示 2D 图像来传递信息，目前常见的 2D 显示屏有 LCD、OLED、LED、MiniLED 和 MicroLED 等[214]。由于显示幅面以及分辨率的限制，在视觉交互过程中 LCD、OLED 和 LED 的 2D 显示屏无法提供更多的信息。随着 MiniLED 和 MicroLED 等技术的发展，2D 显示的性能得到显著提高。MiniLED 背光技术已实现 2000+级分区控光，峰值亮度突破 10000 尼特，MicroLED 直显产品则达成 4000 PPI 以上超高分辨率，可以传递的信息总量大大提升。此外，由于眼球的特殊生理构造，人眼观看到的范围是一个 120° 的弧形区域，为了充分利用人眼的观看区域，研究者提出了曲面 2D 显示器，相比传统的平面显示器，曲面显示器更加贴合人眼的生理弧度，近年来曲面/异形 2D 显示技术发展迅速，能大幅提升使用者的被包裹感和沉浸感。未来 2D 显示技术更是朝着更高分辨率和柔性等方向发展。

由于人眼在现实生活中接受的是 3D 信息，而 2D 显示技术无法传递物体场景的深度信息，存在信息缺失的问题。3D 显示技术的出现，解决了 2D 显示技术无法传递物体场景深度信息的问题。3D 显示是指采用光电和计算机等多种技术手段来模拟实现人眼的立体视觉特性，将空间物体以 3D 信息再现出来，呈现出具有纵深感的一种显示方式，具有强震撼和高逼真等特点。3D 显示技术是元宇宙交互领域的关键技术，通过 3D 显示设备进入元宇宙可以将虚拟世界与现实世界进行密切融合。3D 显示技术主要分为助视 3D 显示技术和裸眼 3D 显示技术[215]。

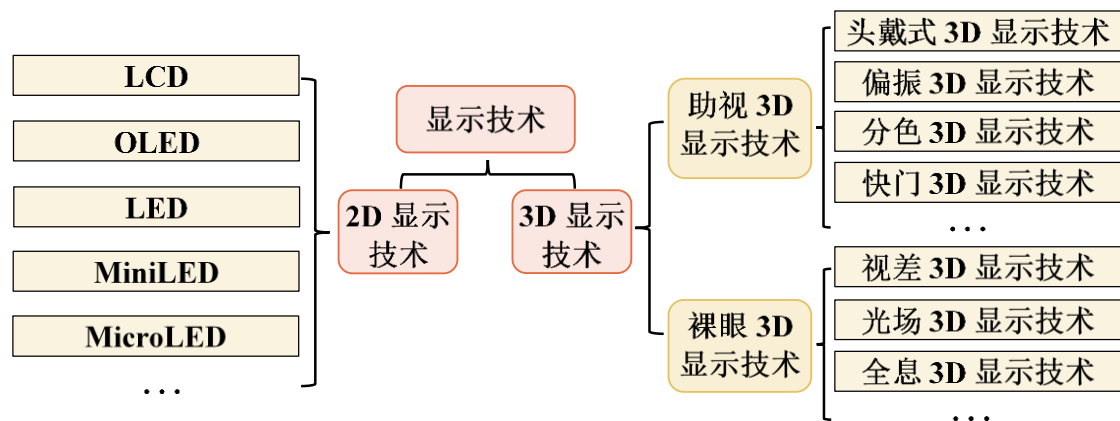


图 2-33 显示技术的分类

助视 3D 显示技术主要包括头戴式 3D 显示技术、分色 3D 显示技术、偏振 3D 显示技术和快门 3D 显示技术。其中头戴式 3D 显示技术是目前元宇宙的主要显示。头戴式 3D 显示技术是将显示设备集成在一个紧凑的外壳中，通过提供空间计算和视觉平台，实现虚实信息的无缝融合，从而为观看者提供更好的沉浸式体验和更自然的人机交互方式[216]。传统的透射式头戴显示技术基于反射和折射透镜元件来实现，视场角受到透镜厚度和焦距的限制，同时系统体积较大，影响佩戴舒适度[217]。近年来，随着光学领域成像技术的不断发展，自由曲面头戴显示技术、波导头戴显示技术、视网膜投影技术和全息头戴显示技术等逐渐成为研究热点[218]。自由曲面头戴显示基于没有任何对称性的复杂曲面来设计，在扩大视场角的同时提高了光学透射率，然而系统重量和体积较大。波导头戴显示技术利用波导元件对传输光线进行全反射，大幅减小了光学元件的厚度，使得系统轻薄化，同时外界场景也可直接进入人眼，实现了虚拟与现实的融合。视网膜投影技术利用 **Maxwellian view** 原理，保证很小光束的光线直接进入人眼即可看到图像，该技术一般采用激光扫面系统作为显示源，存在出瞳小、人眼观看范围受限的问题。全息头戴显示技术基于光的干涉和衍射原理，能够精确调控光波前，提供人眼视觉所需要的全部深度信息，记录并重建出原物体逼真的立体像，为头戴式近眼显示技术的发展提供了新的方案，目前成为了最具潜力的头戴显示技术之一。随着人工智能技术的蓬勃发展，波导全息显示技术与人工智能驱动算法相融合，显著提升了头戴式近眼显示系统的紧凑性[219]。未来应用于元宇宙交互技术的头戴显示技术正朝着轻量化、大视场角和高分辨率等方向发展。

裸眼 3D 显示技术主要分为视差 3D 显示技术、光场 3D 显示和全息 3D 显示技术等[220]。目前应用于元宇宙交互技术中主流的 3D 显示技术为视差 3D 显示技术。视差 3D

显示技术是通过光栅或柱透镜将显示器显示的图像进行分光，根据人眼所在位置，对光线传播的路径进行一定方式的控制，使人的双眼能够接收到不同的图像，经过大脑融合之后形成 3D 视觉。视差 3D 显示技术可以让人区分物体的远近，从而为观看者提供真实的 3D 信息，具有成本低和结构简单等优点。但由于人观看视差 3D 显示器时，聚焦位置固定在显示器所在平面上，在大脑融合视差信息时存在辐辏冲突，因此长时间观看会存在眩晕不适。光场 3D 显示技术通过采集记录人眼观察到的真实 3D 场景的光场分布，在空间中再现完整的光场信息，从而使人双眼的聚焦位置和会聚角度符合现实物体的规律，消除了 3D 眩晕，并具有可多人观看等优点，成为了最具发展潜力的裸眼 3D 显示技术之一。全息 3D 显示技术通常需要借助计算机来生成全息图，同时通过空间光调制器来对输入光进行调制进而再现出 3D 场景。得益于人工智能技术的快速进步，全息 3D 显示技术正处于一个新的发展浪潮之中。然而，全息 3D 显示尚存在计算速度慢、视角小和分辨率低等问题。因此，现有的全息 3D 显示的性能离市场应用需求仍有一定的距离。目前，实时大视角高分辨率全息 3D 显示技术是全息 3D 显示技术的主要发展方向。

2.6.2 多模态交互

在元宇宙中，交互方式的升级是其核心特征之一。这种升级不仅体现在交互界面从传统的二维变为沉浸式的三维，还体现在交互模态从单一模态向多模态的演变。多模态交互，如图 2-34 所示，让用户能够更加自然和直观地与元宇宙中的目标进行互动。

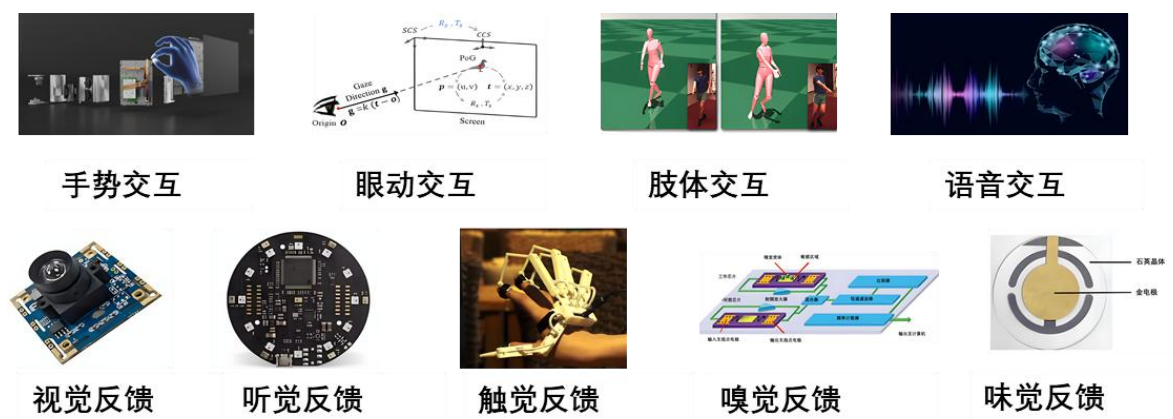


图 2-34 多模态交互

在输入方式方面，用户可以通过手势、眼动和全身动作进行交互，这些方法提升了

用户体验的沉浸感和真实感。此外，语音交互也成为了一种重要的输入方式，由于其应用范围广泛，本文不作详细讨论。

在输出方面，多模态交互不仅限于视觉和听觉，还包括触觉、嗅觉和味觉的体验。其中，触觉交互是当前研究的热点，致力于通过触觉反馈增强用户在虚拟环境中的真实感和互动性。通过多模态的输入和输出，元宇宙为用户提供了更加丰富和多样化的互动体验，开启了全新的数字互动时代。

（1）手势交互

手势交互技术是通过捕捉和识别用户手部动作来实现人与计算机的自然互动。硬件方面，手势识别系统需要高帧率成像设备，以确保流畅的交互体验。Leap Motion 等设备基于结构光和红外技术，通过发射红外光线并捕捉反射数据，精确地重建手部的 3D 模型。软件方面，手势识别技术经历了从传统计算机视觉方法到深度学习和变压器架构 [221][222] 的演进，提升了识别的准确性和鲁棒性。这些进展使得手势识别技术在虚拟对象操作、界面控制和虚拟社交互动等应用场景中，提供了自然、直观且高效的交互体验。

尽管手势交互技术取得了显著进展，但仍面临诸多挑战。首先，如何在自然环境中实现高鲁棒性的手势识别是一个主要难题。现有的许多方法在处理遮挡和光线变化时表现不佳。高动态范围（HDR）传感器、深度传感器、红外摄像头（如 Leap Motion 使用的红外技术）以及先进图像处理算法和机器学习技术的结合，能够在各种光线条件和部分遮挡情况下保持识别的稳定性和准确性。此外，双手交互场景中的手势识别更加复杂，因为两只手之间的相对位置和互动关系增加了算法的复杂性。为了解决这些问题，4DHands [222] 方法探索了关系感知标记和时空交互推理模块，从而提升了复杂场景下的手部重建精度和稳定性。研究者进一步利用 Transformer 架构处理手与物体的复杂交互，例如 EasyHOI [223] 提出了基于大规模视觉模型的手-物交互重建框架，在面对复杂遮挡、强镜面反射、手指与物体的细粒度接触关系等真实世界场景时表现出显著的稳健性。

这类方法的出现表明，随着预训练视觉模型、Transformer 架构和轻量化优化策略的结合，复杂手势与手-物体交互的建模正在从“局部几何推断”迈向“全局语义理解”。通过对更高层特征的捕捉和跨实体关系的统一建模，手部姿态估计在复杂环境下的精度与稳定性将持续提高，如图 2-35 所示。

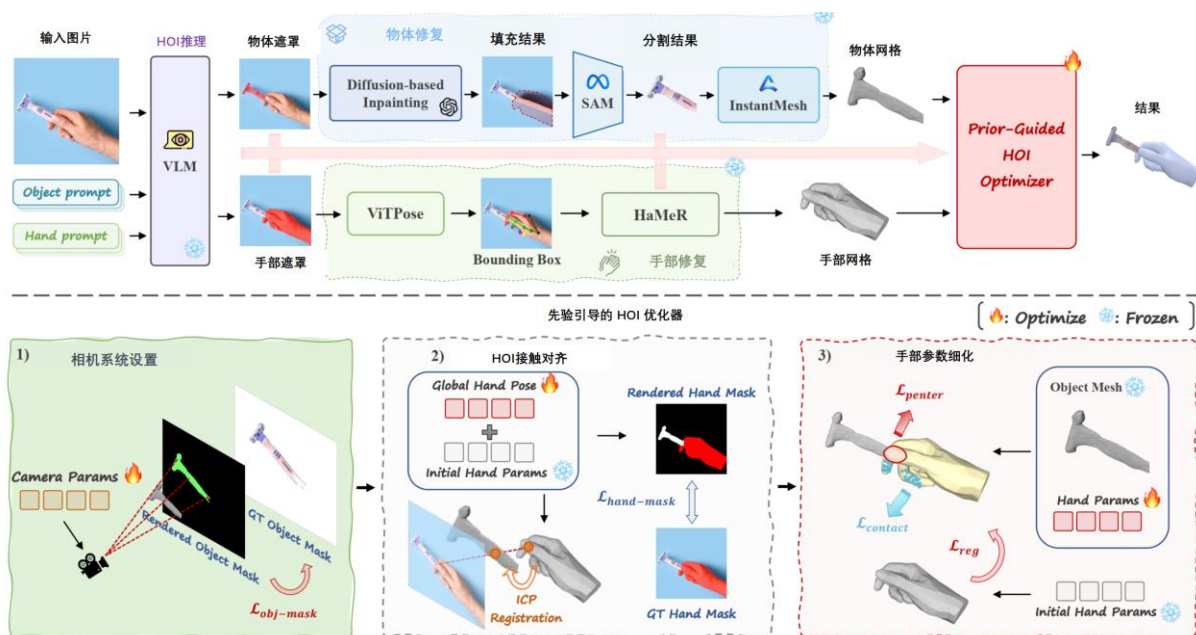


图 2-35 基于大规模视觉模型的手-物交互重建框架[223]

未来,手势识别技术的发展趋势集中在提高精度、降低延迟以及适应复杂环境。2024年发布的消费级空间计算设备(如 Apple Vision Pro)在系统层面采用“注视选中+轻捏确认”的眼-手融合范式,并在渲染管线中普遍集成基于注视的焦点渲染技术,进一步验证了高精度手势与眼动追踪融合作为自然交互主范式的可行性和市场接受度[224]。机器学习和深度学习的持续进步,将推动更复杂和精确的手势识别模型的研发,增强用户在虚拟环境中的自然交互体验。Leap Motion 等基于结构光和红外技术的设备将继续在手势识别中发挥重要作用,通过高精度的 3D 手部模型重建,实现更加细致和准确的手势识别。同时,通过整合多种传感器数据和优化算法,手势识别技术将在不同光线条件和部分遮挡情况下表现出色,并减少感知和计算延迟,提高交互流畅度。手势识别技术的不断进步,将为用户提供更加直观、沉浸和高效的交互体验,推动元宇宙技术的发展与普及。

(2) 肢体交互

肢体交互技术通过检测和解释人体动作及姿态,实现人与计算设备或环境之间的自然互动。常见的技术包括基于视觉、惯性传感器、无线电信号和电场感应的解决方案。其中,视觉传感技术(如 RGBD 和红外摄像头)因其数据收集和标注方便而广泛应用于人体动作识别(HAR)和人机交互(HCI)。惯性传感器技术(如 IMU)因其低功耗、小体积和低成本的特点,在长时间穿戴场景中也有广泛应用。然而,这些技术在实际应

用中仍面临遮挡、光照条件和长期漂移等问题。

最近，研究人员探索了一些新兴的感应技术，以弥补现有解决方案的不足。其中，基于体域电场的感应技术[225]因其非侵入性、高灵敏度等特点，逐渐受到关注。这种技术可以通过检测人体周围电场的变化，识别人类动作和姿态，应用于智能家居、健康监测等领域。

尽管肢体交互技术在多种应用中显示出巨大的潜力，但其广泛部署仍面临诸多挑战。首先是鲁棒性问题。例如，基于视觉的解决方案容易受到光照和遮挡的影响，尽管 ViTPose++[226]等基于 Transformer 的新一代通用体姿态估计模型已大幅提升了其鲁棒性[227])，而惯性传感器的长时间使用可能导致数据漂移。其次是通用性问题，即在不同环境和使用者间的一致性能。环境变化（如家庭电器的开关）和使用者的个体差异（如身高、体重和穿戴物品）都会对感应结果产生影响，从而降低系统的准确性。

此外，基于体域电场的感应技术虽然具有高灵敏度，但也容易受到环境噪声的干扰。例如，电器设备的电磁干扰可能会影响感应结果，导致误检或漏检。因此，提高系统在复杂环境中的鲁棒性和稳定性是当前研究的重点。

未来，肢体交互技术的发展将朝着更高的准确性、鲁棒性和易用性方向努力。集成电容芯片的进步和主动屏蔽技术的应用，有望显著提高基于体域电场感应技术的性能。同时，多传感器融合和连续学习算法的应用，将有助于提高系统的通用性和适应性，使其能够在不同的环境和使用者间保持一致的性能。

在具体应用方面，随着穿戴设备和智能家居技术的发展，基于体域电场感应的解决方案将在健康监测、姿态识别和智能交互等领域发挥重要作用。例如，通过感应电场变化进行步态分析和手势识别，可以实现更加自然和高效的人机交互体验。此外，随着技术的成熟，基于体域电场的感应系统有望在医疗、娱乐和安全等领域得到广泛应用。

(3) 眼动交互

眼动追踪是一种通过检测和跟踪用户眼球的运动，确定用户注视点和视线方向的技术。其工作原理包括数据采集、预处理、瞳孔检测、视线计算、应用映射和反馈机制。首先，系统通过高分辨率摄像头或红外光传感器捕捉眼部图像，利用图像处理算法检测瞳孔位置，计算视线方向并将其映射到屏幕上的注视点。这些信息用于实现无接触的界面操作和用户行为分析。眼动追踪算法包括视线估计部分和标定部分，主要技术路径为瞳孔角膜反射法[227][228]和基于图像的方法[229][230]。前者通过瞳孔位置和反射点

计算视线方向，如图 2-36 所示，后者利用深度学习算法从眼部图像中提取特征进行视线估计，如图 2-37 所示。标定方法包括静态标定、平滑追踪标定、互动标定和隐式标定，确保眼动数据的准确性和可靠性。

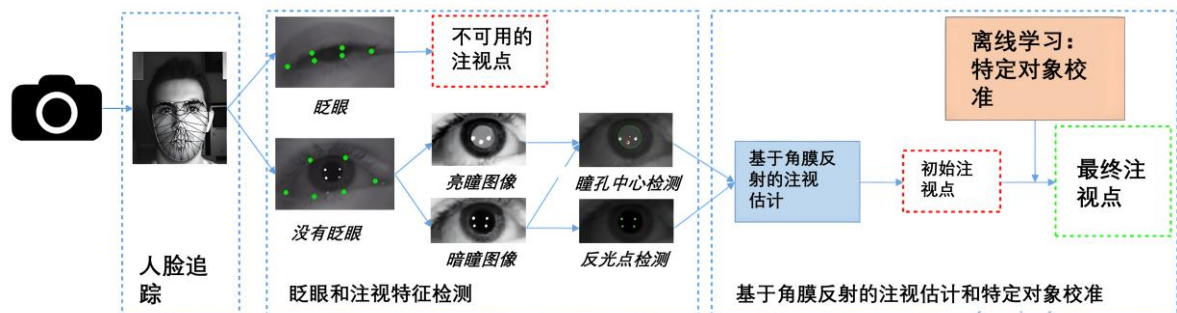


图 2-36 一种单目瞳孔角膜反射法[227]

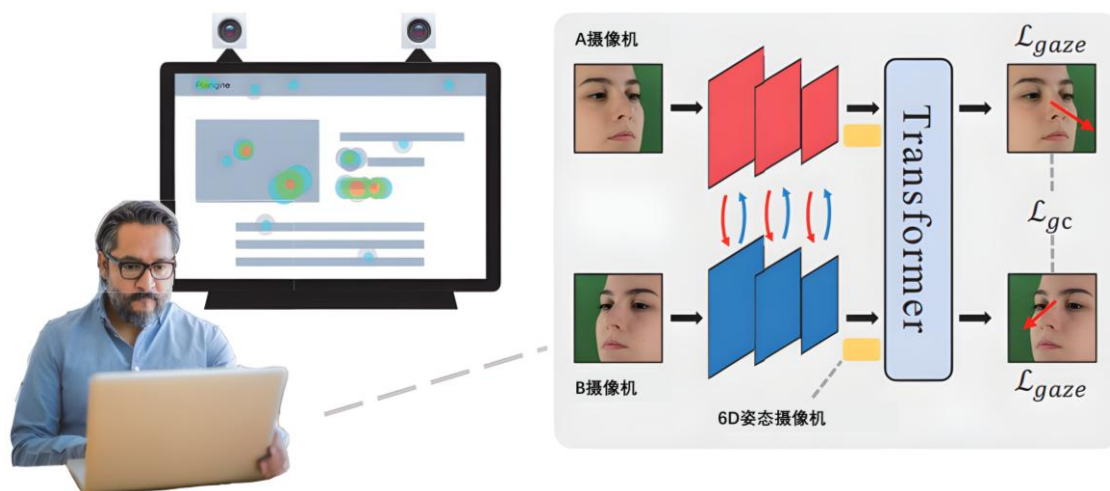


图 2-37 一种双目深度学习视线估计算法[230]

眼动追踪面临的主要技术挑战包括提高精度和响应速度。高精度眼动追踪需要先进的传感器和复杂的算法，但这些技术要求会带来高计算负荷，可能影响实时响应速度。为解决这一问题，需要优化算法效率，提升硬件性能，并利用并行处理和专用加速芯片，以确保在不影响精度的前提下，实现快速、实时的眼动追踪。此外，不同用户在使用眼动追踪设备时存在适应性问题，传统系统应采用个性化校准和自适应算法，确保设备在多样化用户群体中的高精度和一致性。大规模视觉-语言预训练技术也开始用于眼动估计。例如 GazeCLIP[231]通过引入文本提示与图像特征的多模态融合，提高了注视估计在复杂光照、视角变化等条件下的泛化能力，并在跨数据集评测中减少了额外标定的需求，提升了系统在低标注环境下的易用性。

眼动追踪技术的发展趋势主要体现在优化渲染、视线控制和个性化推荐方面。焦点

渲染技术（已在主流 VR/MR 设备中得到商业化应用[224]）通过在用户注视的区域提供高分辨率图像，提升图像质量和渲染效率。视线控制允许用户通过注视实现无接触的交互方式，提高自然性和卫生性。个性化推荐则利用用户的视线轨迹数据，推断其兴趣和偏好，提供个性化内容推荐，提升用户体验。此外，未来的眼动追踪技术还将继续优化精度和响应速度，提升硬件性能和算法效率，确保在多样化应用场景中的广泛应用和普及。

（4）触觉反馈

触觉反馈是指通过物理设备再现虚拟世界中人所接触物体的纹理、温度、振动、形状、刚度、压力等触力觉体验。触觉体验提升了用户的沉浸感，允许用户“触摸”虚拟对象，使人在元宇宙中获得更真实的交互体验。

面向元宇宙的触觉反馈设备主要可以分为穿戴式和裸手式两大类，如图 2-38 所示。穿戴式通常以手套或其他可穿戴装置的形式出现。在穿戴设备的内部通过集成柔性传感器、力反馈结构及温度、纹理、振动等触觉反馈设备，将虚拟环境中的触觉信号准确传递给用户。目前，已经有一些商用产品如 Dexmo[232]、HaptX[233]和 SenseGlove[234]等，实现了基于手套的多种触觉模态融合反馈，但与高保真触觉反馈相比仍存在一定差距。例如，Dexmo 手套能够提供五指力反馈，使用户在虚拟环境中感受到抓握物体的阻力[232]。HaptX 手套利用微流体技术，模拟皮肤的细微触觉，提供高度逼真的触觉体验[233]。在学术界，手套也逐渐从刚性结构向柔性化轻薄化发展，Mate Jam 采用层阻塞的原理实现了手套的柔性化，并可以实现刚度反馈[235]。这类设备的主要优势在于其精确的触觉反馈能力，但其缺点在于通常体积较大，佩戴不够舒适，且价格较高。

裸手式则依靠外部装置产生触觉反馈，用户不需要佩戴任何设备，可以主动探索虚拟世界，其实现方式为通过检测人手位姿，在特定位置放置触觉反馈设备，实现触觉显示。现有实现方式有移动小车[236]、机械臂[237][238]、无人机[239]及一些用于人手精细触觉显示的机械点阵[240][241][242][243]。此外，在产业界，超声波阵列或激光投影仪，直接作用于用户的手部皮肤，而无需佩戴任何设备。Ultraleap 公司研发了一种基于超声波技术的裸手式触觉设备，通过空气中的振动波在用户手上产生触觉反馈[244]。相比于穿戴式触觉反馈设备，裸手式触觉反馈设备大部分仍处于科学研究阶段，距离工程化应用仍有一定困难。

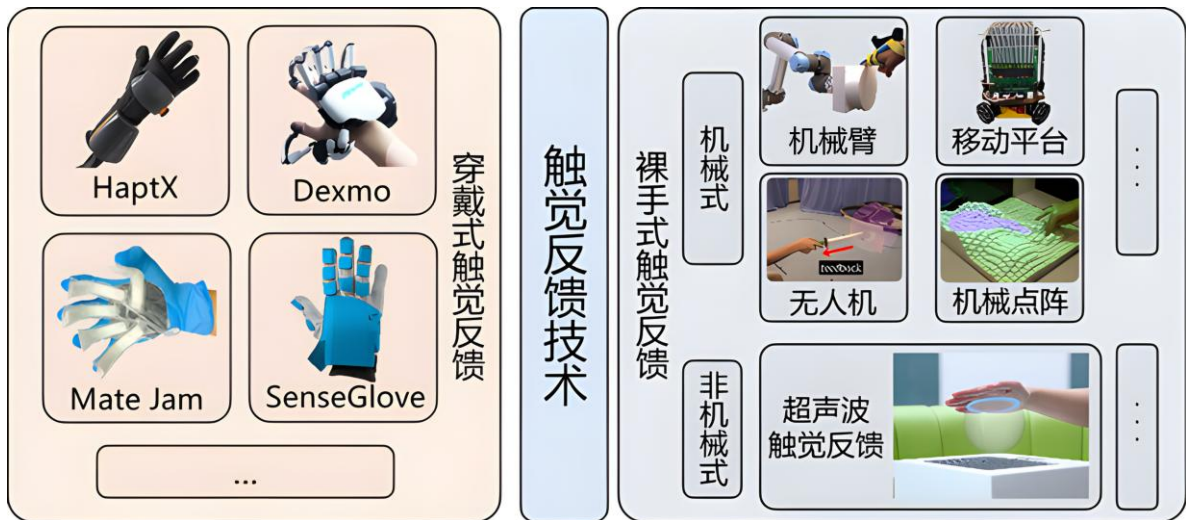


图 2-38 触觉反馈技术的分类

尽管触觉反馈技术有巨大潜力，但仍面临诸多挑战：1) 技术复杂性：为模拟逼真触觉体验，触觉反馈系统需要高灵敏度的传感器感知用户动作，采用高空间分辨率、快速响应的执行器，做出相应触觉反馈。2) 硬件成本：高质量的触觉设备通常价格昂贵，限制了其大规模应用。尤其是在消费级市场，成本效益是一个重要因素。3) 内容兼容性：触觉反馈技术需要与虚拟内容紧密结合，目前许多虚拟环境和应用尚未完全支持或优化触觉反馈，这限制了技术的普及和用户体验的提升。

未来趋势包括：1) 微型化和便携化：触觉反馈设备将变得更加小型化和便携化，使其更易于集成到各种设备中，如 VR 头盔、手套和穿戴设备。2) 软硬件一体化发展：触觉反馈技术将与虚拟环境内容开发紧密结合，形成软硬件一体化解决方案。3) 多模态交互：虚拟世界中的纹理、温度、振动、形状、刚度、压力等多模态反馈将时空一致地呈现给用户，并与其他感官反馈（如视觉、听觉）相结合，将大幅提升元宇宙的沉浸感和互动性。

2.6.3 多模态情感计算

情感计算 (Affective Computing) 是元宇宙交互的重要组成部分，其研究目标在于让计算机具备感知、理解与表达情感的能力，从而构建更人性化、更具亲和力与信任度的人机交互环境[244][245][246]。这一研究范式起源于对人类情感机理与人机协同关系的系统性反思：在沉浸式、强交互的虚拟空间中，只有当系统能够对情绪线索进行稳定、连续且可解释的处理，并以合适的方式回馈或生成情感表达时，用户才会产生“被

理解”的体验与更高的社会在场感[247][248][249][250]。因此，如图 2-39 所示，情感计算不仅关注情绪标签的判别[251][252]，更关注跨时域的情绪动态[253][254]、强度估计[255]以及与语境、意图之间的耦合关系[256][257]，这些能力共同构成元宇宙情感交互的底座。

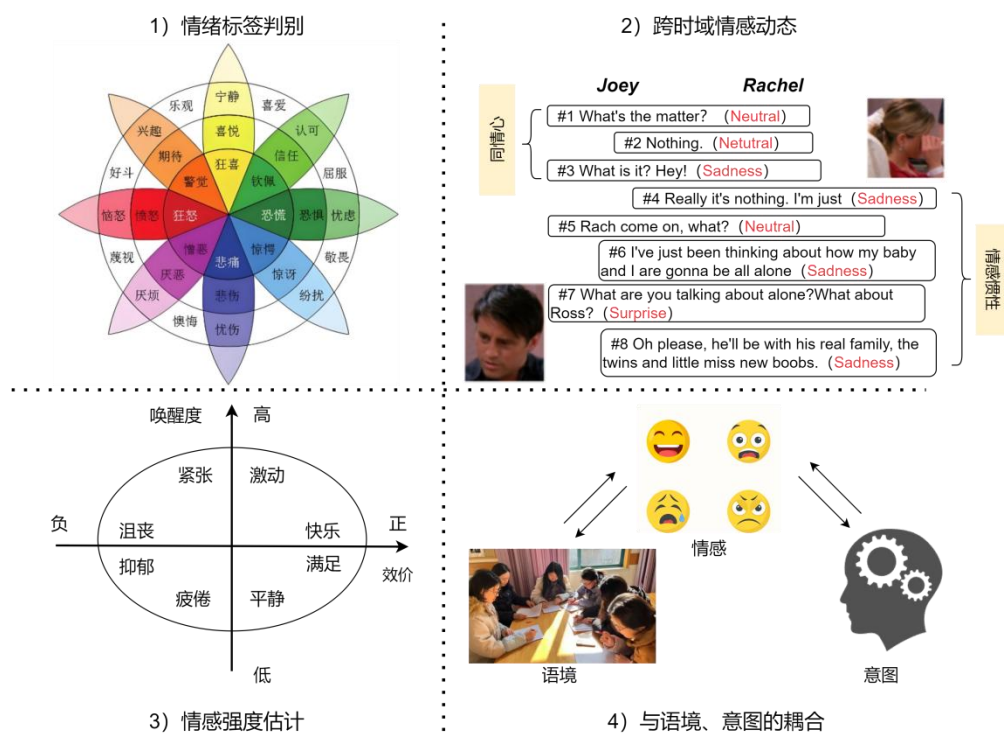


图 2-39 情感计算主要关注：（1）情绪标签判别（2）跨时域情绪动态（3）情感强度建模（4）多因素耦合

人类情感的表达往往以多种模态并行出现[251][253][257]，既包括显性线索（如人脸表情[258]、声学语调[259]、肢体动作[254]与生理反应[260]），也涵盖隐性线索（如文本语义[261]、图像风格[262]）。在显性层面，情绪常伴随可观测的生理与心理变化，但此类信号易受环境噪声、遮挡、刻意压制或设备条件影响而失真[263]，使模型在元宇宙这类高自由度、高变异度场景中难以仅凭显性模态获得稳定表现。与此同时，移动终端与在线社交的普及促使用户以文字、图像、音频与视频分享经验与观点，隐性情感线索因而与数字媒介深度耦合[245]；对语义内容、视觉风格与叙事结构的联合分析可间接推断个体的情绪状态与态度倾向[262][264]，而表情符号、网络流行体与群体互动模式又会进一步调制表达的强弱与方向[265][266]。然而，隐性线索同样受语义歧义与语用反讽[267][268]、跨文化与社群差异[269][270]，以及模态稀疏与缺失等因素[271][272]制约，单独使用会导致偏差与不稳定。基于此，显性与隐性线索并非割裂，而是在真实交互中相互印证、彼此补充；任何单一管道的缺失或退化都可能导致不确定

性并引发对总体情绪状态的误读。由此可见，相较单模态方法，多模态情感识别在信息互补性、模型鲁棒性与泛化性能上更具优势，能够更好地满足元宇宙交互对沉浸性、在场感与社会信任的综合诉求。

在情感计算领域，多模态情感识别作为实现人机情感智能交互的核心技术，其方法体系已形成多元化发展格局。如图 2-40 所示，当前主流技术路径可凝练为三大核心范式：融合驱动型方法、图结构建模方法与生成式建模方法。需着重说明的是，多模态数据固有的互补性与复杂性，使得部分先进技术方案常呈现跨组件融合特征，例如嵌入图结构先验的融合框架、基于生成模型的跨模态融合范式等。本分类体系的界定核心，在于锚定各类方法最具代表性的建模范式与技术内核，而非局限于单一组件的存在形式，确保分类的逻辑性与包容性。

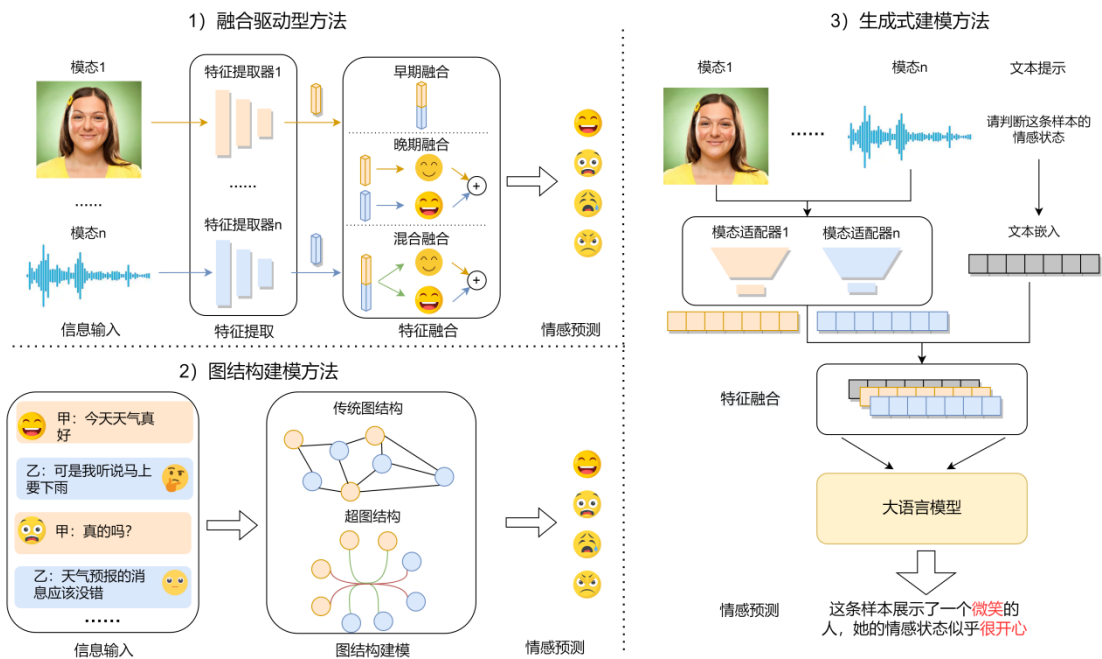


图 2-40 多模态情感识别三大核心范式：（1）融合驱动型方法（2）图结构建模方法（3）生成式建模方法

（1）融合驱动型方法

特征融合是多模态研究中的核心议题之一，也是多模态情感识别中最早被系统化研究、应用最为广泛的一类技术路径。基于融合的方法通常假定不同模态分别承载互补的情绪线索，通过在特征空间或决策空间中整合这些信息，获得更为完整的情感表征，再在此基础上预测情感状态或极性[273][274][275]。在这一框架下，如何以合适的粒度与形式对多模态特征进行对齐与联合建模，成为影响模型性能与鲁棒性的重要因素。

从融合操作在处理流程中的位置划分，主流方法可大致分为早期融合、晚期融合以及混合融合三种范式[273][274]。早期融合亦被称为特征级融合，通过对来自不同模态的输入特征在时间或语义维度上进行对齐与变换，再将其直接拼接或投影到统一的联合空间，形成一体化的多模态表征[276][277]。这一策略在理论上为各模态之间的信息交互提供了充分空间，有利于模型在后续编码阶段同时关注跨模态关联与互补线索，从而提升情感表示的整体表达力。不过，早期融合也面临特征维度急剧增大带来的计算复杂度上升[274]、不同模态节奏与语义对齐难度较高[275]、以及各模态对融合表征的贡献不均衡[278]等问题，若缺乏合理的对齐与加权机制，模型性能容易受到影响。

晚期融合又被称为决策级融合，其典型做法是为每种模态分别构建独立的预测分支，由各自输出情感预测结果，再采用如简单平均[279]、加权投票[280]、或基于置信度/方差的加权规则[281]对多路预测进行整合，得到最终情感决策。相较于早期融合，晚期融合在系统设计上具有更高的模块化与灵活性：不同模态的特征提取与分类器可以独立优化，并且在部分模态缺失、噪声较大或暂不可用时仍能维持基本的推理能力[282]。然而，由于模态间交互主要发生在决策层，晚期融合对跨模态相关性与深层语义耦合的利用相对有限，最终输出在全局协调性与信息利用效率方面往往难以达到最优[275]。

混合融合策略则尝试在统一框架下结合早期与晚期两种思路：一方面在特征层面进行适度的对齐与联合建模，捕获关键的跨模态关联；另一方面在决策层面保留各模态预测分支，通过再融合机制进行多级整合[283]。此类方法在一定程度上兼具特征级信息交互与决策级鲁棒性的优势，通常能够获得更具表现力的多模态情感表征。然而，混合融合在模型设计与计算资源方面的开销也相应增加，如何在性能收益与复杂度之间取得平衡，成为这一路线进一步发展的重要问题。

在按照融合时机对多模态特征融合进行划分之外，还可以从模态贡献与权重分配的角度对融合方法进行进一步区分。根据不同模态在融合过程中的地位与作用，现有基于融合的多模态情感识别方法可分为平等权重融合和模态主导融合两大类。其中，传统的多模态融合方法多属于平等权重融合范式[273][274][275][276][277]。这类方法在设计上通常假设各模态在情感表征中具有相近的重要性，因而在特征拼接、加权求和或联合编码时倾向于采用对称或固定的权重配置，使模型可以在形式上“同等地”接收来自不同模态的信息。这一策略有利于充分保留多源特征，避免在建模初期引入过强的偏置。然而，在实际应用中，不同模态在信息质量、噪声水平、语义完整度等方面往往存在显著差异，平等权重机制在长程训练过程中容易出现“模态惰性”现象[278]：部分模态由

于噪声大、信号弱或学习难度高，实际被模型“边缘化”，对最终决策贡献有限，而另一部分模态则在反向传播中占据主导，导致名义上的多模态融合在效用上趋近于“单模态主导”。

为缓解上述问题，研究者提出了模态主导融合的思路。早期工作多以前景较为清晰、语义信息相对完整的文本模态作为主要模态，将语音、视觉等其他模态视作辅助信号，通过在注意力、门控或残差通道中为文本特征提供增强与校正[284][285][286][287][288]。在这一框架下，文本承担语义主骨架，声学与视觉模态为其补充情绪色彩、说话人状态和细粒度行为线索。随着对模态间关系理解的加深，后续研究进一步提出了动态权重融合策略[278]。这类方法不再预设“文本一定是主模态”的前提，而是通过显式建模任意两两模态之间的相关性或依赖性，在样本或时刻级别自适应地估计各模态的重要程度。据此，模型可以在不同语境下选择更为“主导”的模态，并利用该模态的表征去补充和引导其他模态的特征学习，以获得更加贴合当前情境的多模态联合表示。

(2) 图结构建模方法

早期的情感识别任务多聚焦于单句级别，即在相对独立的语句或短文本上预测情感类别。这类任务通常忽略说话轮次之间的上下文联系以及说话人之间的互动信息，难以充分反映真实对话场景中的情绪变化。随着研究视角逐渐从孤立语句扩展到多轮对话，对话情感识别逐步成为多模态情感识别中的重要方向：模型不仅关注当前语句本身的内容特征，还需要同时关注前后语境、说话人身份以及轮次间的情绪演变模式。在这一背景下，对话被视为由多轮、多说话人参与构成的结构化序列，内部包含丰富的依赖关系与互动线索。为了更有条理地表达这些复杂关系，对话常被组织为图结构，将语句、说话人乃至模态信息抽象为节点或高阶连接单元，以刻画轮次之间、角色之间及模态之间的语义联系与情绪关联。基于图结构的建模方式为对话情感识别提供了超越线性序列的表示空间，使情绪分析能够在更贴近真实交流形态的结构中展开。

在传统图神经网络（Graph Neural Network, GNN）中，DialogueGCN[289]率先将图网络引入对话情感识别任务。与以往基于序列的模型（如 DialogueRNN[290]）相比，图结构能以更加灵活的方式表征对话中的多方交互关系。为整合来自文本、语音与视觉等不同模态的数据，后续研究进一步探索了基于图卷积的多模态融合策略，通过构建多种类型的边与节点，实现模态间信息的关联建模与语义互补特征的整合[291][292]。与

此同时，一些方法将跨模态注意力机制融入图神经网络框架，利用显式注意操作强调模态间差异性与互补性，使模型能够更充分地利用异质特征并增强跨模态语义关联能力[293][294]。

随着研究的深入，传统图结构中“单边仅连接两个节点”的限制逐渐显现。超图神经网络（Hypergraph Neural Network, HNN）因其可在单个超边中同时连接多个节点而受到关注。已有工作基于超图结构构建更灵活的对话表示框架[295][296][297][298]，能够在同一结构中联合建模多模态特征、对话轮次与说话人信息。通过更丰富的连接方式，超图结构能够表达更高阶的协同关系、多粒度语义模式以及跨模态间的多变量交互，使对话情感的整体动态更加清晰可辨。

在深层图网络中，随着网络层数的增加，表征容易趋向同质化，这类现象通常表现为难以维持长程信息的差异性 or 削弱模态间的互补特征[299][300]。为缓解这一问题，近期研究提出了傅里叶图神经网络（Fourier GNN）框架，并将其应用于多模态对话情感识别场景[301][302]。相关方法通过在频域空间中处理图信号，提取高频与低频成分，以突出局部语义变化模式与整体情绪走势。频域表征在多模态语义一致性、变化敏感性与结构层次分析方面表现出较强潜力，使模型能够更好地处理对话中跨轮、跨模态的细微情绪线索。

（3）生成式建模方法

基于生成的方法指的是利用大语言模型生成中间推理或是预测结果的方法。近年来，预训练的大语言模型不仅在自然语言处理任务中取得了显著成功[303][304]，在多模态学习领域也涌现出许多优秀的工作[305][306]，其中不乏与情感识别相关的成果[307][308][309]。此类方法通常采用提示工程或指令调优进行模型的训练。然而，尽管大语言模型表现出强大的通用能力，在特定子任务上仍需要精心设计高质量提示与改进策略，充分激发大模型的潜力，弥合推理能力的差距[308][309][310]。例如，Zhang 等人[308]利用图像和文本数据构建了高质量指令数据集，并使用低秩适应（Low-Rank Adaptation, LoRA）[311]对多模态大语言模型进行微调，使其适应情感识别任务。Dutta 和 Ganapathy[309]则引入了声学 and 文本两种模态，在大语言模型的加持下进行情感识别。此外，Fu 等人[310]引入了基于行为感知的多模态大语言模型框架，该框架视频衍生行为生成模块、行为对齐与细化模块以及指令调优模块构成。前两个模块使模型能够从有限信息中推断人类行为，从而增强其行为感知能力。指令调优模块通过对接的多模

态输入进行对齐和微调，提升模型的情感识别性能。

随着大语言模型参数规模与上下文窗口不断扩张，情感识别的训练与推理成本显著上升，促使研究从“堆算力”转向“降成本、提适配”。受领域专用大型语言模型（domain-specialized LLM）范式启发[312][313]，MSE-Adapter[314]提出了一种轻量化、可插拔的适配架构：由 TGM（Text-Guide-Mixer）与 MSF（Multi-Scale Fusion）两部分构成，其中 TGM 以文本引导对非文本模态（如声学、视觉）进行语义对齐与调制，缓解跨模态语义失配；MSF 则通过多尺度、跨模态特征融合在保持细粒度线索的同时压缩推理开销。该设计在不破坏原主干模型的前提下，有效平衡了计算成本与识别性能，体现出“参数高效迁移+多模态可扩展”的工程优势。与此同时，SpeechCueLLM[315]以“最小侵入”为目标，构造将原始语音线索转换为自然语言提示（prompt）的轻量插件，使 LLM 无需改动架构即可调用语音情感证据，从而以语言中间表征打通语音—文本的推理通道，进一步降低了多模态适配的系统复杂度。

借助大语言模型的链式推理与大词表能力，情感识别正从传统的判别式分类逐步演化为可解释的多模态情感推理：不仅要“判对标签”，更要“说清依据与过程”。与以往仅输出情感类别不同，可解释多模态情感识别旨在借助多模态大模型，用自然语言描述情感形成机制、触发线索与跨模态证据，以提升情感表征的可解释性与可靠性[316]。其中，开放词汇情感识别（OV-MER）[317]通过引入开放标签空间，突破固定类目限制，支持预测任意数量、任意类别的情感状态，更贴近真实互动场景下的细粒度与组合式情绪表达。另一项工作[318]从“数据—模型”双路径切入：构建面向生成式理解的 MER-Caption 数据集，以描述式标注丰富语义证据；同时提出 AffectGPT，通过预先融合策略增强多模态耦合，缓解跨模态对齐与证据整合难题，为可解释情感叙事提供端到端支撑。

总体来看，生成式范式将多模态情感识别任务重表述为文本生成，天然具备跨数据集/跨域迁移与任务指令化的灵活性：一方面，端到端流程简化了输入处理与系统集成；另一方面，由于推理路径更多依赖语言生成，模型对多模态依赖关系的显式建模与对齐细节的可控性相对不足，难以像传统模型那样精确刻画模态间的依赖关系[260]。因此，未来需要在参数高效适配（如更细粒度的 LoRA/Adapter 设计）、可控对齐（如显式跨模态注意与门控机制）、过程监督（如推理链/证据链校验）与不确定性量化（如置信度与校准）等方面细化：既保留生成式框架的通用与可扩展优势，又弥补其在多模态显式推理与可控融合上的短板，从而实现低成本、高解释且可迁移的多模态情感推理系统。

第3章 元宇宙技术平台

元宇宙技术平台是构建元宇宙的重要基础设施，通过数字孪生、数字原生、虚实融合以及虚实联动这四种形式将虚拟和真实进行空间立体的融合，是虚实两个世界信息的整合和相互增强。人和物在这个虚实融合的空间中，可以通过各种各样的互联和互动，以更高的效率、更低的成本、更身临其境的沉浸感和互动体验，解决日常生活、社会运行和产业发展中的诸多问题。元宇宙技术平台的开发还处于起步阶段，但已经吸引了众多科技巨头的关注。国内的联想、咪咕、元客方舟及国外的 Meta、英伟达、微软等公司纷纷投入巨资，在元宇宙共性技术平台、硬件基础设施平台、内容构建平台等布局研发，在工业、教育、娱乐、能源安全、办公、文化旅游等行业进行应用，初步取得了显著的经济效益和社会效益，展现了元宇宙技术在多个领域的强大市场应用潜力。

3.1 典型共性技术平台

3.1.1 超高清视音频技术

超高清视音频技术[319]是元宇宙的底层基础技术之一，元宇宙应用离不开超高清视音频技术的支撑，这既给超高清视音频技术带来新的挑战，也将推动超高清视音频技术发展进入新的阶段。从零售、泛娱乐、社交、教育、政务、工业等元宇宙典型场景入手分析，为满足视觉、听觉以及触觉方面的体验要求，超高清视音频技术在沉浸式视听体验、实时大规模用户多维互动、高效内容生产等方面都有新的需求：

- 沉浸式视听体验：传统二维视音频升级势在必行。从内容生产、高效编码，到沉浸式影像的传输与呈现等方面技术演进，实现虚拟与现实融合的沉浸式体验；
- 实时多维互动：元宇宙的场景中大规模的用户可同时在线并进行实时互动，这对视音频的处理、通信、呈现等实时性、兼容性有较高的要求；
- 高效内容生产：元宇宙时空构建，需要大量的内容，这就需要提升现有的内容生产效率。

超高清视音频技术发展需要在单点技术的发展基础上，融合多种关联技术，形成针对横向技术栈的端到端解决方案，实现元宇宙时代视音频技术的有效连接和共同发展。2023 年第十届中国网络视听大会发布了面向元宇宙时代的超高清视音频技术体系，基

于元宇宙时代的技术需求面向超高清的端到端流程，详细分析了包括内容生成、内容编码、网络传输、内容解码、交互与呈现、用户体验及版权保护各个模块面向元宇宙时代关键技术升级和前沿研究，研究探讨视音频关键技术簇，对于提升元宇宙的视听体验，具有重要的技术支撑作用。

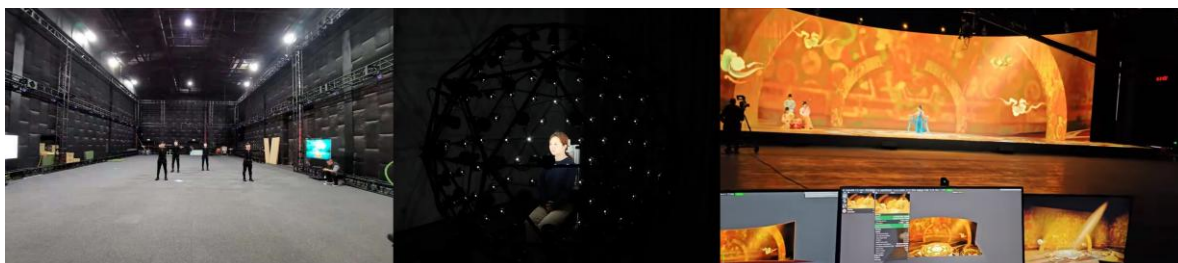


图 3-1 MAC 光场共性技术平台（数字人光场采集系统，动捕系统，XR 虚拟拍摄系统）

在超高清音视频技术方向，一项典型的技术成果就是元客方舟和咪咕文化共同建设的首个光场共性技术平台 MAC[320]，该平台主要为了给科幻、元宇宙等产业提供共性技术服务。针对光场技术门槛高，国内光场系统数据质量差、计算效率低、成熟应用少等问题，该平台采用梯度光场成像、多孔径光场成像及高精度高效 AI 算法技术，开发建设了系列光场采集建模系统、光场数据库、数字人制播应用与评测系统、全息通信应用评测系统，如图 3-1 所示。可以将数字内容制作周期从数月缩短至 2 天以内，将帮助数字内容、全息视频通信等新型元宇宙应用产业发展。

基于“视觉+AI”技术，该平台还打造了光场采集、动作捕捉、4D 全景采集棚、XR 影棚、元宇宙数字人直播基地、数字资产存证确权、数字资产流通等技术服务解决方案[321]，面向元宇宙领域 AR/XR/MR 等虚拟内容、虚实融合内容的智能制作，可以为企业提供优惠的共性技术服务。其中 MR 空间视频重建系统，通过多视点相机和高解析度 HDR 六色 LED 光源阵列采集，实现照片级写实的人物动态重建，千元内可完成 30s 空间短视频制作，制作效率提升 10 倍；此外，还重点打造了 4DLuStage 数字人光场体积视频采集重建系统，参与构建清华、商汤业内头部多视点数据集，基于这些数据集今年已涌现出了大量 Gaussian Splatting 技术相关的顶会论文，助力开创学术前沿，提升我国元宇宙内容产业的竞争力。中央广播电视总台推出的纪念中国人民抗日战争暨世界反法西斯战争胜利 80 周年特别节目《烽火战歌》创新性地融合了 LuStage 体积视频技术，仅 4 天完成数字“聂耳”高精度还原，让抗战精神在科技中重生，为影视制作开启高效智能新时代。

中国移动与咪咕文化提出了面向元宇宙的视频业务要保证用户实时沉浸、无卡顿流

畅的交互体验的需求,构建分层级获取算力,逐层处理,分级服务的网络算力解决方案。在边缘侧通过建设多样化算力、网络内生算力支撑不同场景下的视频算力需求,如与基站结合构建接入侧轻量视频算力提供传输加速、加大上行带宽等能力,与 UPF 结合建设边缘网关视频算力,满足园区级算力需求;在云端构建大规模的视频渲染及视频分析集群算力、高压压缩分布式存储,为高算力类视频业务提供共享算力服务。元宇宙场景下的视频应用有效利用算力网络新底座能力及增强型视频服务能力,极大降低传输时延,支撑大带宽元宇宙的视频类应用,打造沉浸式用户体验。

3.1.2 云渲染技术

云渲染是元宇宙三维虚拟世界高清可视化呈现输出的关键技术,通过云计算提供高清、低延时的视觉输出,克服了本地设备的性能和资源限制,为构建三维数字世界提供了强大支持。云渲染技术的核心在于其混合渲染能力、端云异步渲染技术,以及特别为 AR 和 VR 设备优化的低延时控制技术。通过 AIGC 重建模型、CAD 模型到数字人的高清混合渲染,不仅提高了渲染的精确度和真实感,还极大地提升了渲染效率和系统响应速度。端云协作的异步渲染技术进一步降低了 Motion-To-Photon (MTP) 延迟,使得用户在使用 XR 设备时能够获得即时且连贯的视觉体验,这对于实时互动和虚拟体验的流畅性至关重要。可以预见到,通过不断优化的渲染技术和更深入的硬件集成,将能够支持更多的终端设备,包括 PC、平板及各类 AR 和 VR 设备,使得数字世界的每一个角落都生动而真实。

(1) 隐式模型与 CAD 模型的混合渲染

云混合渲染的研发目标是通过三维模型处理和渲染管线优化,实现对三维物体/场景的低延迟、高质量的渲染,支持 XR 设备同步实时渲染,端云协同渲染,混合渲染,支持多终端、多平台访问。主要解决三个重要技术问题:一是隐式三维模型和传统模型的混合渲染,采用深度预测等技术,实现物体遮挡等效果。二是端云协作渲染,主要解决三维复杂大场景的实时渲染问题,重点解决端云渲染延时。三是 XR 设备低延时渲染,主要解决 XR 类头戴式设备接入渲染画面的显示问题,将结合联想自研 XR 设备,进行端到端优化。

(2) 端云异步渲染

为了在通过 ARVR 等头戴式设备接入元宇宙系统时虚拟画面锚定,即 MTP 在 20ms 内,需要设计将云渲染、编码推流、终端刷新显示过程从串行转为并行处理,云端依据终端上报的位姿信息进行渲染,并传输到帧缓冲区,终端从帧缓冲区中取帧进行二次渲染和屏上显示。这样终端显示刷新不再依赖于云端渲染,从而实现 MTP 小于 20ms 目标。端云异步渲染需要经历了两次渲染过程,第一次是云端根据 AR 设备反馈的位姿信息进行了画面渲染,然后传输到 AR 设备侧缓存;AR 设备本地,从缓存帧中选择一帧,结合当前用户实时位姿,运用 ATW、运动预测等算法,对当前帧进行二次渲染图像矫正,以增强模型的锚定性,解决果冻效应等问题,最终在屏上显示。

(3) 低延时 MTP

MTP 只取决于本地渲染 MTP,本地渲染 MTP 也需要根据 AR 设备显示等条件进一步优化设计。在 Android 系统下,直接利用显示模块,MTP 将超过 50ms,即使采用 ATW (Asynchronous Time Warp):异步时间扭曲技术,这个时间延迟还是超过 36ms,因此,我们设计了一种单显示帧策略,即把原来的双显示 buffer 缩减为单 buffer 处理,这样,可以将 MTP 缩减到 30ms 以下,如图 3-2 所示。为了继续缩减这个时间,需要运用运动预测技术,通过预测头部的转动,来预测显示的时刻,这样,可以进一步将 MTP 缩减到 20ms 以下,如图 3-3 所示。另外,由于 AR LED 显示是逐行刷新,由上向下逐行显示,最上面一行的显示时刻与最下面一行的显示时刻存在差异,导致显示画面变形(Rolling Shutter Effect),我们对每行的显示也进行了位姿的补偿,从而使显示正常。

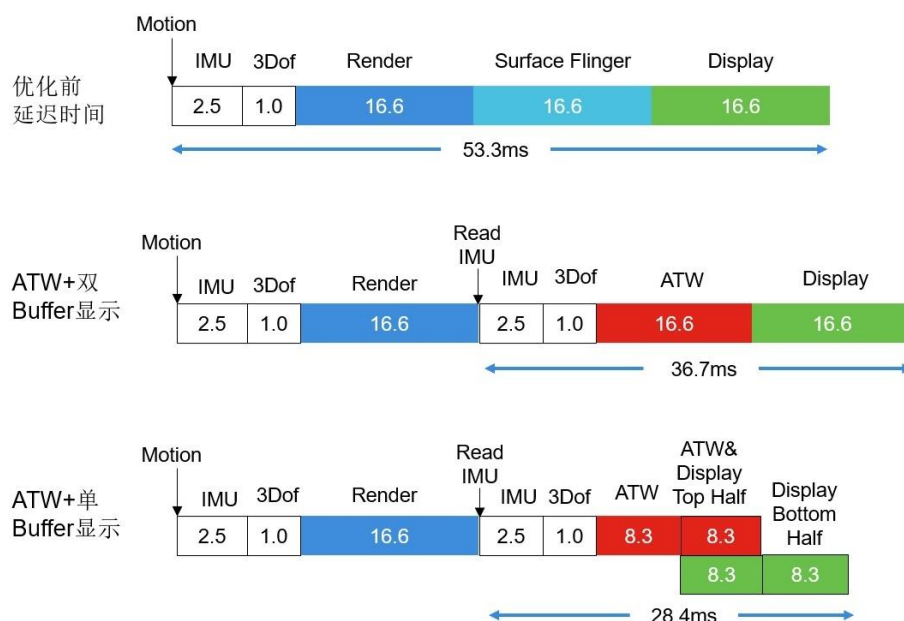


图 3-2 优化显示机制和流程, 缩减 MTP 到 28.4ms

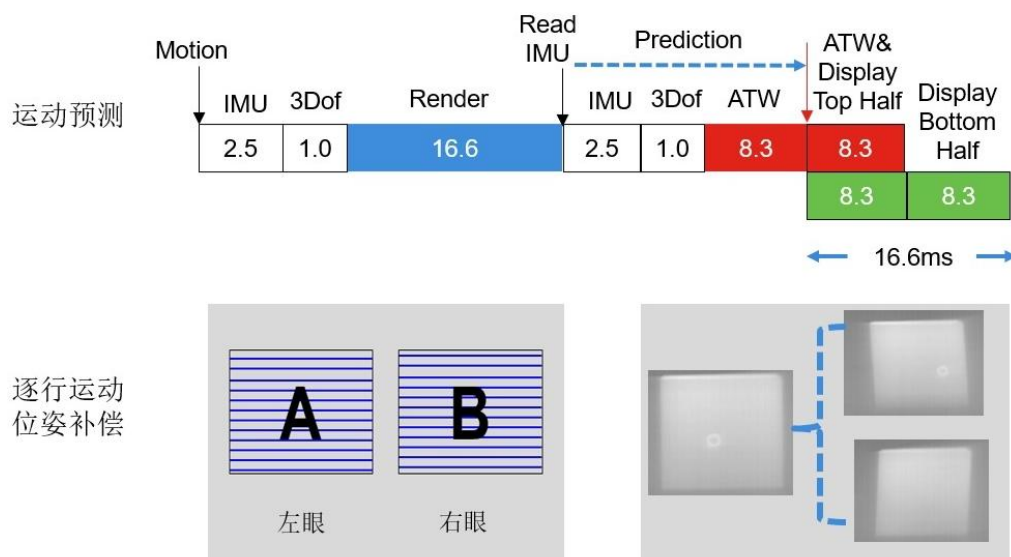


图 3-3 用运动预测和逐行位姿补偿技术，缩减 MTP 到<20ms

(4) 双目交替渲染技术

为了解决 3D 大模型渲染帧在实时传输中较大的带宽占用，设计采用双目交替渲染技术。对于每个渲染周期，云端交替渲染左眼或右眼数据并传输到终端，终端借助上一周期本地缓存的数据和最新拿到的数据进行双目的刷新显示。为了补偿左右眼因间隔一个刷新周期导致的撕裂感，设计采用了双 ATW 矩阵分别应用于左眼和右眼的显示刷新，以达到良好的视觉体验，如图 3-4 所示。由此，在 3D 大模型渲染场景下节省了约 50% 的带宽占用、削减了峰值带宽。

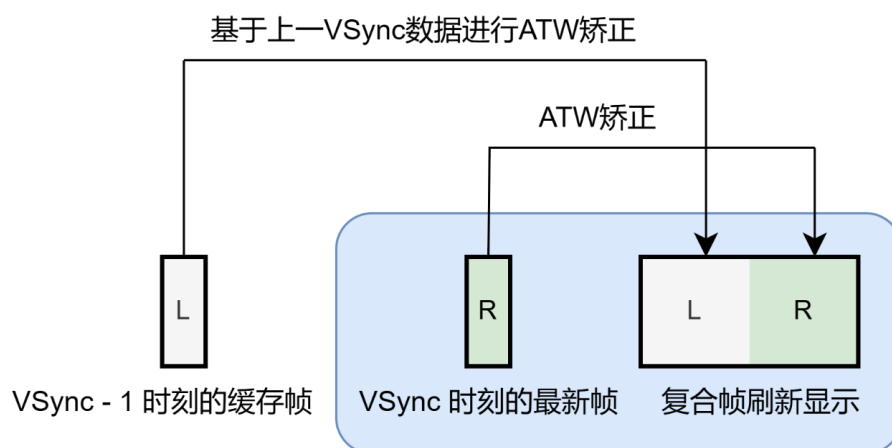


图 3-4 双 ATW 矩阵矫正

3.2 典型元宇宙硬件平台

(1) 虚拟现实（VR）设备

虚拟现实（VR: virtual reality）是一种通过计算机技术，模拟出一套逼真的、交互式的虚拟环境，让用户有一种沉浸在虚拟世界中的感官体验。虚拟现实技术通常使用立体显示器、控制器、传感器和计算设备等硬件设备，以及相应的虚拟现实软件来实现。用户观看立体显示器，眼前的屏幕会显示计算机生成的三维虚拟场景。用户可以通过控制器等交互设备来操作虚拟世界中的对象、进行动作、探索环境等，同时传感器会实时追踪用户在现实世界中的动作和位置，将其反馈到虚拟环境中，从而实现用户在虚拟世界中的自由移动和互动。虚拟现实设备是元宇宙系统的重要入口与交互设备。

虚拟现实设备经历了从实验室阶段（20 世纪 50~70 年代）、产品化实现阶段（20 世纪 70 年代~90 年代）、完善并快速发展阶段（1990 年至今）的发展；2016 年，Oculus Rift 和 HTC Vive 正式发布后，掀起了虚拟现实设备新的浪潮。2017 年谷歌 VR 设备一体机与联想集团合作发布了 Daydream VR 设备一体机，目标是统一基于 Android 系统的 VR 设备硬件参考设计，并搭建完整的 VR 应用生态。2019 年 Oculus Quest 发布，成为首款无线自主 VR 头显，以及联想集团与 Oculus 合作，发布 Rift S VR 设备产品，为虚拟现实的普及带来了重要推动力。

2020 年之后，VR 设备产品跟之前相比，在重量、成像质量、交互和佩戴舒适性等方面都有了长足的进步，一些新品牌的 VR 设备产品也逐渐在市场上被消费者看到。同时脸书公司为了进一步推广并宣传 VR 设备产品，把公司名称也更换成 Meta，寓意更加专注 VR 作为进入元宇宙的专用设备。

2023 年，苹果发布了高端 VR 设备一体机 Vision Pro，能够非常好地支持交互和混合现实（mixed reality, MR）功能，同时搭配苹果的 iOS 生态，已经成为 VR 行业的新的标杆。同时，各家品牌公司也都在不断推陈出新各自的新产品，而且成本更亲民、功能更强、重量更轻 VR 设备一体机，比如联想推出的 ThinkReality 子品牌 VR 设备一体机，为行业用户提供高品质的 VR 设备，而字节跳动旗下的 PICO 推出 PICO4 VR 设备一体机，成为国内最大的 VR 设备一体机品牌。

近年来，VR 头显设计越来越强调轻量化、紧凑化、佩戴舒适度，通过使用微型 OLED (micro-OLED) / 超短焦光学折叠光路（简称：Pancake）光学模组 / 显示与光学结构优化，让头显厚度、重量和体积大幅下降。显示分辨率、刷新率不断提升，部分 VR 头显的分辨率从 4K 向 8K 迈进，刷新率也提升到 120Hz 甚至更高。

目前，市面上比较主流 VR 设备有 Meta 公司的 Quest 系列 VR 设备、HTC 的 VIVE 系列、联想 ThinkReality VRX 和 Rift S 系列、苹果 Vision pro 以及 PICO VR 设备系列

等，如图 3-5 所示。



图 3-5 VR 设备（联想 ThinkReality VRX，联想 Rift S，Pico 4，Vision Pro）

（2）增强现实（AR）设备—空间计算 AR 头显

虚拟现实技术为了追求沉浸感而把用户和真实的世界隔离，用户无法观察到真实的世界。而增强现实（AR）则是一种新兴的计算机应用和人机交互技术，利用传感器、摄像头将真实的环境和虚拟的物体对象实时地叠加到同一个画面或者空间中，使用户看到一个重叠的世界。因此，增强现实是利用现实的信息技术对真实世界的一个补充，而不是单纯地创造一个虚拟的世界，增强现实更多地强调真实性和交互性。增强现实技术近年来成为国内外机构和大厂的开发、研究的热点，并广泛地应用到各个领域，有可能取代智能手机，成为下一代智能计算与显示平台，将物理世界与信息世界联系得更加紧密。

为了区别于近两年快速发展的以信息提示为主的轻量级 AI 智能眼镜，可以将以微软 Hololens AR 头显为典型代表（如图 3-6）的增强现实设备称为空间计算 AR 头显，其显著特点是视场角大、算力强、功能体验以在现实世界叠加锁定三维内容，但体积更大，不太适合长期佩戴。光学与显示以衍射光波导、高亮 Micro-LED 投影等为主，具备较大的视场角和稳定的全彩 3D 叠加能力，配备深度摄像头、环境摄像头、IMU、手势传感器、虹膜识别等，具备完整 SLAM、空间网格重建、手势跟踪与空间交互能力，可实现空间锚定、全息内容固定等完整混合现实体验。主要应用于工业、制造、医疗培训、远程维护、教育演示等高价值专业场景。

这类设备除了微软 Hololens2 外，典型代表设备还有 Magic Leap 1&2、联想晨星 AR 设备等，如图 3-6 所示。未来，这类设备重点在于进一步降低重量与功耗、扩大视场角与自然交互能力，特别是续航时间、佩戴舒适性等方面。



图 3-6 典型 AR 设备（微软 Hololens2，联想晨星 New G2 pro，联想晨星 Tyche）

（3）AI 智能眼镜—轻量化 AI 眼镜

CES2025 将 AI 智能眼镜推至聚光灯下，成为全场焦点。众多终端厂商及上下游产业链参与者纷纷亮剑，带来形态各异、功能丰富的产品。AI 眼镜产业正处于高速迭代期，其核心驱动力来自多模态大模型、轻量级算力芯片、微光学显示技术以及低功耗感知模組的持续突破。产业形态从早期的通知提醒类可穿戴设备，逐步演进到具备实时视觉理解、语言交互、场景搜索、物体识别与多模态信息增强的“轻量级具身智能终端”。在供应链方面，整机品牌（如 Meta、Google、联想、Xreal、Rokid、雷鸟）正与上游光学模组、AI 芯片和语音/视觉模型供应商形成紧密协同。应用场景也由娱乐拍摄扩展到生活助理、运动导航、远程协作、无障碍辅助、实时翻译、安防巡检与企业数字化等方向。随着头戴式设备在佩戴舒适度、续航、散热、算力本地化等问题持续改善，AI 智能眼镜正加速从小众科技产品向大众消费电子和行业级生产力工具转变，预计未来 3 - 5 年将成为最有潜力的 AI 原生硬件品类之一。图 3-7 所示为 Meta、联想及 Rokid 等研发的 AI 智能眼镜产品。



图 3-7 AI 智能眼镜（从左至右：Ray-ban, Lenovo, Rokid）

AI 智能眼镜近三年的研发呈现“轻量化硬件 + 多模态感知 + 云端/本地大模型推理 + 场景化应用”四条主线并行推进。硬件方面，设备从早期的录播类眼镜演进到具备显示、AI 处理与全天候佩戴能力的综合形态，重量由百克级持续下探到 40 - 80 克区间，电池设计采用分体结构、镜腿电池与高能效 SoC 组合，确保在主流使用场景下支持半天至全天续航。光学形态从单目 HUD、双目 Birdbath 小光机逐步扩展到轻量波导方案，FOV 虽仍有限，但亮度、清晰度及户外可视性持续提升，满足日常提示、翻译与导航等轻交互诉求。传感器配置趋于多样化，包括广角摄像头、IMU、阵列麦克风、环境光传感器等，为 AI 视觉识别、语音输入、运动检测与生活记录提供基础数据。算力方面，眼镜本体的 NPU、DSP 与低功耗 CPU 已能支持基础的语音唤醒、离线命令、物体识别和简单图像处理，但更复杂的推理、实时视觉问答、场景总结与多模态理解则通过手机端或云端大模型完成，因此“眼镜端轻推理，结合手机端中推理，以及云端重

推理”的混合架构成为主流。随着 LLM 和视觉-语言模型的快速发展，AI 智能眼镜的能力从“记录与展示”跃迁为“实时理解与辅助”：自动总结会议、识别物体与文本、提供多语言翻译、根据视野内容主动给出提示等均进入可用阶段，使眼镜从信息终端转变为面向工作与生活的“第一视角 AI 助手”。在交互上，语音成为主要入口，辅以轻量触控、手势识别与空间音频，提高移动场景下的稳定性。产业生态层面，消费级市场快速扩大，功能从拍摄与直播扩展至个人助理、健康运动、出行导航与场景级应用；企业侧则将其用于巡检、培训、客服辅助等，强调实时信息查询与流程规范化。整体看，AI 智能眼镜正在从“功能小工具”走向“日常可连续佩戴的智能终端”，其核心推动力来自光学轻量化、电源系统优化、多模态大模型的实用化以及端云协同体系的成熟。接下来，行业将集中在提升全天候佩戴舒适度、增强端侧推理能力、降低能耗、加强隐私与安全机制，以及扩展更具刚需的日常场景，以推动 AI 智能眼镜从先锋用户向大众普及。

（4）机器人设备

机器人+“元宇宙”的想法是进一步扩大人类影响力，为移动性和机器人技术在元宇宙中找到一个角色。通过将机器人链接到元宇宙，人们将实现在现实世界和虚拟现实之间的自由移动。相比元宇宙提供的身临其境的现场感，“机器人+元宇宙”的体验将更进一步，机器人将成为身体感官的延伸，让人们能够通过“机器人+元宇宙”重塑和丰富日常生活。

元宇宙中的虚拟数字世界，为机器人的学习和模拟运行提供了极好的平台环境。机器人可以在元宇宙提供的丰富的虚拟数字化场景中，进行大量的学习，以及业务的模拟仿真，不断地迭代，从而机器人变得更加的“聪明”，更优化地执行业务。如果说 AR/VR 设备是元宇宙虚实融合和虚实交互的入口，那么机器人则是元宇宙虚实联动的载体。机器人系统实现了虚拟世界与真实世界的联动，虚拟机器人可与真实机器人集成并连接。在现实世界中，物理机器人可以同步接收和执行虚拟机器人的虚拟命令，机器人在物理世界中的状态也将实时显示在虚拟世界中。图 3-88 所示为联想上海研究院及杭州云深处研发的足式及轮足机器人系列产品，并结合晨星工业元宇宙平台，并应用于元宇宙园区智能巡检、国家电网变电站智能巡检作业、元宇宙工厂机器人作业等场景。



图 3-8 足式/轮足机器人（从左至右：联想 GS、联想 MX、云深处 X30）

人形机器人的发展将进一步地助推元宇宙的发展，将在元宇宙中成为连接虚拟与现实的关键载体：它们以拟人的感知、运动与交互能力，将虚拟空间中的智能体“具身化”，让数字角色拥有真实的行动、表情与行为逻辑；它们也能作为用户的延伸，在虚拟世界中进行协作、创作、探索与沉浸式体验；同时，人形机器人可把元宇宙中的内容、场景和服务反射到现实场所，实现跨界的工作流与体验流通。通过这一虚实共生体系，人形机器人将重塑元宇宙的交互方式、产业形态与价值边界。总之，人形机器人的发展不仅丰富了元宇宙的内容和互动方式，还为用户提供了更加智能和个性化的服务，极大地提升了元宇宙的整体价值和吸引力。图 3-9 所示为宇树科技、智元和联想研发的人形机器人产品。

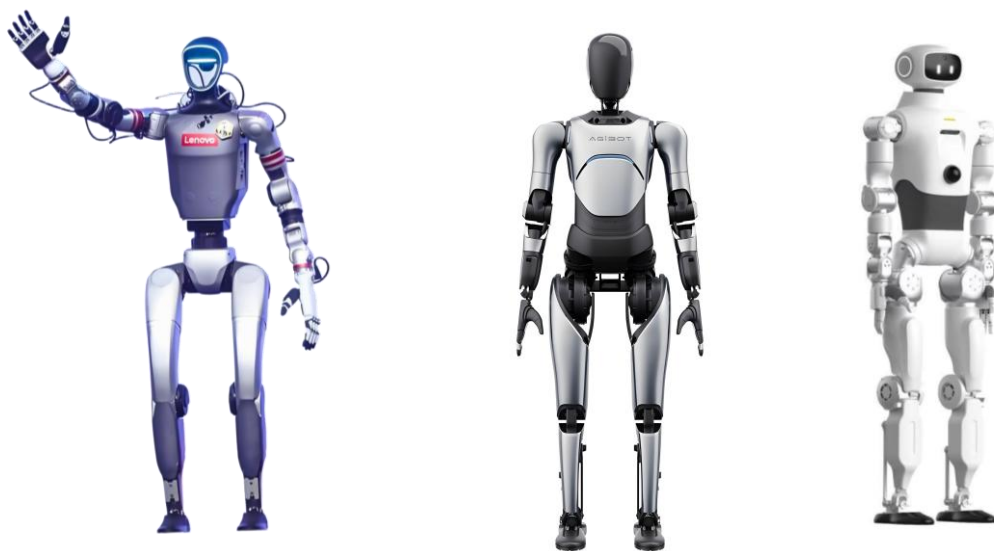


图 3-9 人形机器人（从左至右：宇树 G1、智元远征 A2、联想 LX）

3.3 典型元宇宙内容构建平台

（1）联想晨星工业元宇宙平台 Daystar World

晨星工业元宇宙平台（Daystar World）是由联想开发的面向行业元宇宙解决方案应用的技术服务平台，2022 年首次发布，是基于计算机视觉、空间计算、人工智能、AIGC 等核心技术多核心能力聚合的国内领先工业元宇宙开放平台。如图 3-10 所示，其核心能力包括三维重建、多源融合三维数字地图、数字孪生、XR 内容制作、机器人模拟仿真、AI 视觉检测、实时云渲染等，并提供系列零代码业务构建工具和业务运营工具，支持具身智能机器人、XR、智能相机等设备的连接，实现虚实融合、虚实叠加及联动，帮助客户快速构建和部署元宇宙应用方案。



图 3-10 晨星工业元宇宙平台框图

平台核心功能特征有：1）基于 3D 高斯泼溅（3DGS）技术的高拟真三维重建，实现低成本、快速物理环境的数字化重建；2）零代码元宇宙业务构建，通过可视化拖拽操作及丰富的行业应用模板，显著降低元宇宙方案构建成本及实施周期；3）虚实联动，平台接入了多类智能机器人、相机设备等，实现物理世界与虚拟世界的实时联动与控制。4）全三维体验，高拟真三维孪生空间，结合三维表达的 IOT 设备信息、业务运营告警信息、结果统计等信息，通过平台内置的区域超链接、地图兴趣点等高效交互手段，实现在虚拟空间中的高拟真可视化显示及沉浸感漫游体验。

平台核心技术包括：1）空间计算技术，基于点云数据、倾斜摄影数据、360 视频数据，融合 NeRF 和三维视觉算法的空间重建，构建了“元”空间，面向机器人、AR 设备、无人机等设备，提供全要素（特征/Mesh/纹理/语义）空间计算服务，由此构建了数字底板，形成元宇宙构建平台基础；2）基于视觉大模型的识别检测技术，借助视觉大模型、

正负样本生成、few shot、无监督学习等技术，检测“人”和“物”的行为和异常；3）实时云渲染技术，实现了对三维模型的处理及自动加载（包括基于隐式模型表达的 NeRF 模型），支持隐式模型和 CAD 模型的混合渲染，云端实时高清高精度渲染和低延时推流；4）孪生智能体模拟仿真技术，可构建物理设备的孪生智能体，实现物理仿真、IOT 数据驱动、虚实映射、虚拟部署、仿真运行及反向控制。

平台具备丰富的工具链，如孪生体制作工具、内容连接工具、模型处理与转换工具、机器人仿真部署工具、AR 内容编辑工具、摄像头检测部署工具、数字人编辑工具、多用户互动工具、虚拟展厅制作工具、远程协作工具等。此外，平台具备 SDK 和开放接口 API，开发者可利用其能力开发行业应用工具，形成元宇宙应用生态圈。实现的 IOT 数据统一接入标准，可实现与外部工业设备、AI 平台等的互联对接。

AIGC 技术在平台中广泛应用，可以大幅降低元宇宙构建成本，提升交互体验和智能化应用价值。基于平台开发的教育培训、展览展示、智慧园区、智慧工厂、数字人客服等通用行业元宇宙方案，已经成功应用于国家电网、吉利汽车、上海自然博物馆、无锡商院、中国商飞等标杆客户。通过虚实融合和虚拟模拟验证，帮助企业实现降本增效。针对智能机器人智能作业业务部署复杂、周期长、专业要求高等挑战性问题，机器人业务虚拟部署和模拟仿真系统显得尤为重要，如图 3-11 所示。运用 AIGC 技术，可快速高拟真数字化重建目标环境，并在虚拟数字化空间中部署机器人业务，以及进行机器人业务模拟仿真，然后将优化后的机器人部署方案部署到物理机器人及物理环境。也即，将机器狗方案的部署从基于“物理现场调试”的模式，转变为基于“数据和模拟”的 Sim2real 模式，可极大地提升了机器人业务部署效率，大幅降低了部署调试成本。



图 3-11 机器人虚拟部署与模拟仿真验证

(2) 元宇宙沉浸内容生产及运行基座“咪咕筑境&咪咕绘境”

元宇宙沉浸内容生产及运行基座“咪咕筑境&咪咕绘境”提供一站式元宇宙“人、物、场”沉浸数字内容智作方式，对比传统数智场景制作模式，元宇宙场景构建效率最高提升 300%，大幅降低 UGC（用户生成内容）门槛，重新定义沉浸内容生产方式。

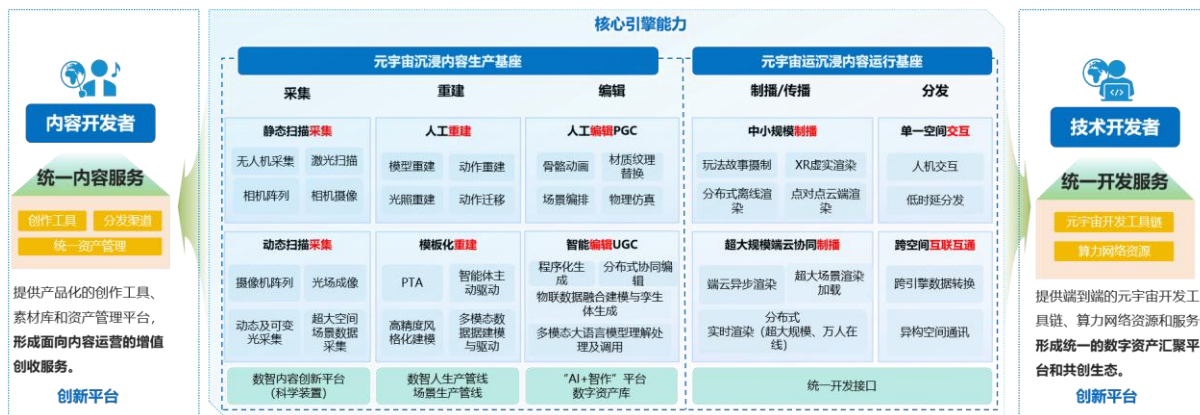


图 3-12 元宇宙沉浸内容生产及运行基座架构图

元宇宙内容生产基座“咪咕筑境”包含光场数据采集、数智人生产管线、场景生产管线、AI+智作平台、数字资产库等功能，整体面向行业内容创作者提供从采集、数字内容重建、资产管理等全流程能力支持。其中，数智内容创新平台（科学装置）可针对“人、物、场”要素进行高精度、多维度的光场信息采集；数智人生产管线涵盖模型拓扑、材质纹理、骨骼绑定和动画驱动等制作全流程，支持 L1-L5 多种精度数字人标准化

高效生产，兼容语音、文本、面部 BlendShapes 等多主流数字人驱动方式；场景生产管线结合 Nerf、3D/4D 高斯喷溅等技术，支持室内外沉浸数字空间场景快速重建，重建效率达 3 小时/万平；“AI+智作”平台内置文生图、图生视频、文生视频等 AIGC 能力，打通“剧本编制-人物创建-场景设计-动画制作-音效配乐-后期剪辑”影音数字作品创作流程；数字资产库对基座上的数字资产集中化管理，支持进一步二创及分发。

元宇宙内容运行基座“咪咕绘境”依托中国移动 5G/5G-A+算力网络能力，以分布式实时云渲染、端云异步渲染技术为核心，融合算网服务能力，可以支撑超大规模场景、超高清模型、超逼真效果的实时交互渲染。对亿级渲染的超大场景，可实现 4K 60 帧的细腻渲染、端到端时延小于 200ms。此外，运行基座建设统一开放接口，可针对数智人、超大场景的 4K/8K 超高清数字内容制作提供即时接入、实时渲染、传输分发等能力支持，同时汇聚多种元宇宙能力及产品服务，满足各类场景运行需求。

依托元宇宙沉浸内容生产及运行基座，进一步构建了“三纵一横”（数智人、数智空间、数智内容及云制作渲染）元宇宙产品服务矩阵。其中，针对不同个人、企业、政府等用户需求，可提供 L1-L5 多种精度的个人、行业、IP 数智人产品；针对不同场地功能需求，可提供景区、社区、会议室等数字空间产品；针对裸眼 3D、VR、AR 等终端设备，可提供配套的沉浸内容产品；针对政府端、企业端沉浸内容需求，可提供定制服务及高效云渲染服务。



图 3-13 “三纵一横”元宇宙产品服务矩阵

目前，针对文博、文旅、文体、文娱、文商等行业场景数字化升级，基座已为超 20 个省/直辖市的政企事业单位提供元宇宙相关服务，面向“元宇宙+景区云游”“元宇宙

+街区漫游”“元宇宙+展览展示”“元宇宙+开幕式”等元宇宙新业态新场景打造了“鼓浪屿元宇宙”“云游·大足石刻”“张家界星球”逾 40 个典型案例。

（3）英伟达 Omniverse 平台

较早布局元宇宙算力技术的英伟达于 2021 年发布了 Omniverse 元宇宙平台，用于虚拟协作与实时物理模拟。Omniverse 是一个开放的、可扩展的 3D 设计协作和仿真平台，旨在增强创作者、设计师和工程师之间的协作能力。该平台通过统一的实时虚拟协作环境，使用户能够在不同工具和软件之间无缝共享和编辑 3D 内容，从而提升工作效率和创意表达。

通过支持多种行业标准和格式（如 USD、MDL 等），Omniverse 为设计和工程 workflow 提供了一个通用的基础，使来自不同专业和背景的用户能够轻松合作。图 3-14 为 Omniverse 平台技术架构，其核心技术及功能模块有：1) USD(通用场景描述)，Omniverse 采用 Pixar 开发的 USD 格式，作为 3D 数据的核心交换格式。USD 提供了高效、可扩展和可互操作的场景描述，支持复杂的场景和资产管理。2) RTX 实时光线追踪，利用 NVIDIA RTX 技术，Omniverse 能够提供高质量的实时光线追踪渲染，使用户在创作过程中实时预览高保真的视觉效果。3) AI 和深度学习，其集成了 NVIDIA 的 AI 技术，用于自动化任务、增强协作和提高生产力，如智能内容生成、物理仿真和自动化 workflow 等。4) Omniverse Nucleus，这是平台的核心协作引擎，负责管理和同步所有用户的修改和更新，确保实时协作的流畅性和一致性。

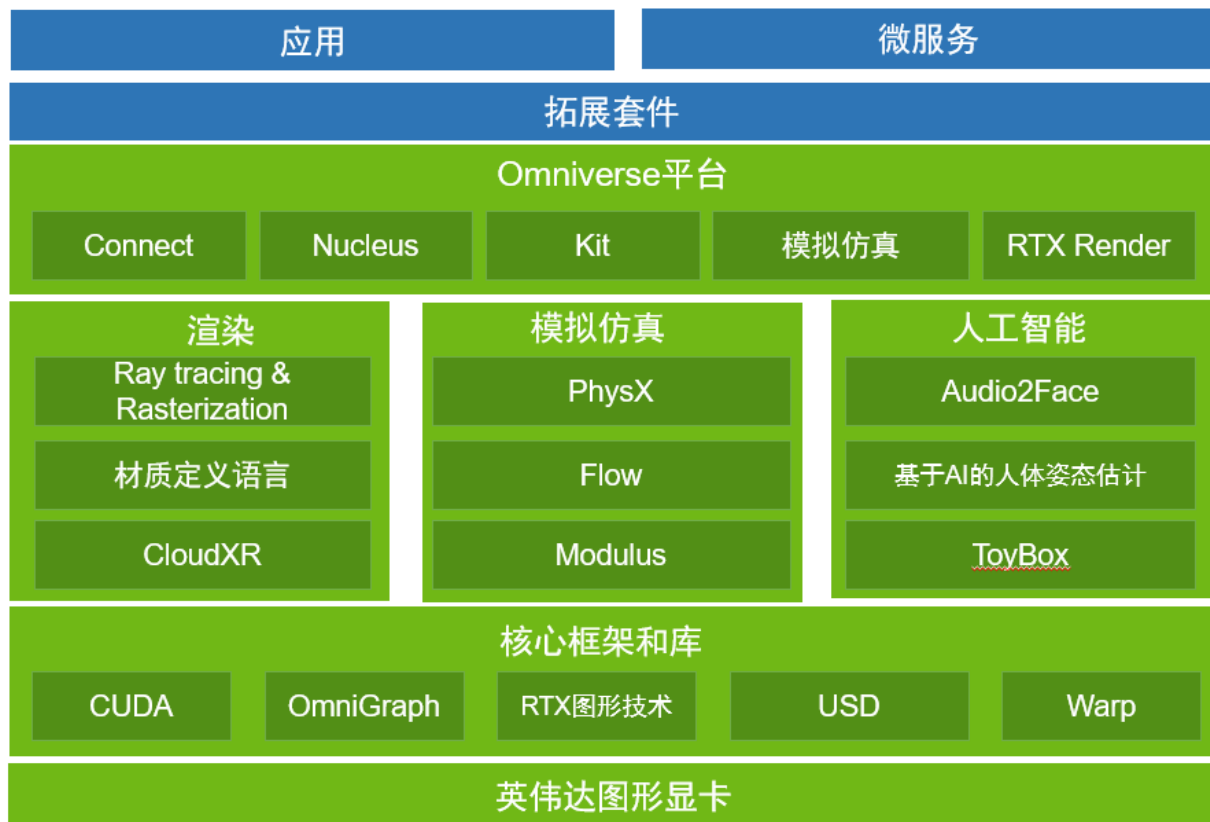


图 3-14 Omniverse 技术框架图

在 SIGGRAPH 2025 大会上，NVIDIA 发布了一系列重磅更新，包括面向大规模场景重建的 NVIDIA Omniverse NuRec 3D Gaussian Splatting (3DGS) 库、面向视觉 AI 的 NVIDIA Metropolis 平台升级，以及全新的 NVIDIA Cosmos 与 NVIDIA Nemotron 推理模型。其中，3DGS 能力的引入尤为值得关注，它为 Omniverse 平台带来了高效率、高保真的三维场景重建能力，使用户能够基于真实世界数据快速构建符合物理规律的 3D 环境，并实现实时渲染和交互。借助这项技术，数字孪生空间的创建不再依赖昂贵的建模流程，而能够以更低成本、大规模地实现高拟真场景的自动化生成，大幅降低了构建高拟真元宇宙场景的成本。

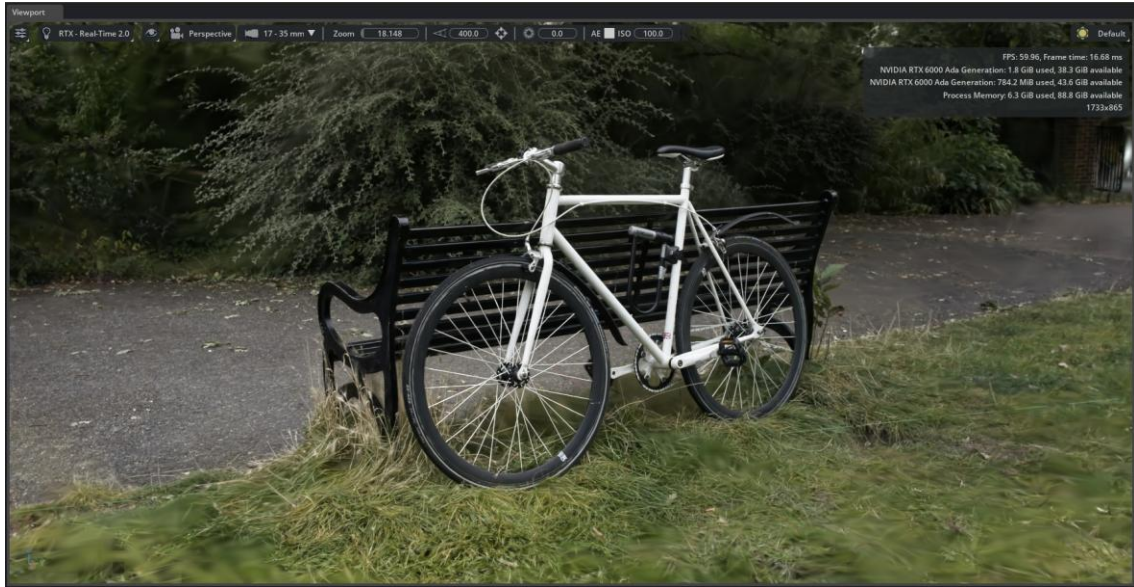


图 3-15 Omniverse 3DGS 生成和渲染高拟真的 3D 场景表达

核心功能特征有：1）跨平台协作：支持多种 3D 设计工具（如 Autodesk Maya、3ds Max、Blender、Adobe Photoshop 等）的无缝集成，允许用户在其熟悉的环境中工作。2）实时同步：所有参与者的修改和更新能够即时反映在共享的虚拟场景中，避免了传统的文件传输和版本管理问题。3）高保真渲染：通过 RTX 技术，用户能够在创作过程中实时看到接近最终效果的渲染结果，大大缩短了设计迭代周期。4）可扩展性：开发者可以通过 Omniverse Kit 和 Connect 扩展平台功能，定制符合特定需求的工作流和工具。

Omniverse 平台在多个行业中取得了显著的成功应用，展示了其强大的协作和实时仿真能力。宝马集团使用 Omniverse 进行工厂的数字化规划和设计。通过创建虚拟工厂，宝马能够在虚拟环境中模拟和优化生产线布局、设备放置和工人操作流程。这不仅提高了规划效率，还减少了物理原型的需求，节省了时间和成本。Bentley Systems 是一家基础设施工程软件公司，使用 Omniverse 来提高其工程和设计团队的协作效率。通过在 Omniverse 中创建复杂的基础设施项目模型，Bentley 的团队能够实时进行设计审查和调整，确保项目的精确性和可行性。这些成功应用案例展示了 Omniverse 平台在不同领域的强大功能和广泛适用性。通过提供一个开放、协作和实时的虚拟环境，Omniverse 正在改变各个行业的工作方式，推动创新和效率的提升。

（4）Meta Horizon World 平台

Horizon World 是 Meta 推出的一个虚拟现实社交平台，旨在为用户提供一个沉浸式的虚拟世界，在这里用户可以通过自定义的虚拟形象进行互动、创作和探索。它不仅是

一个社交平台，更是一个用户生成内容的生态系统，鼓励用户创建自己的虚拟环境和游戏。

Horizon Worlds 依托于 Meta 的 Oculus VR 硬件设备，提供高度沉浸的 3D 虚拟体验，使用户能够通过视觉、听觉和触觉全面感受虚拟世界。利用 Meta 的社交网络技术，允许用户通过语音聊天和手势互动进行实时交流，增强了虚拟环境中的社交体验。Horizon Worlds 提供强大的创作工具，用户可以利用这些工具设计和构建自己的虚拟场景、游戏和物品，推动平台内容的多样化和丰富性。虽然 Horizon Worlds 主要面向 VR 用户，但也支持与非 VR 设备的互操作，使更多用户能够参与到虚拟世界中。

Horizon 平台的核心功能特征包括：1) Avatar 自定义。用户可以创建和定制自己的虚拟形象 (Avatar)，从外观到服装，极大地增强了个性化体验。2) 虚拟世界探索。平台内有各种预设的虚拟环境和用户生成的内容，用户可以自由探索不同的世界，参与各类活动和游戏。3) 创作工具。Horizon Worlds 提供简单易用的创作工具，用户无需编程技能即可设计和构建虚拟物品和场景。工具包括模块化构建、脚本编写和物理模拟等。4) 社交空间。平台设有多种社交空间，如聚会场所、会议室和娱乐区，用户可以在这些空间内举办虚拟活动、会议和聚会。5) 经济系统。Meta 计划在 Horizon Worlds 中引入虚拟经济系统，让用户通过创作和交易虚拟物品获利。

Meta Horizon Worlds 平台主要应用于教育培训、虚拟社交、游戏开发、虚拟会议和艺术创作等应用场景，在社交、教育、娱乐和商业等领域展示了广泛应用和巨大潜力，通过其创新的虚拟现实技术和丰富的创作工具，为用户提供了一个无限可能的虚拟社交和创作空间，正在改变人们互动、娱乐和创作的方式。

3.4 典型开源技术与工具

(1) 交互技术开源工具：MediaPipe

由 Google 开发的一套开源的多模态机器学习框架，旨在简化和加速实时机器学习的应用开发。它提供了一系列预构建的、高效的解决方案，支持多种设备和平台，适用于从移动设备到服务器的多种环境。MediaPipe 采用模块化设计，提供了一系列预构建的组件，如人脸检测、手势识别、姿态估计等，如图 3-16 所示。这些组件可以单独使用，也可以组合起来实现更复杂的功能。通过结合这些模块，开发者能够快速构建复杂的多模态机器学习应用，从而节省大量的开发时间和成本。此外，MediaPipe 还支持跨平台

部署，包括 Android、iOS、桌面和 Web，开发者可以在不同的环境中轻松部署和运行 MediaPipe 解决方案。总之，MediaPipe 不仅提供了强大的功能和灵活性，还通过其高效的设计和广泛的兼容性，极大地简化了实时机器学习应用的开发过程。

代码库地址：<https://github.com/google-ai-edge/mediapipe>

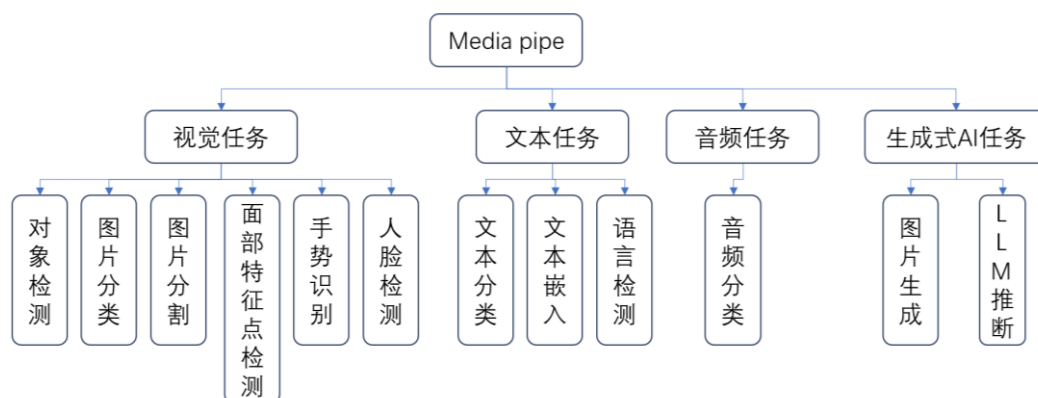


图 3-16 Media pipe 架构图

（2）区块链开源工具

• 以太坊

以太坊是最著名且最被广泛使用的项目。以太坊的官方文档是最好的区块链入门文档之一，这里详细描述了以太坊本身的特征要素，以及相关的开发指南和技术栈，此外还对区块链上最尖端的底层技术，比如扩容技术、预言机等技术有较为详细的介绍。另外，以太坊的项目代码、运行规则、升级提案等都存在其 Github 仓库中，可以帮助开发者更进一步地了解以太坊的特性和发展方向。

文档地址：<https://ethereum.org/zh/developers/docs/>

代码库地址：<https://github.com/ethereum>

• Metamask

虽然只要有网络会写代码就可以直接和区块链交互，但为了方便实用，通常人们还是会使用钱包进行交互。Metamask 是最著名的以太坊（及其兼容网络）钱包，一般我们会使用其 chrome 浏览器版本，因为与浏览器的深度绑定，导致其一度是最为易用的以太坊钱包。

代码库地址：<https://github.com/MetaMask>

• Arbitrum

是采用 optimistic rollup 技术的 ETH level 2 网络之一，也是最常用的 ETH level 2 网

络。Arbitrum 上区块链交易的执行速度和成本都远优于以太坊一层网络，而且大多数常用的以太坊去中心化应用都在 Arbitrum 上部署。

文档地址: <https://docs.arbitrum.io/welcome/get-started>

代码库地址: <https://github.com/OffchainLabs/arbitrum>

- **Chainlink**

著名的去中心化预言机，很多 ETH dapp 都依赖其服务。

文档地址: <https://docs.chain.link>。

代码库地址: <https://github.com/smartcontractkit/chainlink>

- **Solana**

采用历史证明 (PoH) 与权益证明 (PoS) 结合的高性能公链，拥有丰富的 DeFi、NFT 和游戏生态，是以太坊之外最重要的公链之一。

文档地址: <https://docs.solana.com>

代码库地址: <https://github.com/solana-labs/solana>

- **Solang**

通过 Solang，您可以将 Solidity 编写的智能合约编译为适用于 Solana 和 Polkadot 的代码。该编译器基于 LLVM 编译框架，可生成 WebAssembly (WASM) 或 Solana SBF 合约代码。由此生成的输出经过高度优化，能有效节省 Gas 费用或计算资源。

文档地址: <https://solang.readthedocs.io/en/latest/>

代码库地址: <https://github.com/hyperledger-solang/solang>

- **Uniswap**

以太坊上最著名的去中心化交易所，采用自动做市商 (AMM) 机制，为元宇宙中的虚拟资产交易提供了去中心化解决方案。

文档地址: <https://docs.uniswap.org>

代码库地址: <https://github.com/Uniswap>

- **EigenLayer**

以太坊上的再质押 (Restaking) 协议，通过共享安全模型催生主动验证服务 (AVS) 生态系统，为跨链桥、预言机等基础设施降低启动成本和安全风险。

文档地址: <https://docs.eigenlayer.xyz>

代码库地址: <https://github.com/Layr-Labs/eigenlayer-contracts>

- **FISCO BCOS**

首创 DMC 算法大幅度提升性能、推出三种架构形态灵活适配业务需求；全链路国产化，采用国密算法与软硬件体系，支持国产 OS，适配国产芯片和服务端，支持多语言多终端国密接入。拥有覆盖底层+中间件+应用组件的丰富周边组件。

文档地址：

<https://fisco-bcos-doc.readthedocs.io/zh-cn/latest/docs/introduction/introduction.html>

代码库地址：<https://github.com/FISCO-BCOS>

- **Truffle Suite**

旨在简化智能合约和去中心化应用（DApp）的构建、测试和部署流程。通过三个组件的协作，为开发者提供了一套标准化的开发流程。**Truffle**：主框架，负责项目管理、编译和部署。**Ganache**：用于本地运行一个模拟的以太坊区块链，方便快速测试和调试。**Drizzle**：前端库，帮助轻松构建与智能合约交互的用户界面。。

文档地址：<https://archive.trufflesuite.com/docs>

代码库地址：<https://github.com/ConsenSys-archive/truffle>

(3) 三维重建开源工具

- **SFM&MVS 三维重建 COLMAP**

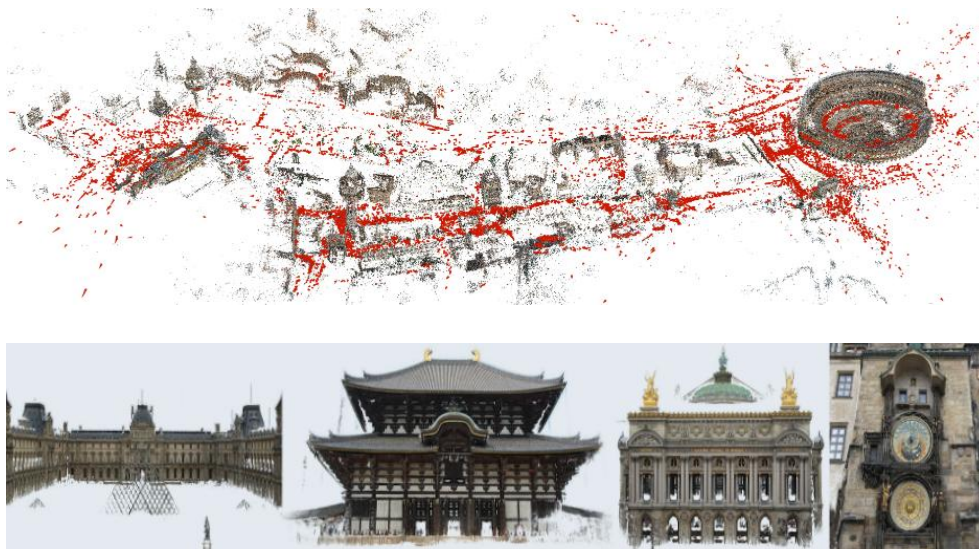


图 3-17 COLMAP 实现 SFM 和 MVS 的稀疏重建和稠密重建

COLMAP 是传统 SFM&MVS 重建的最经典的开源工具，广泛应用于三维重建和摄影测量软件领域。它通过多视图几何技术，从多张图像中提取特征点并进行匹配，生成高质量的三维模型，COLMAP 提供全自动的工作流程，包括特征提取、图像匹配、稀疏和稠密重建以及三维网格生成。COLMAP 不仅用于传统的稀疏和稠密重建，而且还

提供了图像位姿估计的能力，为 NeRF、3DGS 建模提供准确的图像位姿，图 3-17 所示为 COLMAP 实现 SFM 和 MVS 的稀疏重建和稠密重建。COLMAP 支持多种图像格式，并配备交互式 GUI 工具，便于用户查看和编辑重建过程，COLMAP 支持多平台，包括 Windows、macOS 和 Linux。

代码库地址：<https://colmap.github.io/>

- 神经光照场三维重建 NeRFStudio

NeRFStudio 是 NeRF 等基于可微渲染的三维重建算法建模的一套代表性开源工具，提供了一套完整的工具链，从数据准备、模型训练到结果可视化和导出，旨在简化和优化 NeRF 的工作流程。NeRFStudio 支持各种图像数据集格式，并提供数据预处理工具，方便用户进行特征提取和图像校准，通过集成多种深度学习框架，NeRFStudio 能够高效地训练 NeRF 模型，支持多种内置基于可微渲染的三维重建算法实现（如 NeRF，Nerfacto，Instant-NGP，Mip-NeRF，TensorRF，Splatfacto）和新的三方算法（如 ZipNeRF，Instruct-GS2GS 等），并提供超参数调整和训练监控功能，并且内置交互式的可视化工具，用户可以实时查看重建的三维场景和生成的视图合成效果，如图 3-18 所示。NeRFStudio 兼容多种操作系统，包括 Windows、macOS 和 Linux，提供了跨平台的开发和使用体验。

代码库地址：<https://github.com/threestudio-project/threestudio>

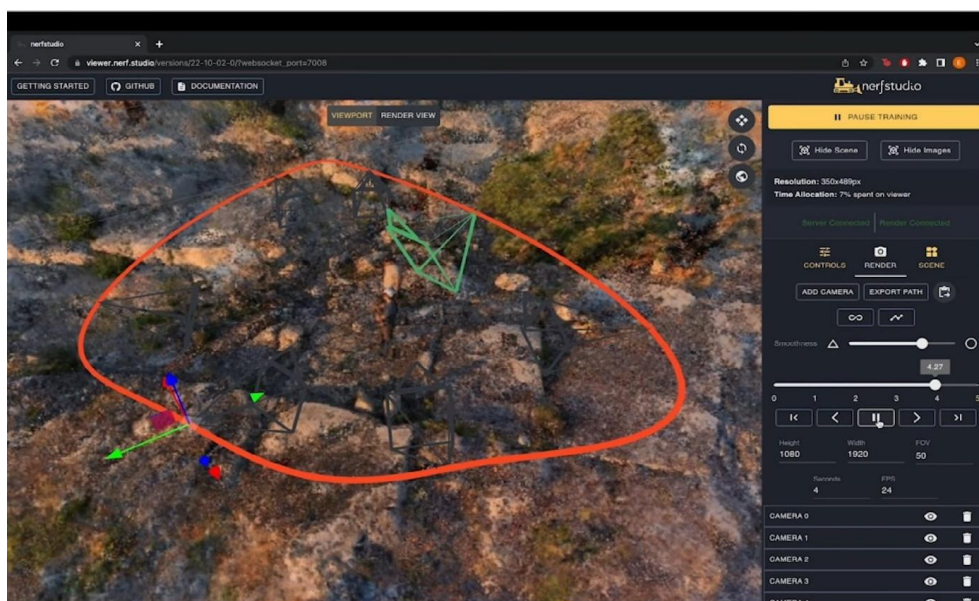


图 3-18 实现基于可微渲染引导的三维重建

(4) AIGC 开源软件: ThreeStudio 和 ThreeStudio-Extensions

这两个开源工具是开源 3D 内容生成统一框架。通过利用 2D 文本到图像生成模型,实现从文本提示、单张图像或少量图像生成高质量的 3D 内容。该框架集成了多种 3D 生成方法和扩展,如 Animate-124、DreamCraft3D 等,如图 3-19 所示,支持从简单物体到复杂场景的生成。用户可以根据具体需求选择不同配置,并支持通过 Gradio Web 界面进行操作。ThreeStudio 提供详细的安装指南和丰富的配置选项,适用于多种硬件环境,特别是 NVIDIA GPU。其主要应用领域包括 3D 建模、虚拟现实、游戏开发等。

代码库地址: <https://github.com/threestudio-project/threestudio>

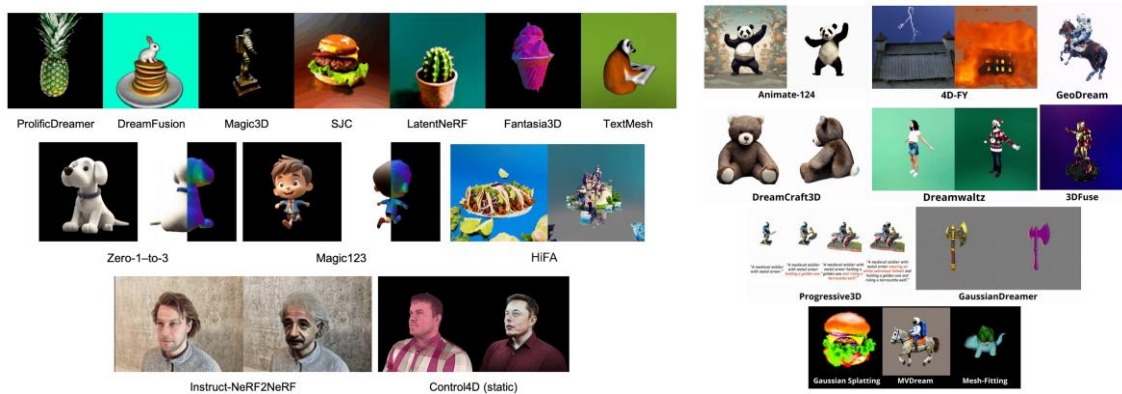


图 3-19 左: Threestudio 实现; 右: Threestudio-extensions 实现

(5) 数字人开源工具：SMPL-X

SMPL-X 是由加州大学伯克利分校的研究团队开发的，它是 SMPL (Simple Skinned Multi-Joint Model) 的扩展版本。SMPL-X 可以基于单视图进行人体模型重构与姿态估计，得到更为精细的人体形态描述，包括肌肉细节、性别差异以及面部表情，使得模型更接近真实人体。

SMPL-X 的核心在于一个参数化的三维人体模型，该模型由几千个顶点构成，通过线性混合形状和皮肤权重来模拟身体形状和姿态的变化。此外，模型还包括了一套灵活的面部表情系统，基于 Facial Action Coding System (FACS)，可以精确地控制 6890 个面部特征点。

该项目采用了优化的泊松方程求解器来实现高效的形变计算，同时利用神经网络进行先验估计，大大提高了从 2D 图像或 3D 扫描数据中恢复人体姿态和形状的精度。此外，SMPL-X 还支持与常见的动捕设备和软件无缝对接，提供了便捷的数据输入和输出接口。

代码库地址: <https://github.com/vchoutas/smplx>

(6) 开源大语言模型

• GLM (General Language Modeling)

GLM 模型系列是由智谱 AI 公司开发的自然语言处理模型,源自清华大学的技术成果,其开源版在全球的下载量超过 1000 万,成为目前下载量和开源影响力最高的国产大模型之一。智谱 AI 从 2020 年开始研发 GLM 预训练架构,2021 年训练完成百亿参数模型 GLM-10B,2022 年合作研发了中英双语千亿级超大规模预训练模型 GLM-130B。2023 年,智谱 AI 推出千亿基座对话模型 ChatGLM,并两次升级。2024 年,GLM-4 模型推出,支持更长的上下文,具备更强的多模态能力,推理速度更快,支持更高并发,大大降低推理成本。GLM-4 模型的整体性能相比上一代有大幅提升,特别是在中文能力方面可以与 GPT-4 相媲美,并能够根据用户意图自动理解、规划复杂指令,自由调用网页浏览器、代码解释器和多模态文生图大模型来完成复杂的任务。智谱 AI 还推出了个性化智能体定制功能,用户即使没有编程基础,也能通过简单指令创建属于自己的 GLM 智能体,降低了大模型的使用门槛。

在线链接: <https://github.com/THUDM/GLM-4>

• Llama (Large Language Model Meta AI)

Llama 是 Meta AI 开发的先进自然语言处理模型,旨在推动 AI 在语言理解和生成领域的前沿。2023 年 2 月,Meta 推出了 LLaMA 大模型,使用了 1.4T token 进行训练,虽然最大模型只有 65B,但在相关评测任务上的效果可以媲美甚至超过千亿级大模型,被认为是近期开源大模型百花齐放的开端之一。Llama 整体 Transformer 架构与 GPT-2 类似,但所采用的 Transformer 结构和细节与标准的 Transformer 架构有所不同,包括采用了前置层归一化 (Pre-normalization) 并使用 RMSNorm 归一化函数 (Normalizing Function)、激活函数更换为 SwiGLU,并使用了旋转位置嵌入 (RoP)。由于自然语言处理的需求量巨大,Llama 系列模型及其生态快速发展,主要版本如下: 1) Llama 1: 初代 Llama 模型,展现了强大的语言生成能力。2) Llama 2: 改进版本,增强了理解和生成能力,优化了训练算法和数据集。3) Llama 3: 进一步提升了模型的性能和应用范围,针对特定任务进行了微调。

在线链接: <https://llama.meta.com/>

(7) 协同安全开源工具

- **uPort**

uPort 是一个去中心化身份解决方案，旨在为用户提供自主掌控的数字身份，支持用户在不同平台间进行安全的身份验证和数据共享。

代码库地址：<https://github.com/uport-project>

- **Ceramic Network**

Ceramic Network 是一个去中心化的身份和数据存储网络，支持数据的持续更新和共享。该网络为开发者提供了构建动态去中心化应用的基础设施。

代码库地址：<https://github.com/ceramicnetwork>

- **Torus**

Torus 是一个密钥管理系统，致力于简化用户身份认证和密钥管理过程。通过 Torus，用户可以使用现有的社交账户登录去中心化应用，确保便捷性和安全性。

代码库地址：<https://github.com/torusresearch>

- **NuCypher**

NuCypher 是一个提供代理重加密技术的加密服务网络，旨在确保数据的安全共享和隐私保护。该技术允许数据所有者在不直接分享密钥的情况下，委托代理重新加密数据。

代码库地址：<https://github.com/nucypher>

- **OpenZeppelin Contracts**

OpenZeppelin Contracts 是一个智能合约安全模板库，提供了经过审计的、可复用的智能合约组件，帮助开发者构建安全可靠的去中心化应用。

代码库地址：<https://github.com/OpenZeppelin/openzeppelin-contracts>

- **Mythril**

Mythril 是一个智能合约安全分析工具，旨在检测智能合约中的潜在漏洞和安全问题。它支持静态分析和符号执行，帮助开发者识别和修复安全漏洞。

代码库地址：<https://github.com/trailofbits/mythril>

- **Aztec**

Aztec 是一个隐私保护基础设施，提供了零知识证明技术，确保区块链上的交易隐

私。该协议支持机密交易和数据保护，为去中心化应用提供隐私保障。

代码库地址: <https://github.com/AztecProtocol>

- **OpenSSL**

OpenSSL 是一个强大的开源工具包，用于实现 SSL 和 TLS 协议。广泛应用于网络通信中，确保数据传输的安全性。

代码库地址: <https://github.com/openssl/openssl>

- **Tor Project**

Tor Project 是一个免费的开源软件，能够保护用户的匿名性，防止网络流量被追踪。适用于在元宇宙中保护用户隐私。

代码库地址: <https://www.torproject.org/>

- **Snort**

Snort 是一个开源网络入侵检测和防御系统，可以实时监控网络流量并识别潜在的安全威胁。

代码库地址: <https://github.com/snort3/snort3>

- **Suricata**

Suricata 是另一个开源 IDS/IPS 引擎，提供高性能的网络安全监控功能。

代码库地址: <https://github.com/OISF/suricata>

- **星际文件系统 (InterPlanetary File System, IPFS)**

IPFS 是一个去中心化的分布式文件系统，旨在实现更快、更安全和更开放的互联网。

开发文档地址为: <https://docs.ipfs.tech>

代码库地址: <https://github.com/ipfs>

- **OpenRefine**

OpenRefine (以前称为 Google Refine) 是一款强大的数据清洗工具，适用于处理大型数据集。它支持数据转换、清洗、格式化等功能，可以处理复杂的数据清洗任务。

开发文档地址为: <https://openrefine.org/docs>

代码库地址: <https://github.com/OpenRefine/OpenRefine>

- **Libsnark**

Libsnark 是一个用 C++ 编写的开源库，提供了构建 zk-SNARKs 所需的基本工具。

代码库地址: <https://github.com/scipr-lab/libsnark>

- **HElib**

HElib 是一个实现同态加密（HE）的开源代码库，能实现加、减、乘的全同态加密操作，保障计算过程安全。

代码库地址：<https://github.com/HomEnc/HElib>

- **PrimiHub**

PrimiHub 是一个开源隐私计算平台，它支持多方安全计算、联邦学习等多种隐私计算技术，保障数据保密性及完整性。

代码库地址：<https://github.com/primihub/primihub>

- **DataHub**

DataHub 是一个开源的元数据平台，旨在帮助企业管理和治理其数据资产。它由 LinkedIn 开发，能够捕获、存储和查询各种数据资产的元数据。

代码库地址：<https://github.com/datahub-project/datahub>

- **Pillow**

Pillow 是 Python Imaging Library（PIL）的一个分支，提供了强大的图像处理能力，可以用于在图像上添加水印。

代码库地址：<https://github.com/python-pillow/Pillow>

（8）世界模型开源工具

- **真实世界世界模型：YUME**

YUME[87]由开源社区开发的一套生成式世界模型框架，旨在简化和加速交互式虚拟世界（Interactive World）与可探索场景（Explorable Environment）的构建。它提供了一系列预构建的、多模态输入支持的世界生成能力，能够根据文本、图像或视频创建具有连贯结构、可导航、可操作的三维动态场景。YUME 采用模块化设计，提供了一系列核心组件，如场景理解、世界构建、交互控制和动态渲染等。这些组件可以单独使用，也可以组合起来，实现更复杂的世界模型功能。通过组合这些模块，开发者能够快速构建交互式虚拟环境、沉浸式体验系统或智能体训练世界，从而节省大量的开发时间和成本。此外，YUME 支持多模态输入，并预留跨平台扩展能力，能够在桌面端、训练平台和未来的移动设备环境中部署和运行。总之，YUME 不仅具备强大的生成能力和灵活性，还通过其模块化结构和对多场景的兼容性，极大地简化了世界模型应用的开发过程。

代码库地址: <https://github.com/stdstu12/YUME>

- **Matrix-Game 系列**

Matrix-Game[88]世界模型是一类聚焦游戏场景智能生成的特色化世界模型，其核心价值在于以“矩阵化规则映射”为技术核心，为《我的世界》（**Minecraft**）、《神庙逃亡》等不同类型的游戏提供高效、多样且符合玩法逻辑的场景生成方案。与通用世界模型不同，它将游戏的核心玩法约束（如地形通行规则、障碍物分布逻辑）、视觉风格标准（如像素化或写实化纹理）及交互反馈机制，拆解为可量化的矩阵参数，通过构建“元素组合矩阵”与“规则约束矩阵”的双维框架，实现场景生成的精准可控。例如在《我的世界》中，它可将方块类型、生物群系特性、资源分布概率转化为关联矩阵，快速生成兼具探索性与合理性的随机地形，同时确保矿脉位置、村庄布局符合游戏平衡规则；在《神庙逃亡》这类跑酷游戏中，则通过动态调整“障碍物类型-出现间隔-路径分支”的矩阵参数，生成难度循序渐进且避免重复的跑道场景，适配玩家的操作节奏。此外，该模型还能融合玩家行为数据优化生成逻辑，比如根据玩家在游戏中的偏好路径，动态调整矩阵权重，让生成的场景更贴合不同玩家的游玩习惯。这种将游戏场景生成需求与矩阵化建模深度结合的特性，使其成为游戏开发中降低场景设计成本、提升内容多样性的关键工具，为 **procedural content generation**（**PCG**，程序化内容生成）技术提供了高效的实现路径。

代码库地址: <https://github.com/SkyworkAI/Matrix-Game>

第4章 元宇宙典型应用

元宇宙作为新质生产力重要应用场景，各部委各地市为促进元宇宙的发展应用，累计出台相关政策超 150 项。23 年工信部等 5 部委联合印发《元宇宙产业创新发展三年行动计划（2023—2025 年）》、25 年 8 月国务院印发《关于深入实施“人工智能+”行动的意见》、25 年 9 月商务部等 8 部委印发《关于大力发展数字消费共创数字时代美好生活的指导意见》，明确提出加快实施“人工智能+”消费提质重点行动，加快人工智能与元宇宙等技术融合和产品创新，为元宇宙产业与实体经济深度融合、消费市场提质升级提供了政策支撑与发展路径。在应用场景创新方面，明确指出需“人工智能+”消费提质，鼓励运用元宇宙等技术创新表现形态，创新数字内容消费；鼓励运用人工智能、虚拟现实、增强现实等信息技术打造沉浸式、体验式的多元化消费场景。

元宇宙的应用已经深入千行百业。在文旅元宇宙中，用户可以虚拟游览全球名胜古迹，体验历史文化，推动旅游业新模式和文化遗产保护。教育元宇宙通过沉浸式学习环境，使学生能在虚拟教室中与全球师生互动，提高了学习的趣味性和效率。工业元宇宙通过数字孪生和虚拟现实，为工业生产提供解决方案，提升了生产效率和安全性。办公元宇宙通过协同办公和 AI 辅助，促进全球团队协作，提升办公效率。体育元宇宙让用户以更沉浸的方式观赛，参与多样化的数智竞技，增加了体育活动的趣味性和参与度。娱乐元宇宙为用户提供虚拟演唱会、游戏和社交活动，扩展了娱乐产业的边界。金融元宇宙通过区块链技术，提供安全透明的交易环境，促进数字资产管理和流通。供应链元宇宙利用虚拟现实和区块链技术，提升了供应链的可视化和透明度，优化物流和库存管理。农业元宇宙通过虚拟现实和物联网，实现作物实时监控和智能管理，提高了农业生产效率和可持续性。

元宇宙技术的快速发展不仅赋能了各个行业，提升了行业效率，更加速了数字经济的创新与变革。随着技术的不断进步，元宇宙将在更多领域展现出其无限潜力，推动整个社会的运行效率迈上新台阶。

4.1 文旅元宇宙

文旅行业作为传承文化血脉、拉动经济增长、满足民众精神需求的重要支柱，在新时代发展中却面临着诸多亟待突破的瓶颈与挑战：1）体验形式固化单一：当前多数文旅

场景仍以“观光式”“讲解式”为主，游客多处于被动接收状态，难以深度参与其中，对文化内涵的感知流于表面，无法形成强烈的情感共鸣与记忆点。2）文化传播效能不足：大量非物质文化遗产、历史典故等文化资源因缺乏生动载体，难以突破时空限制触达更广泛受众，部分小众文化资源甚至面临传承断档的风险。3）资源开发与承载失衡：热门景区每逢节假日便陷入“人满为患”的困境，游客体验大打折扣；而一些偏远地区的优质文旅资源则因地理位置、宣传力度等因素，面临开发不足、鲜为人知的尴尬局面。以 VR 大空间、虚实融合交互、数字导览员等为代表元宇宙技术实现沉浸式场景构建、高质量实时渲染等，为破解上述文旅发展难题提供了全新的技术路径与应用形态。根据 VR 陀螺统计，截止 2025 年 8 月，全国超 2000 家大空间体验门店，预计到 2025 年年底，市场总门店数量或突破 3000 家。2025 年 3 月，国家电影局发布《关于促进虚拟现实电影有序发展的通知》，规范 VR 大空间向影院体验发展。截至 2025 年 9 月，共有 48 部虚拟现实电影完成立项并获得“龙标”资格，涵盖历史、科幻、文化等多种题材，标志着 VR 大空间开始进入规范化、高质量发展阶段。

（1）《琴岛奇遇》VR 大空间

基于 5G-A、高精度 3D 建模、分布式实时渲染等研究成果，中国移动咪咕深入布局 XR 沉浸式业态，打造了《琴岛奇遇》VR 大空间。《琴岛奇遇》以首个鼓浪屿 IP 元宇宙原创故事《奇妙家庭音乐会》为蓝本，通过“八闽文脉·记忆”“乐动琴岛·守护”“钟声依旧·时光”“以乐会友·圆满”四大篇章生动串联起鼓浪屿的如画景致和音乐秘境。游客只需穿戴 VR 眼镜和手柄，便能以“家庭音乐会嘉宾”身份跟随“时空领航员”穿越时空隧道，以第一视角开启寻找、守护、修复的鼓浪屿时空之旅。游客通过本案例既能畅游鼓浪屿秀美风光和人文建筑，也能通过触发多重剧情，加深对鼓浪屿深厚历史文化底蕴的了解。本案例已获得 2025“金海豚奖”虚拟现实数字作品大赛“闽台文化虚拟现实（VR）数字作品金奖”“VR 大空间 2024-2025 年度应用潜力奖”，入选“福建省文化和旅游厅 数字文旅应用场景优秀案例”“2025 UWA 超高清产业应用案例”。



图 4-1 “琴岛奇遇”大空间

（2）《敦煌奇境》VR 大空间

中国移动咪咕公司携手敦煌研究院、零卡文化运用 5G+XR 技术突破传统展览局限，打造了《敦煌奇旅》VR 大空间项目。作为国内首个聚焦第 220 窟的多媒体展览的延续之作，《敦煌奇旅》体验时长约 30 分钟，围绕“一念”“一夜”“千年”及“绝响”四大篇章展开，观众将扮演洞窟守护者的角色，踏上帮助翟通寻找第 220 窟壁画粉本的冒险之旅。在旅途中，不仅可 360° 自由漫游于大唐盛世，领略长安盛景，还能深入第 220 窟与壁画互动，触发丰富的知识点，并解锁唐韵音律的惊喜彩蛋。本项目通过高精度画面、富有创意的故事线、易于操作的互动设计，生动展现了第 220 窟壁画和大唐乐舞景象，让观众在享受视听盛宴的同时，深入感受唐代长安与敦煌两地的历史渊源，以及敦煌作为文化交汇地的重要意义。基于“咪咕未来空间”计划，本项目自 2024 年 11 月亮相以来便已在多个省市的中国移动营业厅成功落地应用，包括北京西单大悦城厅、广州全球通大厦厅、江苏泰州厅、四川眉山厅、上海长寿路厅、惠州厅、南宁厅等标杆厅店，初步形成了覆盖华北、华东、华南、西南的示范网络。



图 4-2 《敦煌奇旅》大空间

4.2 教育元宇宙

教育行业在当今社会扮演着至关重要的角色，当前我国教育仍然面临一些主要问题和挑战：1) 教育资源不均衡：城乡、地区之间教育资源的分配不均衡是当前中国教育面临的主要问题之一。一线城市和发达地区的教育资源相对充裕，而一些农村地区和欠发达地区的教育资源严重匮乏。2) 教育质量参差不齐：我国教育体系存在着以升学率和考试成绩为导向的倾向，导致一些学校过度注重应试教育，忽视了学生的全面发展和创新能力的培养。教育与就业之间脱节。3) 教育与就业之间存在着一定程度的脱节，一些学生在校期间接受的教育与实际工作需求不相符，导致毕业生就业能力不足，出现了大量的“应届生失业”现象。元宇宙技术，包括 ARVR 设备、显示技术、SLAM、实时交互、虚拟培训学习、虚实融合指导、智能教具、智能考核判断、AIGC、数字人等，在解决上述挑战中提供了新工具新形态。

(1) VR 虚拟教学和模拟训练

通过元宇宙 VR 教学平台，学生可以进行各种科学实验和职业训练，无需真实的实验室或工作场所。例如，学生可以在虚拟实验室中进行化学、物理或生物实验，模拟各种工程项目的设计和操作，从而获得实践经验和技能培训。

咪咕以北京新首钢高端产业综合服务区的炼钢厂旧址场景为基础，开发了覆盖小初

中研学体系,并通过咪咕数智竞技科技中心提供研学服务。其中,《钢铁是怎样炼成的》沉浸体验课程采用游戏化交互设计,学习者通过 VR 设备即可在 360° 环绕的熔炉场景中自主探索钢铁工业奥秘,从历史、物理、化学、政治等各个角度学习钢铁制造的过程,实现知识传授与科技体验的深度协同。

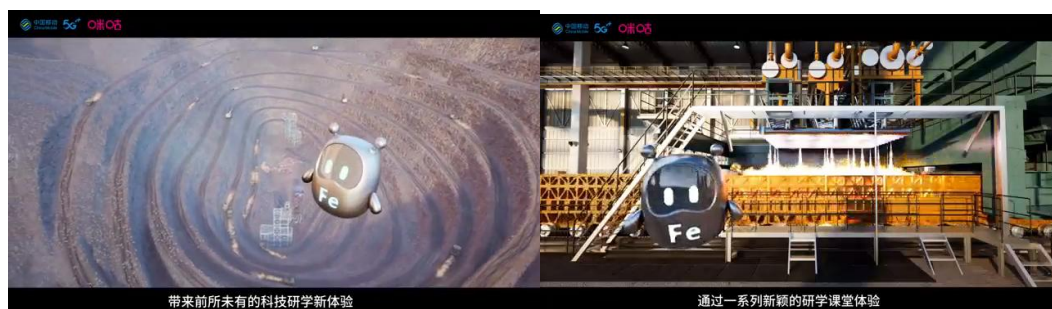


图 4-3 《钢铁是怎样炼成的》沉浸体验课程效果

(2) AR 实训和智能考核

通过元宇宙 AR 教学平台,学生可以进行实操训练,通过虚实融合的指导操作方式,强化学习记忆,加深实操印象,同时结合元宇宙视觉检测考核能力,实时检测学生操作正误并实时反馈。

2022 年 10 月,南方电网应急实训部引入了 AR 实训及考核方案。如图 4-4 所示,该方案基于应急培训需要的实物、人员、流程、环境等,构建了智能教具,基于 AR 设备实现在真实环境中增强现实指导信息,同时支持远程专家接入指导;基于 MR 摄像机,以第三视角直播现场虚实操作过程,可以让其他学员通过直播平台查看学习。该应用在教学技术成熟度、课程内容具体化、AR 技术与培训需求的融合度上在当时都达到了国内一流、国际领先水平,获得广东电网各地市局专家团队的一致认可,并入选亚洲创新成果展览。

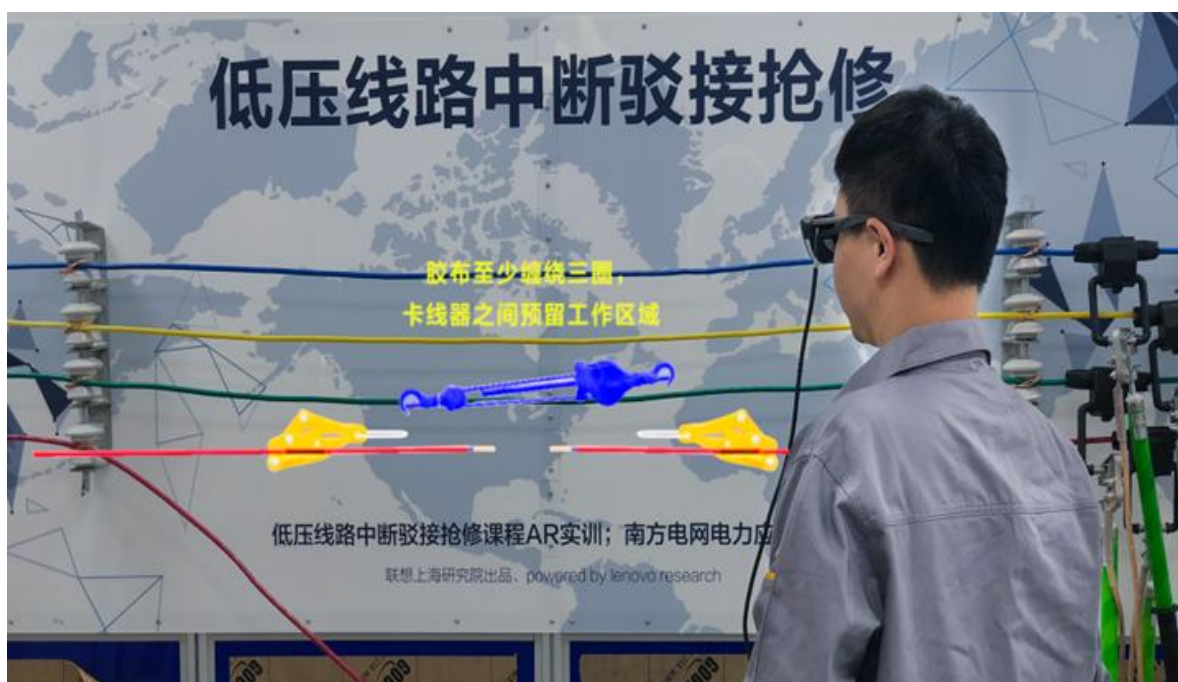


图 4-4 基于 AR 设备的南网应急 AR 实训

丰田北美汽车公司拥有数万名团队成员，每年在美国、加拿大和墨西哥生产 100 多万辆汽车。确保所有员工清楚如何组装最新型号需要一个强大的培训计划，通过大量的教学来帮助新人在不中断生产过程的情况下进行学习。2021 年 10 月，丰田汽车北美公司在 HoloLens 2 部署了 Dynamics 365 Guides 和 Dynamics 365 Remote Assist，如图 4-5 所示。培训师可以快速轻松地创建和更新内容，并同时向多个参训人员讲授内容。另一方面，公司可以跟踪相关的学习进度，以确定优化培训的机会。当需要实时专家支持时，团队成员可以使用 Dynamics 365 Remote Assist 来快速与远程专家对接，并在真实工作环境中启用 3D 注释。

目前元宇宙技术在教育领域已经展现出巨大的潜力，但仍然存在较多的局限性，包括技术成熟度不高、应用开发门槛高、ARVR 终端硬件普及率低、缺少优质的课程内容和教学设计、缺乏统一的标准和规范等问题。

随着 ARVR 设备、显示技术、SLAM 技术、AIGC 等元宇宙技术的不断成熟和在教育领域应用场景的探索，元宇宙教育应用的应用开发成本、设备成本将进一步降低，更多的优质课程内容将不断涌现，我们相信未来教育行业将迎来更加均衡化、个性化、智能化的变革。



图 4-5 基于 Hololens 和 Dynamics 365 Guides 的汽车公司员工培训

4.3 工业元宇宙

工业作为国民经济的支柱产业，承载着实体经济高质量发展的核心使命。我国已成为全球制造业大国，2024 年制造业附加值（2015 年不变美元价格）达 5.06 万亿美元，是美国的 2.13 倍，制造业总体规模连续 15 年居全球首位，为向“工业强国”跨越奠定坚实基础。但在发展进程中，我国工业仍面临诸多亟待突破的挑战：1）生产效率与质量瓶颈：部分传统制造环节依赖经验驱动，设备运维以事后维修为主，易出现停工损失，精密制造领域的工艺精度与产品一致性仍有提升空间；2）创新转化效率偏低：产学研融合不够紧密，实验室技术向工业生产线转化的周期长、成本高，关键核心技术的产业化应用存在壁垒；3）资源协同能力不足：供应链各环节信息孤岛现象突出，生产计划与市场需求衔接不及时，跨区域、跨企业的产能调配与资源共享效率低下。《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》明确提出促进实体经济和数字经济深度融合，工业互联网作为重要支撑，2024 年核心产业规模已达 1.53 万亿元，标识解析体系基本形成。而工业元宇宙在工业互联网基础上实现进阶升级，深度融合虚拟现实（VR/AR）、数字孪生、区块链、物联网、大数据、人工智能等技术，为破解工业发展难题、构建高级形态数实融合工业生态提供了全新路径。

(1) 工业互联网工厂

2024 年 5 月，微网优联携手中移物联、成都移动、华为等单位，建设了四川首个 3C 电子行业“5G+全连接工厂”。该项目融合中国移动 5G 专网运营平台 OneCyber、园区智服平台 OnePower、工业互联网平台 OnePark 等三大平台打造新型工业互联网基础设施，将企业生产数据、设备数据等数据沉淀为数据资产，以数据驱动企业在生产过程中的决策和运营，实现产线级/车间级/工厂级等生产现场的广泛连接、信息和运营深入融合、数据要素充分利用，项目实现设备 100%互联、工厂自动化率提升 33%、人均生产效率提升 43%、应用上线周期缩短 30%、安全巡检投入降低 60%，以工厂的工业互联网升级实现降本增效。



图 4-6 微网优联“5G+全连接工厂”



图 4-7 中移物联园区智服平台 OnePower 整体架构

（2）虚拟部署与模拟仿真

随着机器狗智能巡检在南网东莞局的试点成功，东莞局机器狗“吠云”成为全国率先实现变电站设备巡视全覆盖的单台智能装备。2025 年为加快推广试点成果，助力南网“机器代人”的变电站巡检任务目标达成，南网携手联想等单位，在南网 5 省区域（广东、广西、云南、贵州、海南）20 余所变电站上线了电力机器人智能巡检系统，实现了四足机器人自主巡视。项目上线的过程中，基于元宇宙技术的机器人虚拟部署与模拟仿真技术，为机器人的批量部署实施快速上线提供了坚实的基础。项目实施阶段，通过激光雷达及 RGB 相机对变电站进行扫描重建，生成 1:1 高拟真的虚拟三维场景，如图 4-8 所示；基于孪生体、空间计算技术，将物理世界中的巡检机器人在虚拟空间中构建，并在虚拟空间中进行任务部署，包括停靠位置生成，拍摄范围生成，机器人行走路径规划生成及运行模拟，如图 4-9 所示；在虚拟空间中得到准确的部署数据并一键同步至物理环境机器人，开启机器人自主巡检巡逻，将传统的耗时耗力的机器人勘探部署环节由线下搬至线上，实现了部署效率 600%的提升和部署成本 180%的降低。



图 4-8 1：1 高拟真虚拟三维电站



图 4-9 机器人任务虚拟部署及模拟仿真运行

(3) 车间智能监控与运维

2024 年 3 月，中烟下属某卷烟厂细支烟制丝车间上线一套开放性的孪生智能监控运维系统，实现制丝的精益化智能管理。在图 4-10 所示系统中，基于三维重建技术重建了虚拟工厂模型，系统结合工业相机、深度相机等实现对产线生成的实时图像采集，及接入烟厂设备 IOT 平台，实现了 3000kg/h 制丝线的统一管理，生产过程实时同步及生产调度产前仿真，提升了生产组织的有效性和产品质量的稳定性，实现了设备故障管理、物料平衡分析、人员安全作业管理。



图 4-10 卷烟制丝车间孪生智能监控运维系统

凭借着数字孪生工厂、虚拟规划与模拟仿真、智能监控与运维、远程协作与培训、数据分析与优化等应用的落地，元宇宙技术在工业领域已经展现出巨大的潜力。随着元宇宙构建成本的降低和元宇宙智能应用的价值释放，将有更多的工业企业进行元宇宙数实融合数字化转型。面向未来，我们可以预见工业元宇宙将向工业运行机理、仿真预测、自主决策等纵深方向持续发展并发挥更大的价值。

4.4 办公元宇宙

办公元宇宙指借助虚拟现实、人工智能、语音识别、空间定位等新兴元宇宙技术突破时间和物理空间的限制，为用户带来最接近面对面工作的体验，达到提升办公、生产、沟通、协作效率的目的。新冠疫情以来，全球远程办公需求激增，催生了更多的办公元宇宙需求场景，也带来了更多的挑战。

(1) 远程协作平台

Horizon Workrooms 是一个虚拟会议空间，能够不限地点支持您和您的同事开展更好地协作，如图 4-11 所示。以卡通形象加入 VR 会议，或者通过您的电脑拨号加入会议。利用虚拟白板直观展示您的想法。将您的电脑和键盘带入 VR 世界，与其他人开展协作，或者只需进行富于表现力的对话交流，尽享身临其境般的协作体验。借助 Horizon

Workrooms 提供的 VR 空间，团队成员可以联系交流、开展协作和集思广益，即使天各一方，也能与团队成员围坐在桌前畅所欲言。

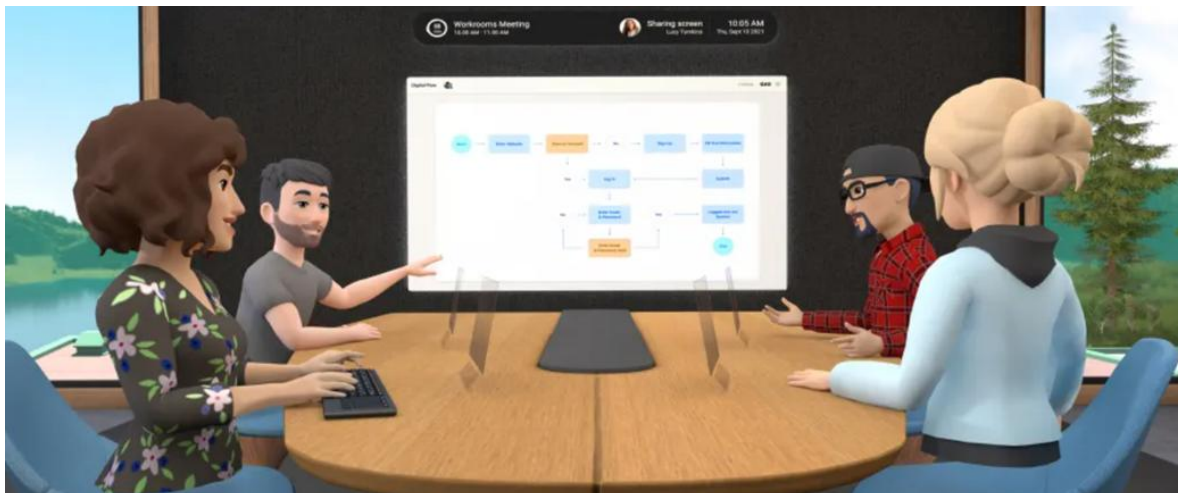


图 4-11 Horizon Workrooms 虚拟会议空间

(2) AI PC

联想在 2024 年 4 月份的联想创新科技大会上一口气连发 10 余款 AI PC 产品，宣告 AI PC 时代的正式到来。联想 AI PC 配备个性化的个人智能体——联想小天，它是一个具备内嵌个人大模型与用户自然交互的智能体，由嵌入本地的天禧大模型驱动，通过自然交互，在工作、学习和生活等诸多场景中带来全新的 AI 体验，搭载了十余款堪称“杀手级”的联想小天核心应用，包括 AI 画师、AI PPT、文档总结、知识问答、AI 识图、会议纪要等等，且还在不断增加。用户只需要通过快捷键一键调出联想小天，即可以使用语音或键盘输入等多种自然交互方式使用这些 AI 技能。

微软于 2024 年 5 月发布新一代 Copilot+PC，接入 GPT-4o 提供各种实时 AI 能力。可以在打游戏的时候和 AI 对话，实时翻译字幕、以及上文提到的 recall anything 等。其革新主要体现在系统层面，同时端侧也深度集成了 40 多个 AI 模型，使得 Copilot 具备更强大的 AIGC 能力，如图 4-12 所示。

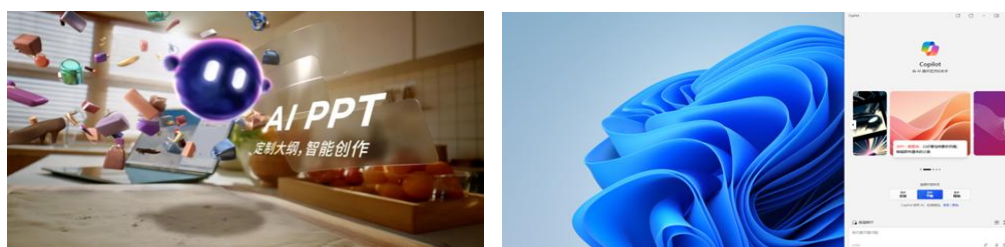


图 4-12 联想 AI PC-小天智能体（左），微软 Copilot PC（右）

凭借着元宇宙会议虚拟空间、远程协作平台、AIPC、AI 办公系统等元宇宙软硬件平台的落地，元宇宙技术在办公领域已经展现出巨大的潜力。当下，以虚拟会议系统为代表的软件平台还存在着虚拟角色恐怖谷效应，以 AIPC 为代表的硬件平台还存在着应用不足、价格高昂等问题。面向未来，我们可以预见办公元宇宙将向 AI 智能体、虚拟办公协同等纵深方向持续发展并发挥更大的价值。

(3) AI 辅助办公的新进展

2024 年以来，AI 技术的突破性进展为办公元宇宙带来了前所未有的变革。端侧 AI 算力实现质的飞跃，M4 Max、AI395 等处理器开始具备强大的 AI 能力，使垂直领域端侧大模型部署成为现实。高性能笔记本可在本地推理 32B 模型（约 15 tps），足以胜任日常办公任务，企业无需依赖云端服务即可获得智能办公体验，满足数据安全和离线需求。同时，DeepSeek 等开源模型使中小企业定制化部署成为可能：采用双路 Xeon 处理器搭配 3TB DDR4 内存即可运行 671B 参数大模型，虽然推理速度仅为 2 tps，但随着厂商解决内存带宽瓶颈、或采用“内存驻留+GPU 加速”等优化方案、甚至利用 V100 等二手服务器集群等经济方案，小型企业可在可控成本内实现本地大模型部署。软件层面，Cursor、Claude Code 等 AI 编程助手结合能力突飞猛进的大模型、RAG 技术和 MCP 协议，使人工智能与日常办公场景深度绑定。这些工具已从简单代码补全进化为能够理解上下文、自主规划任务、多步骤协同工作的智能代理，实现了“AI 辅助的零代码基础编程”。硬件方面，Vision Pro 等 MR 设备通过直观手势交互和空间计算能力降低了虚拟世界操作门槛，使虚拟办公和会议场景成为可能。

基于这些技术支撑，许多企业当前已开始将 AI 技术用于辅助办公。实践表明，完全不懂编程的金融交易员通过 AI 对话式交互，实现了从数据爬取、清洗、批量处理到报表生成的完整自动化流程；程序开发人员利用 AI 工具理解项目结构、自动重构代码、生成测试用例，开发效率获得数倍提升。类似应用正在各行业快速普及：财务人员自动处理发票识别和账务核对，市场人员批量生成营销内容和数据分析报告，行政人员实现会议纪要自动生成和日程智能管理，研究人员进行文献检索和报告撰写。企业可建立本地化 AI 办公助手系统，既保证数据安全，又能根据特定业务定制优化，形成专属知识库和 workflows。AI 已从辅助工具转变为深度参与业务流程的“数字同事”，为办公元宇宙实现奠定了坚实基础。

4.5 娱乐元宇宙

元宇宙的概念最初源自科幻小说，但其实际发展极大地借鉴了视频游戏领域的技术与设计理念。以虚拟形式存在的游戏不仅为元宇宙提供了理想的试验场，还成为了元宇宙展现的主要形式之一。实际上，在元宇宙这个概念变得广为人知之前，具备元宇宙特征的应用几乎全部来自于游戏领域。

自 20 世纪 90 年代起，随着互联网的发展，一些开放世界多人在线游戏开始展现出元宇宙的特性。经过二十多年的演进，这种结合已形成显著规模并走向成熟，催生了诸多创新的发展路径。接下来将通过时间线回顾，并结合国际和国内的实践经验，展示一些具有代表性的娱乐元宇宙案例。

(1) EVE Online

在 2000 年前后，互联网的快速发展促成了大批具有元宇宙特征的网络游戏，除了著名的 *Second Life*[322] 以外，冰岛公司 CCP Games 开发的 *EVE online* 也是其中的佼佼者。它以庞大的虚拟宇宙、复杂的经济系统和玩家之间高度自主的互动而闻名，这些特点使其成为探讨元宇宙的极好案例。

EVE Online 构建了一个庞大且复杂的虚拟宇宙，包含超过 7000 个星系，每个星系都有自己的行星、空间站和矿场。玩家在这个沙盘式世界中扮演太空飞行员，可以从事探索、采矿、制造、贸易和参与 PvP/PvE 战斗等活动[323]，如图 4-13 所示。*EVE* 的经济由玩家驱动，涵盖从原材料采集到高级产品制造所有环节，市场价格由供需决定[324]。游戏中的社交结构复杂，玩家可以选择成为守法公民、加入玩家军团、成为星际商人或海盗等社会角色，玩家间的合作与对抗形成了丰富的社会动态。*EVE Online* 的时间与现实世界同步，一场战役甚至可能持续数日。玩家行为不断影响着虚拟世界的发展，谱写了这个世界中的历史[325]。

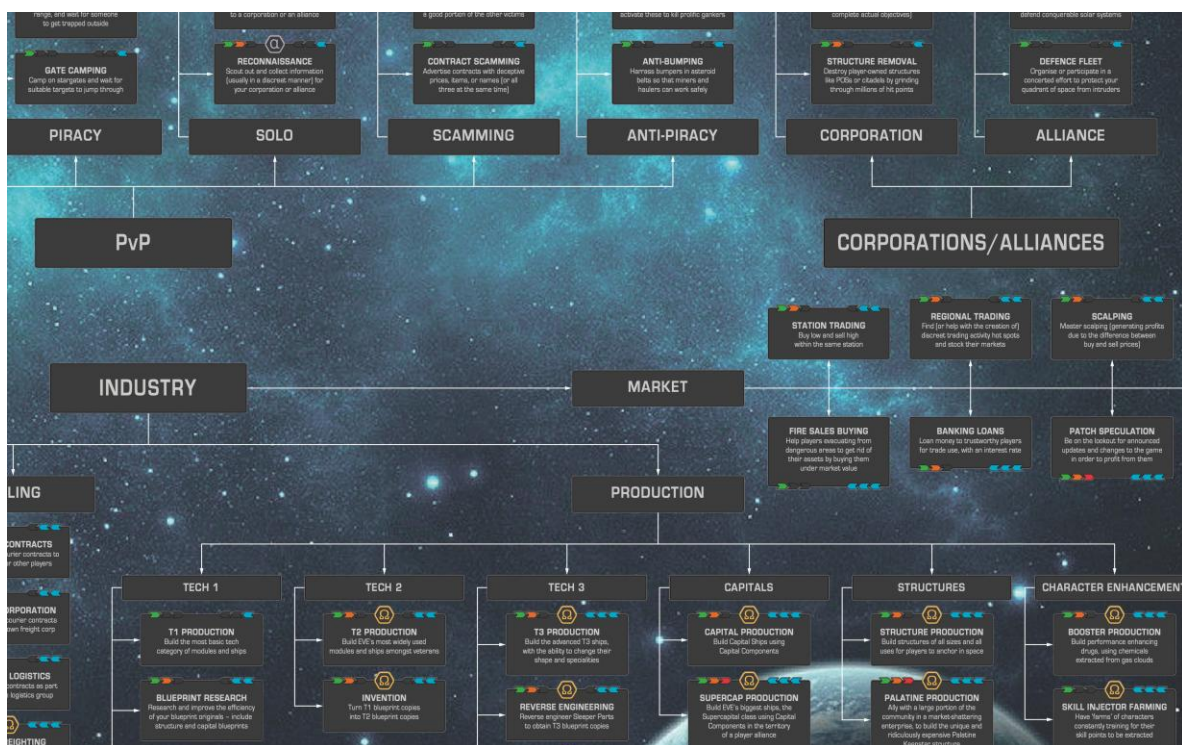


图 4-13 玩家可以在 EVE Online 世界中从事活动一览表

通过这些特性，EVE Online 展示了一个虚拟世界可以如何具备元宇宙的核心特征，尤其是在经济系统、社交互动、身份表达和系统可持续发展方面。虽然不具备现代元宇宙可能包含的所有技术如 AR 或 VR，但它的复杂性和玩家的深度参与已经非常接近元宇宙的概念[326]。

(2) Roblox

2021 年，知名游戏公司 Roblox[327]上市，迅速成为元宇宙在投资领域的代名词，这也成了元宇宙发展的重要里程碑，标志元宇宙概念开始受到商业上的认可[328]。

诞生于 2004 年的 Roblox 目前已经成长为全球最大的多人在线游戏创作平台，它不仅仅是一个游戏，更是一个创意孵化器，让每个玩家都能成为游戏世界的创造者。如图 4-14 所示，在 Roblox 中，玩家可以体验到各种类型的游戏，从经典的赛车、射击到角色扮演、解谜等，种类繁多，应有尽有。但 Roblox 的魅力远不止于此。其核心在于提供了强大的开发工具集，让玩家可以设计自己的游戏规则、建造独特的环境、编写引人入胜的故事线。这种创造力的自由发挥，使得 Roblox 成为了一个充满活力的社区，每天都有新游戏诞生。



图 4-14 Roblox 世界中多样化的应用

Roblox 提供了一个沉浸式的虚拟世界，其中内容主要由用户创造，涵盖游戏和各种虚拟物品。它不仅支持跨平台访问，还内置了虚拟经济系统和丰富的社交功能，使用户可以在连续的虚拟环境中进行交互、交易和社区建设，展现了元宇宙的核心特征：创造性、经济互动和社交互动[329]。Roblox 证明了虚拟世界的价值，以及用户生成内容(UGC)在游戏行业中的巨大潜力。其货币系统和经济模型，为玩家提供了将创意变现的机会，让游戏创作成为一种可能的职业。

(3) 中国移动 M-Space 元宇宙空间

2024 年 10 月，江苏移动联合咪咕打造的国内首个 3D 实时交互的“AI+权益+数字藏品”M-Space 元宇宙空间，聚合“1 个数字家+N 个数字城市街区场馆”创意场景，提供丰富数字权益、精美数字藏品与资产、趣味数字空间互动三大玩法体验，为用户提供一个获取业务资讯、拓展社交互动、聚合权益资产等涵盖多样玩法的娱乐数字空间。



图 4-15 中国移动 M-Space 元宇宙空间

以上娱乐元宇宙的应用案例主要通过游戏平台构建虚拟世界，为用户提供了有别于传统娱乐方式的沉浸式体验。自由的虚拟环境激发了用户内容创新的主动性，让用户直接参与游戏内容的塑造。强大的社交互动使得用户的体验更加丰富多样，而不仅限于游戏发行商提供的游戏内容[330]。此外，很多游戏元宇宙内设有自己的经济系统，玩家可以通过交易虚拟商品或服务来赚取虚拟货币，实现与现实世界价值的转换。

但是，由于技术限制，目前游戏元宇宙仍受限于图形质量、服务器容量和实时互动的技术限制，影响用户体验的连贯性和沉浸感。随着用户在虚拟世界中参与程度的加深，数据安全和个人隐私也成为重大问题。此外，虚拟经济的波动、法律和监管框架的缺位，使用户的资产面临较高风险，**Second Life** 的衰落就是典型案例[329]。中心化元宇宙中虚拟资产的归属问题，也使得用户在面对开发商/运营商时处于弱势地位。

在可预见的将来，增强现实和虚拟现实会发展并完善，伴随更高效的数据处理和通信技术，元宇宙在用户感官方面的沉浸感和互动感会进一步提升。结合区块链技术和去中心化治理架构的发展，将推动更深层次的社会和经济互动，包括新经济模型的建立和全球社交网络的扩展[331]。同时，法规更新和道德准则的建立将确保用户隐私和安全，教育和职业发展也将通过元宇宙获得新的机遇。这些都预示着元宇宙将通过技术创新、社会互动和教育机会为未来社会带来更深远的变革。

4.6 体育元宇宙

当前，元宇宙与体育的深度融合正以前所未有的速度重塑体育体验方式，其发展趋势主要体现在体验沉浸化、赛事虚拟化、空间数字化和服务个性化四个核心维度。通过虚拟现实和增强现实技术，用户可以身临其境地参与虚拟赛事或通过数字孪生技术在元宇宙中体验骑行、滑雪等运动场景；传统体育赛事与电子竞技的结合催生了虚拟体育联赛，例如 PGA、NFL 等机构已布局元宇宙赛事以拓展粉丝经济；区块链和 AI 技术赋能的空间数字化不仅重构了体育消费生态，还衍生出数字藏品、智能场馆管理等新商业模式；而基于用户数据的个性化服务则通过智能硬件和算法推荐，实现了运动方案定制和沉浸式社交互动。这些变革不仅打破了物理空间限制，更以虚实融合的模式推动体育产业向沉浸、开放、智能的方向演进。

(1) 基于 XR 头显的沉浸观赛

咪咕视频 Vision 版基于苹果 Vision Pro 头显，独家支持多赛同看、多视角切换、沉浸式交互操作等功能，通过空间感知技术与场景重建能力，用户可自由组合观赛画面，在现实空间中实现“上帝视角”观赛体验。针对篮球、足球等核心赛事，平台提供战术沙盘等特色功能，极大提升赛事信息获取效率与观赛深度。



图 4-16 咪咕视频 Vision 版沉浸观赛

(2) 数智竞技

咪咕数智竞技科技体验中心是一个集文化、体育、科技、教育于一体的创新场所。依托中国移动 5G、算力网络和分布式云渲染等前沿技术，打破了传统竞技模式，打造了“比特骑行”“火线精英”“钢铁突袭”等自主可控、自由发挥、即时对抗、实时在线的全新竞技体验。

“比特骑行”融合 5G 技术、云游戏技术和增强现实技术，将传统体育竞技数字化、智能化，用户通过大小屏幕和 AR/VR 眼镜的助力完成骑行运动，不仅能够与全国参赛者共同飞驰在竞速赛道之上，还能悠闲漫游于风景名胜之中，实现足不出户完成体能训练的目标。



图 4-17 “比特骑行”体验界面

“火线精英”基于大空间 VR 定位技术和穿戴式 VR 交互竞赛设备，实现参赛队员在数字世界中高强度 PVP 对抗的团队竞技项目，则能够推动传统游戏电竞的体育化、公益化、健康化发展，以创新的玩法和模式，打破市场对于体育竞技的刻板印象。



图 4-18 “火线精英”体验现场

“钢铁突袭”利用 VR+AR+MR、大空间定位，动作捕捉等技术，参赛人员身穿数智装备，可在数实融合的空间中进行对抗竞技，完全实现了科技与体育的完美融合。



图 4-19 “钢铁突袭”体验现场

(3) 沉浸数字健身

为满足全年龄段用户的多元化健身需求，咪咕开发并集成超 300 款互动内容，构建出丰富的数字健身生态。其中《天天八段锦》由北京体育大学专家亲自完成动作捕捉与授课，数智人教练古逸飞提供专业指导，课程结合写实 3D 自然景观与流畅动画，营造出身临其境的沉浸式练习环境，实现数字技术与健康生活的完美融合。



图 4-20 咪咕《天天八段锦》体验界面

4.7 金融元宇宙

元宇宙是互联网、区块链、5G、人工智能等技术发展到现阶段的“集大成者”，从场景、互动到消费，元宇宙在场景上的叙事能力将发挥重要作用。当前，金融机构之间服务和产品同质化严重，在机构纷纷寻求发展各自优势的努力中，元宇宙将成为企业脱颖

而出的突破口，金融业本质上是为资金流通提供服务，而金融元宇宙[332]带来的新的资产内容和流通形式将为其带来新的增长点。

（1）数字员工

目前银行机构的数字员工在国内应用较多，用虚拟数字人技术打造数字员工，在品牌营销、客户服务、业务办理等领域发挥作用。例如百信银行的虚拟数字人艾雅，作为人工智能（AI）虚拟品牌官，在直播、短视频新媒体领域发挥作用，如图 4-21 所示；宁波银行推出 001 号数字员工“小宁”，主要承担大堂经理负责客户服务工作，并且内置知识库，能够回答 550 个常见业务问题。虚拟数字人垂类核心企业 50 余家，主要集中于北京、广东、上海、浙江等地。



图 4-21 百信银行虚拟员工艾雅

（2）虚拟营业厅

客户可以在营业厅中办理业务，获得更加沉浸的体验。例如 2022 年工商银行[333]推出“VR 元宇宙虚拟营业厅”（如图 4-22 所示），以 VR 智能眼镜为媒介，可为用户打造身临其境的沉浸式体验，目前已率先在河北省雄安分行试点使用。



图 4-22 工商银行虚拟营业厅

(3) 数字孪生

在保险元宇宙场景内，可以通过“数字孪生”技术重现应急避险等保险场景，为精准地调节承保费率提供新技术支持。模拟避险模拟出险事故过程及场景，引导应急避险防灾减损；基于“数字孪生”技术的原理，在保险的事前投保阶段，仿真模拟出险事故过程及场景，对保险主体进行教育警示。引导客户进行事前保障及优化自身的内部抗风险管理能力，通过采用应急避险手段和策略，实现防灾减损的共赢目标。风险仿真让客户感知风险，实现沉浸式、场景化营销；元宇宙世界构成采用大量数字化技术，本质上就是一种全真的地球生态仿真。通过对此类数字技术的复用，可以在保险风险仿真领域大量应用，这种沉浸式的场景化体验，极大地丰富了保险营销的内容手段，在人身险、财险、养老险等各个主体保险营销方面，带来革命性的手段变革。

(4) 金融机构营销

金融机构在已有的元宇宙场景中开展品牌宣传、投资教育等活动，达到营销获客、提升品牌体验的目的。比如汇丰银行在沙箱（Sandbox）中开设了虚拟运动场、韩国国民银行在元宇宙平台 Gather 上创建了“虚拟城镇（virtual town）”。元宇宙在营销方面的应用主要是虚拟数字人，虚拟数字人目前主要应用在品牌营销、广告传播、虚拟主播等传媒娱乐领域。虚拟数字人为企业提供服务应用核心在于精准传递品牌价值，扩大品牌销售市场。

（5）真实世界资产代币化

真实世界资产代币化（RWA）是指通过区块链技术将真实世界资产和数字世界资产进行孪生的资产表达形式，代表对实体资产的所有权或权益。狭义 RWA 指基于真实世界资产在数字世界生成可交易的链上权证（Token），广义 RWA 则被视为联通实体与数字资产的价值化新赛道，通过资本要素市场赋能数据等其他要素市场，充分释放资产价值。巡鹰集团旗下巡鹰出行成功完成基于换电实体资产的 RWA，涉及金额数千万港元。该项目通过发行 RWA 基金产品实现资产通证化，借助数字革命重构产业价值链，通过区块链、大数据等技术构建产融结合新生态，以 RWA 模式释放换电实体资产潜能，既为国际资本搭建了可追溯、可验证的绿色投资通道，更在技术攻坚、模式革新、全球资源整合三维度形成突破，驱动低速电动车行业向技术密集型、资本友好型产业跃迁。

元宇宙在金融行业的普及应用还面临着很多挑战性的问题。首先，技术的成熟度和稳定性是一个关键因素。元宇宙需要强大的技术支持，包括虚拟现实、增强现实、区块链等技术，这些技术的成熟度和稳定性直接影响到金融服务的质量和安全性。其次，数据安全和隐私保护是另一个重要的问题。在元宇宙中，用户的大量个人信息和金融数据将被收集和处理。如何确保这些数据的安全，防止数据泄露和滥用，是金融行业必须面对的挑战。再者，监管合规也是一个待解决的问题。元宇宙中的金融活动可能涉及多个国家和地区，不同地区的法律法规可能存在差异，如何制定统一的监管框架，确保金融活动的合规性，需要世界范围更广泛的协作。最后，用户体验和接受度也是一个挑战。元宇宙中的金融服务需要用户适应新的交互方式和技术，提高用户的接受度和满意度，这对元宇宙内容技术架构中的各环节提出了更快、更准、更智能、更易用的综合性能要求。

4.8 供应链元宇宙

2025 年，商务部、国家发展改革委、教育部、工业和信息化部、交通运输部、农业农村部、税务总局、国家数据局等 8 部门联合印发了《加快数智供应链发展专项行动计划》，提出“到 2030 年，形成可复制推广的数智供应链建设和发展模式，在重要产业和关键领域基本建立深度嵌入、智慧高效、自主可控的数智供应链体系”[334]。供应链元宇宙依托交互传感、区块链、人工智能、机器视觉、虚拟现实和增强现实等新一代信息技术，集成供应链各环节量化作业数据，实现供应链运行数据化、模型化、可视化、智

能化，提高分析预测、决策支撑、风险管控能力，降低企业运营成本，提高生产效率。供应链元宇宙已在供应链生产、物流、服务等多个领域展现出其革命性的应用潜力，本章节旨在探讨元宇宙技术在供应链领域的创新应用，代表性应用包括如下三个方面：

（1）供应链智能生产

在数字化浪潮的推动下，内容生产方式正在经历从文字、图片、视频到 3D 信息流的演变。如图 4-23 所示，京东推出全球首个裸眼 3D 商品方案——“京东立影”[335]，率先开启“3D 信息流时代”，为用户带来全新体验，为品牌带来新的增长空间。该技术引入了移轴渲染，增强 3D 展示的立体效果，创造出逼真的立体视差。同时，结合深度剔除技术，在 App 层面实现了“破窗”展示方式，进一步增强了空间感和表现力。用户打开手机京东 App，就能看到扑面而来、360 度全方位展现商品的立体视觉，仿佛在手机空间里搭建了一个三维立体橱窗，享受好玩、好逛、好买的沉浸式购物体验。京东立影联动华为、小米、大疆、LAMER、DIOR、海信、伊利等多个超一线品牌均已经上线全新效果，涵盖 3C 数码、家电、时尚、商超等多个领域。数据显示，SK-II 采用裸眼 3D 广告后品牌点击率提升 60%，华为耳机借助 3D 广告推动销售转化提升 37%，使用 AR 试妆、AR 摆摆看的用户订单转化率分别达 24%、27%。

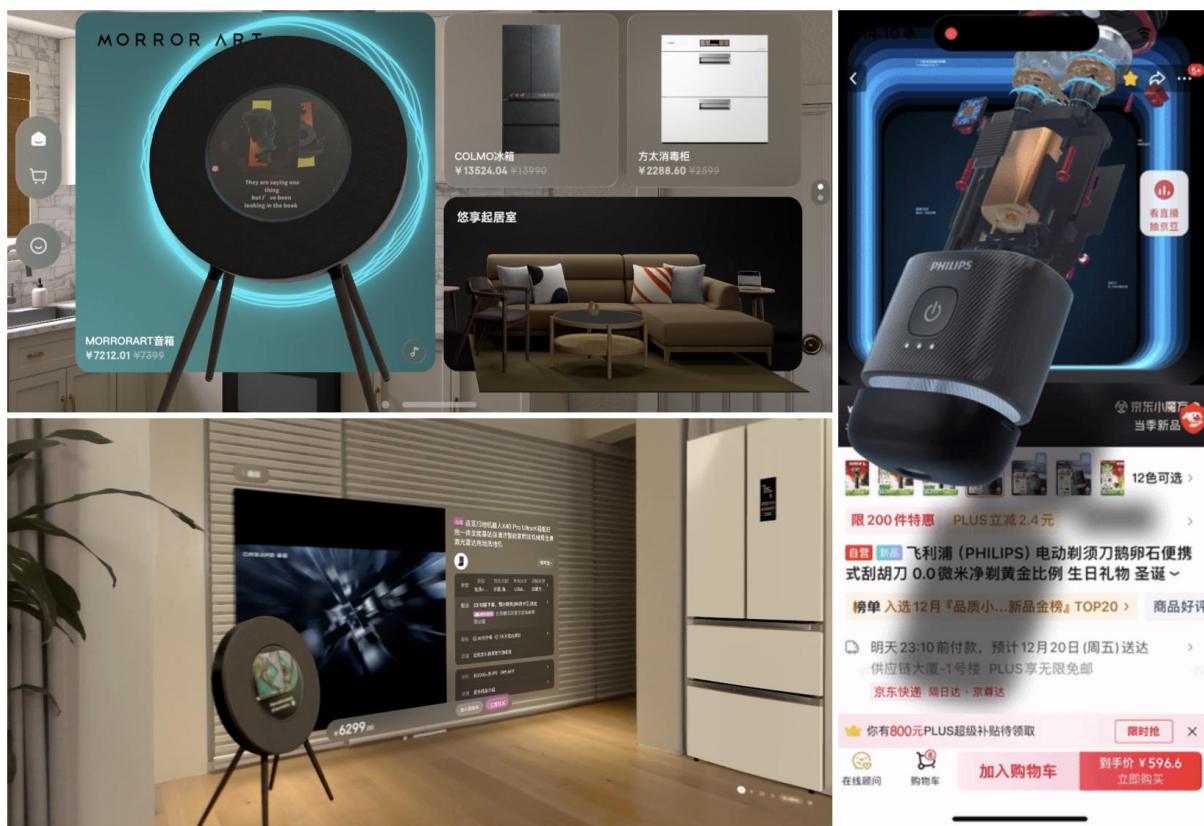


图 4-23 裸眼 3D 商品方案-京东立影

（2）供应链智能物流

物流的本质就是时间和空间管理的系统，元宇宙可以加快物流时效，缩短物流的时间，同时可以使空间利用率达到最高。京东打造了基于大模型的供应链智能物流产品——“京东物流超脑” [336]，为商家、产业带打造高效稳定的供应链支撑。如图 4-24 所示，该产品将大模型与数字孪生技术深度结合，将数据驱动的网络规划模型与仿真模拟结合，实现仓储网络、运输网络、配送网络的智能规划，在平衡成本效率体验的基础上驱动供应链全局最优。在交互层面，用户无需具备专业建模能力，可以直接描绘希望呈现的仓储布局效果，系统将快速生成三维可视化方案，并根据用户描述，进行局部调整。

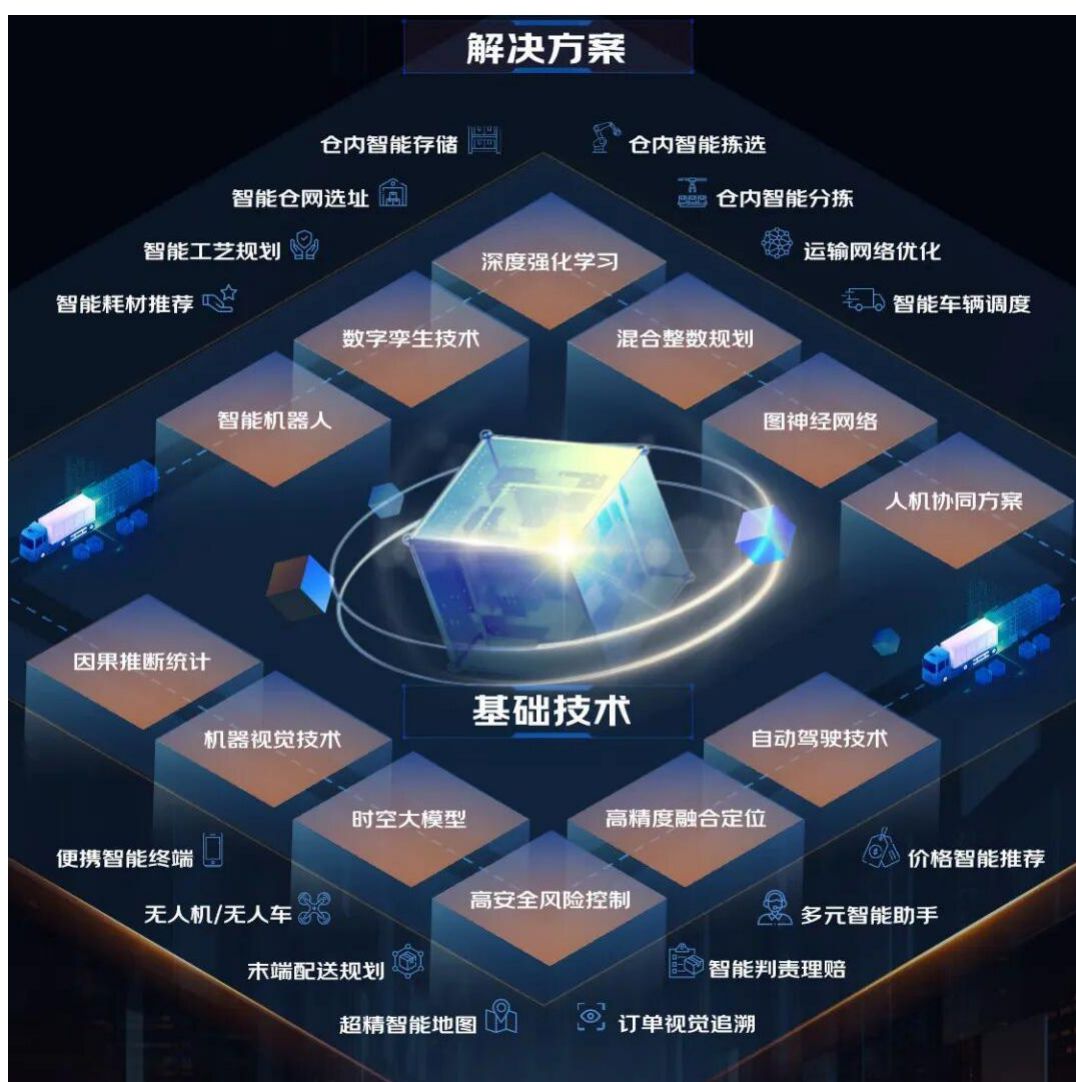


图 4-24 基于大模型的供应链智能物流产品-京东物流超脑

在决策层面，可以高效进行不同布局对比、归因分析和方案推荐，通过大模型分析、理解当前仓储 3D 模型的异常运营问题，给出改善性建议，变被动调整为主动干预，显著提升运营效率。京东物流超脑已覆盖近百个应用场景，累计调用 19 亿次，分拣环节

的日均降低小件操作 650 万次，使拣货效率、一线员工效率均提升 10%，不仅实现 95% 的自营订单、87% 的乡镇街道在 24 小时内送达，更让近百座城市的消费者享受到最快“分钟级”收货服务，对于降低全社会物流成本，科技驱动行业高质量发展有重要意义。

（3）供应链智能服务

虚拟数字人是元宇宙不可缺少的一部分，它不仅是元宇宙的原住民，还是自然人在元宇宙中的数字化身，是我们在元宇宙中自我呈现与形成自我认同的重要方式，是人和人、人和物的交互载体。如图 4-25 所示，京东打造了多模态交互式数字人平台，通过语音合成大模型和通用数字人大模型，仅需输入自然语言或 1 秒视频/照片，即可生成 4K 超高清、100% 口型匹配的数字人，彻底告别数字人制作对线下拍摄、建模与耗时训练的依赖。京东基于该平台为供应链服务场景打造数字人主播，首次实现了数字人大规模产业化应用，实现 7×24 小时不间断直播带货，帮助商家把直播成本降低至真人的十分之一，带货效果超越 80% 真人主播，平均提升直播间转化率 30%。平台已服务超 4 万家品牌商，累计直播时长超 500 万小时，2025 年双 11 期间带动 GMV 超 23 亿元[337]。京东数字人项目还荣获 2024 年度吴文俊人工智能科学技术奖科技进步特等奖。



图 4-25 京东多模态交互式数字人平台

机器人是元宇宙虚实联动的载体。京东打造了 JoyInside 贴身智能平台，具备领先的产业应用能力、高情商对话体验、多元场景适配、海量角色选择四大优势，将角色大模型驱动的对话智能体植入智能硬件载体，如机器人、机器狗、玩具等，为硬件提供“大脑”能力，赋予“灵魂”和“智慧”，提升产品竞争力和用户粘性。平台目前已接入众擎、云深处、Fuzozo、元萝卜、凯叔讲故事等 30 多个品牌[338]。

元宇宙技术已经成为一种改造社会的生产力，从降本增效逐渐辐射到生产关系与生产管理的改造，从一个产业逐渐辐射到多个产业，乃至整个社会，推动着供应链的生产、物流、服务体系一体化发展，促进社会秩序的变革，建立了高效、和谐和公平的新型社会秩序。展望未来，我们期待供应链元宇宙能够为构建国内大循环为主体、国内国际双

循环相互促进的新发展格局提供更加有力的支撑。

4.9 农业元宇宙

农业元宇宙利用环境感知物联网与空间计算技术构建了一个虚拟孪生世界，基于 3DGS 数字孪生农场整合现实数据。AI 育种平台模拟基因型表达加速理想作物设计，智能体分析数据提供最优决策并指挥智能机器人执行任务，形成“感知-映射-模拟-决策-执行”的闭环。这将农业生产从经验转向科学实验，结合新兴技术催生新业态，解决劳动力老龄化、可持续发展及气候变化等问题，尤其在中国推动从经验到数据驱动的转变[339][340]。

核心技术包括：AI 作为“大脑”，通过多源数据分析实现智能决策，支持作物管理与可持续发展；智能体扮演“AI 管家”，自主生成农事处方调度执行，实现闭环管理；3DGS 技术高保真映射物理农场，提供精准可视数字化底座，模拟不同灌溉策略对作物生长的影响；具身智能技术赋予农业机器人更强的感知和作业能力；空间计算技术关联数字与物理世界，提升作业精度效率。这些技术共同促进了农业全流程智能化管理[339][341][342][343]。

在海外和国内，已经有多个将前面提到的智能机器人，AI，3DGS 等技术应用到农业领域的成果和项目，其中有学术界的研究，也有企业界的案例，这些项目从不同的角度展现农业元宇宙技术的未来发展趋势，也展示了技术为农业带来的推动与变革。下面就是一些典型项目介绍。

（1）AGROS 项目与 ROBS4CROP 农业机器人系统

在瓦赫宁根大学牵头的 AGROS 项目研究中，智能算法结合基于模型的数字孪生，可以实现智能温室控制，让自主种植变得更智能[344]。多种技术共同构建起农业元宇宙的完整技术生态，为智慧农业的发展提供全方位支撑。瓦赫宁根大学牵头研发 ROBS4CROPS 农业机器人系统，构建了一个由智能工具、自动驾驶车辆和规划装置三个要素组成的机器人系统。现有的农业工具和拖拉机经过升级，能够与现有的农业机器人协同工作，成为机器人系统的一部分（减少更换所需的投资）。这个系统在实际操作中得到了测试和展示，涵盖了种植园农业（荷兰）、大田蔬菜种植（法国）、葡萄栽培（德国）和水果种植（西班牙）等领域[345]。

（2）nexeed 设施农业 AI 生产力平台

摩尔线程和埃舍尔科技共研的 nexeed 设施农业 AI 生产力平台，通过融合尖端 AI 与机器人技术，为全球高价值作物种植者提供从感知、决策到执行的端到端智能解决方

案，通过三维可视与取证完成“看得见”、多模态识别与预测实现“算得准”、边缘闭环与机器人协同助力“执行快”、A/B 验证与知识沉淀达到“可复盘”，构建“感知-决策-执行-复盘”业务闭环，实现生产力的指数级跃升和农业的可持续发展。



图 4-26 nexeed 数字花卉智能 AI 温室

(3) 弥勒市新哨镇韭黄农业全产业链数字平台

云南省弥勒市新哨镇建设了韭黄农业全产业链数字平台，通过数据为决策层提供信息支撑；全流程追溯管理系统则确保生产信息可追溯、可信任、可验证；再结合 AI 数据分析，持续优化产业数据资产；并构建基于溯源体系的特色联农带农服务体系，借助数字化手段增强农业产业效益，推进乡村振兴战略的实施。



图 4-27 韭黄农全产业链数字平台

农业元宇宙的应用对推动农业现代化转型与可持续发展具有深远意义：通过数据驱动的精准管理，大幅提升资源利用效率与劳动生产率；通过优化资源利用与减少化学品投入，助力农业可持续发展；通过模拟预测与前瞻规划，增强农业经营韧性。

农业元宇宙代表着农业数字化的前沿方向，通过多项核心技术的协同创新与深度融合，正推动农业生产方式与价值链的深刻变革，不仅为应对当前农业面临的各类挑战提供了创新解决方案，也为构建更加透明、高效、可持续的全球农业食品系统提供了全新可能。面对技术成本、数据安全等挑战，需政府、产业界与科研机构协同努力，通过技术创新、标准制定与人才培养，共同构建高效、透明、可持续的智慧农业新生态。

第5章 元宇宙安全治理：挑战与策略

元宇宙作为未来的新兴数字娱乐世界、生活世界乃至产业发展新空间，在虚实孪生乃至虚实融生的过程中，也创造了新型的安全问题，尤其是数据安全、网络安全和国家安全等方面的安全风险。因此，元宇宙空间也需要建立规范的公共秩序，不应违背公序良俗，在遵守现行伦理道德标准的基础上，探索法律规范创新，加强数字内容风控，确保数字公正，促进技术向善，创设与元宇宙相匹配的“元规则”[346]。

5.1 元宇宙的安全风险

5.1.1 数据安全与治理

随着元宇宙时代的到来，在线社交、游戏、媒体、医疗、教育、工作、生活等诸多元宇宙场景需要大量采集、存储、分析个人数据（信息），且海量数据具有迭代性和不可删除的特性，其数据规模，信息完整性是前所未有的，因此也伴生着前所未有的数据安全挑战[347]。元宇宙中涉及的三维建模、三维显示、三维音频以及体感交互等技术能够全方位地收集人体生物信息，包括眼球位置信息、肌电图、脑电图、心电图等，个人隐私极易被毫无保留地暴露在元宇宙平台的视野中，用户的隐私安全也会受到严重的威胁，网络暴力和安全问题的形式将花样百出，规模和危害程度远高于现有形式[348][349]。同时，元宇宙中部署的大规模人工智能模型在训练过程中可能接触到大量敏感信息，甚至会有国家保密信息，存在泄露保密数据的风险。另外，数据垄断是元宇宙治理的新型隐忧。数据被认为是数字时代的“石油”，但数据的传播效率远远大于石油，对于数据资源的管控已经不能使用传统工业时代管理有形资产的方式。存在互联网平台为提升收益垄断用户数据并提高进入壁垒及转换成本的风险，“赢者通吃”的局面将损害用户个人利益和社会公共利益[350]。因此在元宇宙距离生活愈发切近的背景下，强化隐私保护和数据加密技术发展、建立完善的数据安全和隐私保护制度尤为关键。

5.1.2 网络安全

元宇宙作为虚拟互动平台，面临多种网络攻击风险，包括钓鱼攻击、中间人攻击、分布式拒绝服务（DDoS）攻击、恶意软件以及身份盗窃等。这些攻击不仅可能导致用户

经济损失、个人隐私泄露和虚拟身份被盗，还可能引起服务中断，削弱用户对元宇宙平台的信任。特别是通过智能合约漏洞进行的攻击，可以直接影响区块链上的资产和合约执行。智能合约是一种运行在区块链上的程序，它自动执行、控制或记录合同条款的履行，但也可能存在漏洞，这些漏洞源于编码错误、逻辑漏洞或设计不足。智能合约攻击是指恶意用户利用智能合约中的漏洞来进行未经授权的操作，通常的目的是窃取资金、操纵合约行为或破坏合约的正常运行。这些攻击利用合约代码的弱点，通过精心构造的交易来触发这些漏洞。例如，著名的“DAO 攻击”就是通过重入攻击漏洞，攻击者能够重复提取 DAO 合约中的资金，最终导致数百万美元的以太币被盗。

5.1.3 社会安全

元宇宙可能带来就业替代和社会结构的深刻变化，导致传统行业人力需求减少、虚拟空间犯罪、游戏成瘾乃至虚实混淆等问题。一方面，元宇宙将催生出新的数字创意岗位，特别是推动数字创意者成为自由职业者，推动数字经济小微企业的发展[351]。同时，元宇宙可能带来一些新的社会问题。近年来以“元宇宙”名义金融诈骗的方法层出不穷，有编造虚假元宇宙投资项目、恶意炒作元宇宙房地产圈钱、恶意炒作元宇宙房地产圈钱等。此外，元宇宙逼真的沉浸感和超低的延迟体验带来了更高的“成瘾性”，更容易造成青少年过度沉迷，对其成长带来不利影响。尤其是在元宇宙中虚拟互动更加激烈和真实，甚至直接传导连通用户大脑感知，相关技术容易被不法分子用于制造带有成瘾性的“数字毒品”，预防或阻断上述成瘾情况，其挑战则更加艰巨。元宇宙具有让用户长期沉浸其中的“吸引力”，使用户认知和行为可能与现实世界的人们脱节，带来明显的代际差异。

5.1.4 国家安全

元宇宙作为互联网、人工智能技术迭代的新大陆，也可能逐步成为国际意识形态竞争的角斗场，直接关系到国家的精神文化和意识形态安全。美国社会学家卡斯特在描述未来信息社会时认为：“信息时代的权力战斗是文化的战斗，媒体是他们主要的战场和战斗工具”[352]。元宇宙凭借沉浸式叙事和切身体验特性，能够打破时空的限制，成为敌对势力开展信息战、欺骗战和心理战的潜在场所。它能够通过推波助澜热点事件，引导和煽动群体性事件，进而可能演变为线上线下结合的“颜色革命”，对现实社会造成实质性破坏[353]。另一方面，随着元宇宙具有一定的隐蔽性和虚拟性，社会交往成本降低，

个体价值追求会被充分激发和深度挖掘，各种形式的社团、组织和党派可以频繁交流，刻意构成信息茧房，加深人们的认知偏差，甚至控制群体成员的精神和意识，让人们听到不到也看不到外界的信息，元宇宙就可能成为极端主义的温床，进一步加剧社会意识形态系统的潜在风险，从客观上增加了元宇宙空间意识形态安全治理的不确定性[354]。元宇宙中的意识形态治理需要未雨绸缪，重视潜在危机。

5.2 元宇宙安全治理的现实挑战

5.2.1 数字平台的安全漏洞

（1）技术架构的安全隐患

人工智能的神经网络算法基于黑盒模型，因此缺乏可解释性，极易受到对抗攻击。“对抗攻击”指的是通过故意制造输入数据的微小扰动，使得 AI 模型做出错误的判断，在图像识别、语音识别等领域中较为常见。例如，对抗性图像可能使自动驾驶汽车错误识别交通标志，导致安全事故。由于近年来人工智能算法、模型、应用发展演化速度较快，如何判断人工智能是否具备可解释性仍缺乏统一认知，难以形成统一判别标准。

人工智能模型的训练和交互过程存在大量数据泄露的风险点。如果训练数据中包含敏感信息，且这些数据未经适当处理，就可能导致数据泄露。一方面，随着人工智能模型的增大与开发过程的复杂化，数据泄露的风险也随之增多，且很难完全消除。另一方面，交互式人工智能的广泛应用使得数据流向模型的过程变得更加简单。用户在与 ChatGPT 等人工智能对话模型交互时，往往容易放下警惕，有可能不经意间泄露个人隐私、商业机密或科研成果[355]。例如，员工在使用人工智能辅助办公时，可能会输入涉及企业机密的信息，从而导致重要数据的泄露。

（2）元宇宙犯罪的新形态

元宇宙技术催生了全新的犯罪形态，这些犯罪可能涉及人身安全、财产权益和金融市场的稳定，将使色情赌博、网络传销、非法集资、电信诈骗、黑灰产等传统网络犯罪活动逐步渗透至虚拟世界，甚至可能会催生新的网络犯罪类型。新犯罪形态具有隐蔽性强、成本低和破坏性大的特点。数字身份和数字资产盗取可能是未来元宇宙主要犯罪形式之一[356]。数字身份盗窃涉及非法获取和使用他人的个人信息进行网络身份冒充，可能导致信誉损害和财产损失[357]。数字财产盗窃则是指通过黑客攻击、诈骗等手段非法

占有他人在网络上的虚拟资产，如元宇宙装备、虚拟货币。这类犯罪不仅侵犯个人隐私和财产安全，还可能引发进一步的网络诈骗和经济犯罪，严重时还会影响到整个网络经济的信任体系。Meta 的一款 VR 游戏《地平线世界》（Horizon Worlds）的线上性侵事件表明，元宇宙中也将可能发生虚拟性骚扰和暴力事件，由于元宇宙的匿名性和沉浸感，某些用户可能在元宇宙平台上进行某种形式的性暗示，受害者可能会感到极大的不适和心理压力，这种影响有时甚至会延伸到现实生活中。

5.2.2 数字资产的安全保障

在元宇宙背景下，数字资产的安全保障面临诸多挑战，尤其是在数字货币交易和区块链技术的应用方面。数字货币的匿名性和不可逆性使其容易被用于洗钱和诈骗，同时，区块链尽管被视为安全的技术，却可能因智能合约中的漏洞而遭受攻击。此外，数字资产如加密货币的钱包密钥一旦丢失或被盗，资产的恢复几乎不可能实现。面对这些风险，元宇宙的治理需求较为迫切，包括加强技术安全性，如对智能合约进行严格的安全审计；建立健全的监管框架，比如国际合作制定跨境交易的规范；提高用户的安全意识，传播教育安全存储私钥和识别钓鱼网站的应对策略等。通过这些措施，可以提升元宇宙数字资产的安全级别，保障用户的资产安全，推动元宇宙健康可持续发展。

5.2.3 国际合作与全球治理

（1）跨境数据流动的安全

借助去中心化和随地特征，元宇宙相关技术和业务生态将以全球化模式发展，将不可避免地涉及数据跨境问题，对数据安全保障带来严峻挑战。不同国家和地区的数据保护法规存在差异，若缺乏统一的标准和监管，会出现隐私泄露、数据滥用、跨境犯罪等风险，同时带来数据主权与管制、治理趋于区域化、监管和执法难等挑战。首先，许多国家对数据的管控成为国家治理的重要内容，将现实地缘政治之间的博弈映射至数据治理领域，如何收集、存储、保护、使用数据正成为突出的地缘政治问题。其次，法律规范缺乏全球标准和框架。跨境数据流动治理中最重要的影响因素莫过于各国传统政治经济利益划分所连带的路径依赖与权力外溢，国家安全与稳定、社会价值实现与隐私保护、企业的经济效益，如何在跨境数据治理的复杂性和多元化下协同发展。最后，对数据的

监管与执法不可能离开技术的支撑。典型的例子就是数据流动溯源，尽管采用区块链、人工智能等安全算法号称能有力保障数据要素安全流动，但是数据监管和执法目前仍面临难题[358][359]。

（2）国际规范的协调与制定

元宇宙监管的国际规范制定面临着主权、技术、法律、利益相关方、监管能力、道德伦理等多重挑战。各国关于元宇宙的不同规则使技术本身和相关技术人员受到不同国家法律的约束，对国家造成了司法管辖权上的挑战。一是主权和管辖权的冲突。不同国家和地区对于数据主权、网络主权等概念存在分歧，各方可能会基于国家安全、经济利益等诉求，在制定国际规范时存在利益冲突和博弈。这可能导致规范缺乏广泛共识，难以达成一致。二是技术标准的不统一。元宇宙涉及多种前沿技术，如虚拟现实、区块链等，各国在这些技术的发展阶段和应用方向存在差异。要在国际层面统一技术标准和协议极具挑战性。三是法律体系的差异。不同国家和地区的法律体系存在差异，在制定跨境规范时可能面临法律冲突和监管碎片化的问题。如何在不同的法律框架下寻求平衡和协调是一大挑战。四是监管存在局限性。元宇宙的虚拟性和跨境性给监管带来了很大挑战，许多国家的监管体系和能力还无法完全应对元宇宙带来的新问题。如何提升监管能力，建立有效的跨国执法合作机制也是关键。五是道德伦理问题的复杂性。元宇宙涉及隐私保护、数字人权保护等伦理问题，不同文化背景下的价值观和道德标准存在差异，在国际层面达成共识可能存在难度。

5.3 元宇宙安全治理的策略与建议

5.3.1 技术层面的创新与应用

加强关键技术研发与安全测试，提升底层平台的自主可控性与稳健性。 坚持自主创新和开放融合并重，激发我国元宇宙产业的技术层面高质量发展。加强关键产业技术研发，建立覆盖元宇宙各关键技术领域的安全标准体系，包括网络安全、数据安全、虚拟资产管理、人机交互安全等多个方面。建立全生命周期的安全测试机制，涵盖元宇宙技术研发、产品设计、系统部署等全生命周期，包括代码审计、渗透测试、压力测试等多种手段，确保安全问题在早期被发现和修复。培养专业的安全人才队伍：大力培养具备元宇宙安全专业知识和实践经验的人才，包括网络安全工程师、数据安全分析师、虚

拟资产管理专家等。建立健全的职业培养体系和人才激励机制。加强用户安全意识教育：提高公众对元宇宙安全风险的认识，培养用户的安全意识和自我保护能力，减少因用户误操作而导致的安全事故。政府、企业应该共同开展安全教育宣传活动。

强化数据加密与访问控制体系，完善身份认证、权限管理与隐私保护机制。数据安全在元宇宙产业生态中承受着多方面的风险，需要通过不断完善法律法规，加大对市场操纵等恶性违法违规行为的打击力度，建立能够应对不确定性和管控随机风险的一体化监管制度，加速弥合数字鸿沟，提升数据加密与访问控制水平。可建立多因素认证机制，针对元宇宙中的各类用户身份，应该采用密码、生物特征、硬件令牌等多种身份验证因素相结合的多因素认证机制，限制数据的收集、使用和共享范围，大幅提升身份认证的安全性。建立动态访问控制机制：根据用户、设备、环境等因素，实施动态的权限分配和访问控制，确保只有经过授权的主体才能访问相应的数据和功能。同时要定期评估和调整访问控制策略。将数据按照敏感程度进行分层存储，对于高敏感数据采取更加严格的加密和访问控制措施，降低数据泄露的风险。同时要确保不同层级数据之间的安全隔离。建立数据溯源机制：对于元宇宙中的各类数据，建立全面的溯源机制，记录数据的来源、流向、操作等信息，以便于事后的安全审计和责任追究。定期对元宇宙中的数据加密和访问控制机制进行安全测试和审计，发现并及时修复存在的漏洞和隐患，持续提升安全防护水平。

5.3.2 法律层面的完善与实施

健全数据治理相关法律法规，强化数据采集、加工、共享与跨境流动的规范化管理。由于近年人工智能的飞跃式发展，元宇宙所涉及的数据安全问题已有相关法律法规进行规约。如中国发布的《数据安全法》、《个人信息保护法》，欧盟《一般数据保护条例（GDPR）》，新加坡《个人数据保护法》等，对数据的收集、存储、使用、共享等环节提出了明确的安全要求，要求企业采取加密、访问控制等技术手段，保护数据安全。同时，要求企业在使用个人数据训练算法时，必须征得用户授权，并采取匿名化等措施保护个人隐私。在数据跨境流动方面，中国《数据安全法》中的数据出境安全评估制度，欧盟 GDPR 中的跨境数据传输规则等法规旨在保护数据安全的前提下，促进数据的有序共享和流动。一些行业也出台了专门的数据管理法规，如金融业的《个人金融信息保护技术规范》、医疗健康领域的《医疗机构数据安全管理办法》等，对行业数据的安全和

共享做出具体规定。数据安全和数据共享的相关法律法规体系正在不断健全和完善，为数据资产的安全管理和有序流动提供重要法律依据。一些国家针对互联网平台算法的垄断行为，出台了相关的反垄断法规，如中国《反垄断法》、美国《数字市场竞争法案》等，旨在防止算法滥用导致的市场垄断。

完善虚拟资产与知识产权保护框架，明确数字内容的权属、交易和使用规则。在元宇宙中，虚拟资产作为一种非实物化的数字资产，在确权和交易保护方面将会面临更加复杂的法律问题。一方面，元宇宙中的虚拟资产是归属于个人用户还是服务提供平台目前缺乏法律上的界定。另一方面，元宇宙中的虚拟资产是否受法律保护尚不明确，相关法案对虚拟资产保护的观点尚未统一。虽然《民法典》第 127 条规定：“法律对数据、网络虚拟财产的保护有规定的”，最高人民法院联合国家发展和改革委员会共同发布的《关于为新时代加快完善社会主义市场经济体制提供司法服务和保障的意见》中强调，要“加强对数字货币、网络虚拟财产、数据等新型权益的保护，充分发挥司法裁判对产权保护的价值引领作用”等，但目前尚未出台专门针对虚拟资产交易保护的法律法规。

元宇宙中的知识产权问题主要涉及著作权、商标权侵权等。著作权侵权的场景主要包括：未经著作权人授权将其著作改编为元宇宙影视或游戏；未经著作权人授权私自引用其文学、音乐、影视等作品内容。以国内“元宇宙侵权第一案”为例，2022 年，有用户在原与宙公司旗下的元宇宙平台上发布了《胖虎打疫苗》作品，虽然该用户是在平台上最早发布的人，但并非作品原创者，他还在该平台上向其他用户收取 899 元的费用，于是《胖虎打疫苗》的版权所有公司奇策将原与宙公司告上法庭。中国的《著作权法》、美国的《数字千年版权法》等，初步针对人工智能等数字原创产品做出了说明，明确了数字版权的范畴和保护措施。但元宇宙中面临的知识产权问题能否完全适用于现有法律法规还需与时俱进的研究分析。

构建适应虚拟交互场景的社会风险治理机制，加强对欺诈、成瘾、虚假信息等新型风险的预防与处置。面对元宇宙可能出现的新型诈骗等风险，2022 年 2 月，我国银保监会发布了《关于防范以“元宇宙”名义进行非法集资的风险提示》。关于元宇宙的反垄断与公平竞争，各国出台了相关法规，如美国的《数字市场竞争法案》、中国的《反垄断法》等，要求平台企业确保公平公正的竞争环境。网络安全管理方面，中国的《网络安全法》、美国的《网络安全改进法案》等法律法规的完善，要求企业采取技术措施保护元宇宙系统和网络的安全。中国的《消费者权益保护法》，要求元宇宙企业尊重用户隐私，保障用户合法权益。

第6章 总结与展望

随着时间推移，元宇宙的市场热度逐渐回落，产业态势也趋于理性，各方普遍达成了“元宇宙是一场长期演进”的共识。正如扎克伯格所强调，元宇宙更像一个时间概念，是一种“奇点时刻”的未来形态，只有在软硬件基础设施、技术体系与应用生态全面成熟后，才能真正实现虚实融合的深度交互。在此之前，产业的重点仍将是底层技术能力、核心设备、行业应用与创新产品的持续孵化。当前，全球元宇宙产业正处于加速演进阶段，其发展将为制造、教育、医疗、艺术、游戏等众多行业释放巨大增长空间。这一趋势也对科技研发与人才培养、产业生态构建、治理机制与法律体系等方面提出更高要求。只有在制度、技术与产业等层面协同发力，构建与元宇宙特征相匹配的发展理念与治理体系，才能抓住智能经济与第四次工业革命带来的重大机遇，充分发挥元宇宙对经济社会发展的赋能效应。

展望未来，在技术迭代、市场扩容与应用需求的共同作用下，元宇宙将持续演进，新技术、新产品、新模式与新业态也将在这一概念框架下不断涌现。从产业和资本市场的动态来看，元宇宙赛道正逐渐成型，一批具备中短期落地价值、长期想象空间的领域将加速成长，有望孕育出一批“专精特新”的小巨人企业。同时，产业龙头的持续投入以及政府与资本的协同支持，将进一步增强对消费型元宇宙、工业元宇宙等重点业态发展的信心，推动元宇宙产业迈向更加健康、稳健与可持续的新阶段。

6.1 元宇宙技术总结与展望

元宇宙技术体系将在“计算、感知、生成、协同、交互”五大核心能力的带动下持续演进，并从早期的分散式探索迈向软硬件深度协同的体系化阶段。随着人工智能成为元宇宙的关键动力源，虚拟空间的构建方式、运行机制与交互逻辑将迎来系统性重构，技术路径也将由“功能叠加”加速转向“智能协同”，呈现出算力网络化、感知全域化、生成智能化、协同可信化与交互自然化的总体趋势。与此同时，软硬件一体化将成为未来发展的基础支撑：终端设备、传感模组和渲染硬件的持续迭代进一步提升虚拟空间的沉浸质量与运行稳定性，而算法模型、生成引擎与场景编排体系的快速演化则形成更强韧的能力底座。随着这一体系不断成熟，元宇宙将逐步从沉浸式体验入口演化为支撑数字经济的智能基础设施。

元宇宙计算技术的持续革新将成为支撑元宇宙规模化发展的关键牵引力。依托分布式并行计算架构、异构资源调度机制与低时延网络体系，元宇宙的计算底座正具备处理海量并发与复杂逻辑的能力，并从以图形渲染为中心的单一负载模式，迈向“渲染—智能—仿真”深度融合的混合计算范式。依托于强大的分布式计算底座，元宇宙数字空间涌现出四大核心计算场景：首先是图形渲染，云边端协同架构结合 3D 高斯泼溅技术，突破了实时渲染的画质与效率瓶颈；其次是人工智能，从大模型到智能体与世界模型的跃迁，重塑了元宇宙的智能化内核与生成能力；再次是科学计算，AI 与物理模拟的深度融合正在革新虚拟世界的仿真范式；同时，超高清流媒体与以 AI 眼镜为代表的端侧交互计算，则进一步确立了以沉浸式体验为中心的算力流向。这些典型场景不仅验证了分布式并行计算架构的有效性，也反向推动了底层算力网络向更低延迟、更高智能、更强协同的方向持续演进。

元宇宙感知技术将从单一模态的精准采集迈向全模态的统一理解，形成由光场、声场、触觉场与行为场协同驱动的立体化感知体系。依托可微光学、神经重建与多模态大模型，感知系统将摆脱传统逐点采集，通过异构传感网络、算摄一体化以及光电混合计算等新型架构实现跨尺度、跨维度的实时数字化，并在采集过程中同步完成场景重建、语义分析与行为预测，使内容生成从被动记录迈向主动理解与智能构建。通用大模型将统一视觉、音频、触觉与动作语义的表达方式，使感知系统能够以更接近人类认知的方式解读空间、行为与情绪。同时，轻量化边缘推理、时空同步采集与隐私保护技术的成熟，将进一步推动其在真实场景中的规模化部署，为数字孪生、沉浸交互、具身智能和复杂系统模拟提供高可信输入。

元宇宙生成技术如今已取得了量变到质变的突破，多模态生成技术能够从文本、图像或视频输入中直接生成高保真的三维模型和场景；三维模型生成质量在细节纹理和几何精度上显著提升，支持动态光影效果和物理属性模拟；AI 驱动的内容生成工具开始集成低代码平台，使非专业用户也能通过自然语言描述快速创建三维场景，大幅降低创作门槛；高斯泼溅技术通过离散点云优化和可微渲染，在保持高真实感的同时，将渲染速度提升至实时水平，已陆续在空间视频、文旅、工业仿真等场景应用落地。世界模型作为元宇宙生成新范式正在快速兴起，它致力于让 AI 系统不再仅仅是处理信息的工具，而是能够在内部构建一个对现实世界的模拟，从而像人类一样进行“思考”和“预演”的主动参与者。在世界模型的驱动下，生成技术不再局限于产生静态的文本、图像或视频，而是能够创造出符合物理规律、具有时空一致性的动态交互环境。这意味着人们有

望在不久的将来按需生成复杂的、可探索的虚拟世界，为游戏、影视制作、元宇宙应用等带来革命性变化。

元宇宙协同技术正加速迈向规模化应用阶段，为元宇宙构建可信、透明、可持续的制度基础。目前，多个关键方向出现突破性进展：**Layer 2** 扩展方案正进入大规模部署阶段，基于 **zkEVM** 的 **Rollup** 网络使交易成本下降 **90%** 以上；模块化区块链架构从概念走向落地，数据可用性层与执行层的解耦提升了技术栈组合的灵活性；**AI** 代理与 **DAO** 的深度融合进入验证期，智能合约自动化审计与提案分析工具率先投入使用；零知识证明技术在身份认证与隐私保护场景取得突破，使用户能够在不暴露敏感信息的前提下完成跨平台可信交互；现实世界资产（**RWA**）代币化进程加速，合规框架的逐步完善推动传统资产安全融入元宇宙经济体系；后量子密码算法也开始试点部署，为应对未来量子计算威胁提供前瞻保障。这些技术进展将推动虚实融合从概念验证迈向产业化落地，为元宇宙协同体系的长期演进奠定坚实基础。

元宇宙交互技术将从界面驱动迈向意图驱动，实现由“人适应机器”向“机器适应人”的关键跃升。未来的交互体系将依托多模态融合构建更自然的感知—理解—响应链路，通过大模型提升语音、视觉、动作等多模信号的语义解析、情境推理与行为预测能力，使交互方式从操作式输入转向自然化、具身化的意图表达。随着近眼显示设备向轻量化、大视场角和视网膜级分辨率演进，沉浸显示将突破光场与全息技术瓶颈，有效缓解视觉辐辏冲突并提升沉浸舒适性。交互智能也将与具身情感计算深度融合，实现可解释的情绪识别、意图理解与主动协作，高精度、可穿戴的触觉反馈设备将进一步强化沉浸体验与真实在场感。总体来看，交互技术的演进将推动虚拟空间从视觉呈现场景转向具备情感感知、任务协同与认知推理能力的智能环境，使元宇宙成为人类认知、表达与创造能力的自然延伸，并逐步融入生产生活与社会运行体系，为构建自然、可信、低门槛的虚实融合体验提供核心技术支撑。

总体而言，元宇宙正由技术驱动阶段迈向平台化、系统化和智能化发展新阶段。计算、感知、生成、协同与交互五大核心能力的持续融合，将推动元宇宙从沉浸式体验空间演进为支撑产业升级的核心基础设施。随着行业应用持续深入，元宇宙有望成为数字经济高质量发展的重要引擎，并将在未来的经济建设、文化生产与社会治理中发挥日益突出的作用，并逐步成长为贯通物理世界与数字世界的智能化基础设施。

6.2 多方协同机制建设与元宇宙产业高质量发展

推动“专精特新”企业的培育与发展，打造行业领军者。在元宇宙领域，中小型企业占据了主要地位。因此，必须充分利用中小企业数量众多的优势，打造一批具备高度专业化、精细化、特色化和突出创新能力的中小企业，以此作为元宇宙产业做大做强的核心路径。为了达成这一目标，需要整合现有的元宇宙企业资源，准确识别每个企业的优势和薄弱环节。中小企业应该在各自的细分领域中深耕，致力于提升技术和市场资源的积累，以实现“补短板”、“锻长板”和“填空白”的目标。通过专注于各自的优势产品，强化产业链的构建，为元宇宙产业中的中小企业提供更多机会，使其逐步成长为专精特新企业。这些中小企业应持续加大创新投入力度，深化在单一细分领域的专业知识，为元宇宙产业链上游企业提供更加优质和精良的产品与服务。在优势产品领域逐步形成明显的竞争优势和规模经济效应，为大型企业和项目提供关键零部件、元器件及配套产品。努力将这些企业培育成掌握独特技术的“单打冠军”或在特定领域提供全套解决方案的“配套专家”。

进行深入且实际的产业链规划与布局。政府部门需绘制元宇宙产业链的详细图谱与发展规划蓝图，并持续更新产业链分布图、全景透视图以及招商指南图等，以确保全面、动态地掌握产业链现状。实施创新的头部企业培育计划——“揭榜挂帅”行动。在这一行动中，政府将挑选并推荐领军企业作为特定产业链的领航者，并根据企业的实际需求与特性，制定明确的发展目标，提供个性化的扶持措施。同时，优先满足这些链主企业在创新、人才、土地和融资等方面的需求，确保其在产业链中的核心地位稳固。为了进一步加强产业链的整合与协同，需要支持这些链主型企业通过产品带动或资本运作，联合上下游企业、研究机构等建立产业链共同体。此外，需要拓展上下游对接的平台和渠道，加强产业链各环节的紧密合作，提升本地产业的协同配合能力。

重视人才及团队的引进与培养。依托高层次人才、领军人才等政策，积极引进国内外元宇宙领域的高精尖人才与技能型人才，并支持这些人才在元宇宙领域创新创业。对于掌握国际领先核心技术或拥有自主知识产权的元宇宙相关人才团队，对其成果转化项目给予倾斜性扶持，并为高管和骨干技术人才提供连续多年的人才补贴，补贴金额最高不超过其上一年度对经济的贡献。此外，对于元宇宙产业的领军人才，不仅要提供创业资助与直接股权投资支持，还需为符合条件的杰出人才、优秀人才、精英人才提供不同档次的人才租房、购房补助，并在子女教育、医疗保障、养老等方面给予全方位的支持。

加强技术创新激励。对于元宇宙企业增加研发补助,鼓励其在数字孪生、人工智能、AR/VR/MR 等核心技术的研发上加大投入。对于主营业务收入较高的元宇宙企业,经过认定后,根据其当年的实际研发投入给予多年且高比例的补助。将元宇宙相关技术研发纳入科技计划体系,推动重要技术领域的研发实力和技术积累达到领先水平。为促进元宇宙技术的集成应用和融合创新,鼓励高校、科研院所、创新平台、互联网龙头企业以及元宇宙技术创新型企业等进行合作,开展技术共享、联合适配和协同攻关等活动。加强对元宇宙关键硬件的支持,包括高端芯片和传感器的制造,以及 VR/AR/MR 终端设备的开发。为吸引更多企业投入到元宇宙产业中,需要支持 VR 终端和内容生产等领域的龙头企业建立总部或研发中心。

加强对该领域的监管和引导措施。应当加快相关法律法规的制定,涵盖元宇宙资本运作、虚拟金融支付和信息安全等方面的内容,以确保虚拟世界中的活动合法、有序进行,避免信息泄露、盗用等安全风险的产生。为此,应着力打造数字信任的安全保障体系,通过加速安全芯片、零信任网络、可信计算和量子通信等技术的研发和产业化,提升云、网、端的数据安全服务与保障能力。同时,针对虚拟空间中的对象信息,建立完整的数字身份标识解析规则,以满足虚拟空间中的复杂流程需求。此外,应探索治理与监管沙盒机制,为新技术和新模式提供试验田,并建立元宇宙行业的相关标准和规范,强化平台主体行业的监管,加强伦理制约,防止投机炒作。推进国家数据基础制度的先试先行,建立健全数据安全、权利保护、跨境传输管理、交易流通、安全认证等数据制度和标准规范体系。加快数据要素的价值开发,建设区域性、行业级的数据交易平台,探索基于 NFT 的数字资产交易场景建设,以推动元宇宙行业的健康发展。

加强元宇宙产业的扶持力度。首先,出台资金引导政策,通过建立产业引导基金,采取“母基金+直投”的方式联合社会资本,打造一只专注于元宇宙产业的基金。该基金将主要用于支持元宇宙初创项目和重大项目,以推动产业生态建设。此外,设立子基金,专注于早期和长期投资,为元宇宙产业提供持续稳定的资金支持。支持元宇宙企业及服务机构集聚,对于在元宇宙应用创新中心新注册并租赁自用办公场地的重点企业,将根据房租补贴标准进行分类补贴,以鼓励企业在此类创新中心集聚发展。对于成功在主板上市的企业,给予一定的奖励。同时为推动元宇宙技术的创新与应用,需要设立奖励机制,涵盖元宇宙标志性场景、关键共性技术与通用能力的价值创新、公共服务平台建设以及特定研究方向的元宇宙相关项目等三个层面。

6.3 元宇宙应用规范制定和评测体系完善

如上文所述，随着元宇宙的快速发展，越来越多的应用涌现出来，这些应用不仅给用户带来了无限的可能性，也面临着诸多挑战。为了确保用户能够在规范、友好、安全的环境中体验元宇宙，我们需要有明确的规范和有效的评测机制，确保元宇宙生态系统的健康发展和持续优化。本章将探讨如何建立该类规范和评测平台，以及展望未来如何通过相关规范的制定，以及相关评测机制推动元宇宙生态系统的可持续发展。

在元宇宙应用中，内容的规范是确保用户体验、安全性和可持续性的重要基石。未来对于元宇宙相关的应用和内容，应该定义在法律允许范围内的内容类型和禁止的内容类型，特别是在内容生产方面，包括但不限于文字、图片、视频形式等。另外，对于未来新兴的元宇宙内容制作服务和产业，应该在法律层面包含知识产权保护措施，确保用户的创作受到合理的保护，规范相关的使用。

在元宇宙生态系统中，确保用户数据的安全和隐私至关重要。这包括数据的收集、存储、处理和共享等各个环节的规范。为防止隐私受侵、黑客攻击和虚假信息的传播，必须建立强有力的安全措施，保护用户免受各种潜在的威胁。例如，欧盟出台的《通用数据保护条例》（GDPR）[360]保障了数据处理的合法性、透明性和安全性。此外，不同年龄群体的数据隐私保护也需特别关注，例如美国的《儿童在线隐私保护法》（COPPA）[361]，专门保护 13 岁以下儿童的在线隐私。我国已颁布《数据安全法》，为相关问题提供了法律保障。因此，未来元宇宙的实际应用应严格遵循相关法规，确保所有用户的数据隐私和安全得到充分保护，为用户提供一个安全可信的虚拟环境。

在元宇宙生态系统中，确保所有用户都能平等地访问和使用虚拟环境至关重要。可访问性规范应涵盖从用户接口到功能实现的各个方面，确保无论用户的身体能力、技术水平或设备条件如何，都能顺利进入和使用元宇宙应用。在应用的开发与设计的同时，应遵循无障碍设计原则，例如提供语音交互、触觉交互和实时翻译等功能，以帮助视力、听力或语言障碍的用户顺利使用。此外，还应提供多语言支持和文化适应性，尊重和包容不同背景的用户。

在未来元宇宙的应用中，还应建立可靠的元宇宙法律和监管框架。未来元宇宙不仅是一个虚拟的空间，更是一个完善的社会化平台，用户在其中可以进行各种形式的交流和互动，如聊天、交易、合作等。为了确保这些互动是有序和有效的，必须建立明确可靠的元宇宙法律和监管框架。例如，社交互动是元宇宙用户体验的重要组成部分，制定

元宇宙的社交互动规范有助于构建一个和谐、健康和积极的虚拟社交环境。这些规范应涵盖用户行为、互动规则和内容审核等方面，确保每个用户都能在元宇宙中享受到积极的社交体验，应制定明确的行为准则，禁止欺凌、骚扰、歧视和仇恨言论等不良行为，鼓励尊重和友善的互动。

在确保元宇宙应用规范的同时，建立全方位评测平台是至关重要的。同时，其评测标准的制定可以为元宇宙的应用提供系统性的测试和评估，从而保障测试和评估的质量和稳定性，确保元宇宙环境的健康发展和持续改进。

相关部门可以牵头制定元宇宙的合规评测标准，这些举措将有助于企业和研究机构，助推元宇宙产业朝着规范化和健康发展的方向迈进。通过制定合规评测标准，可以确保元宇宙的研发过程和实际应用全流程符合道德和法律要求。这包括物理世界和虚拟世界中数据采集和使用的透明度、合法性、隐私保护措施，以及对敏感主题和内容的处理原则。同时，基于评测标准的指南提供在开发和部署元宇宙应用的研究机构和企业提供的最佳实践参考，有助于推动行业的规范化发展和可持续进步。制定元宇宙应用的评测标准和具体方法，在不同应用领域和任务中评测其表现，帮助用户和开发者更好地了解 and 评估系统的性能和可靠性，从而为选择合适的应用场景提供参考。评测平台应提供标准化的评测指标体系、测量指标和分析方法等，促进技术性能的客观比较和提升。例如，对于元宇宙相关的体验，可以从帧率、跟踪精度、系统稳定性和延迟相关的客观测量指标入手，结合沉浸感、舒适度、可用性、情感参与度和认知负荷等主观指标，完善标准化的评测体系。对于元宇宙密切相关的显示设备使用体验，可以制定明确的评测标准和指标，包括沉浸体验质量输入参数、呈现体验质量输入参数、交互体验质量输入参数等。

同时，评测平台应涵盖多样化的评测任务，包括用户交互体验、内容质量、跨平台兼容性、可访问性、安全性和伦理等不同领域，以全面评估系统在各种任务上的表现，推动多领域的研究和应用探索。

此外，在上述标准制定过程应该探索符合我国元宇宙实际应用背景的评测规范和方法论，明确评测过程中的数据准备、评估指标及测试方法等具体细节，以确保标准的实用性和适用性更符合我国元宇宙应用需求。

通过上述平台的建立及其评测标准的制定，可以为元宇宙的应用提供系统性的测试和评估，确保元宇宙应用符合各项规范和标准，确保元宇宙环境的健康发展和持续改进，进一步推动未来元宇宙技术的创新和实际应用，助力产业的整体进步。

参考文献

- [1] 王文喜, 周芳, 万月亮, 等. 元宇宙技术综述. 工程科学学报 44(04): 744-756 (2022).
- [2] The Metaverse: Concepts and Issues for Congress.
<https://crsreports.congress.gov/product/pdf/R/R47224>
- [3] 中商产业研究院 《2024-2029 全球与中国元宇宙市场现状及未来发展趋势》.
- [4] 史元春. 元宇宙需要人机交互的突破. 中国人工智能学会通讯(6): 26-33 (2022).
- [5] Yuchen Guo, Tao Yu, Jiamin Wu, Yuwang Wang, Sen Wan, Jiyuan Zheng, Lu Fang, Qionghai Dai. Artificial intelligence for Metaverse: a framework. CAAI Artificial Intelligence Research 1(1): 54-67 (2022).
- [6] Minrui Xu, Wei Chong Ng, Wei Yang Bryan Lim, Jiawen Kang, Zehui Xiong, Dusit Niyato, Qiang Yang, Xuemin Shen, Chunyan Miao. A Full Dive Into Realizing the Edge-Enabled Metaverse: Visions, Enabling Technologies, and Challenges. IEEE Communications Surveys & Tutorials 25(1): 656-700 (2023).
- [7] Lik-Hang Lee, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, Pan Hui. All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda. arXiv preprint arXiv: 2110.05352, 2021.
- [8] Zryan Najat Rashid, Subhi R. M. Zebari, Karzan Hussein Sharif, Karwan Jacksi. Distributed Cloud Computing and Distributed Parallel Computing: A Review. IEEE International Conference on Advanced Science and Engineering 2018: 167-172.
- [9] Neelesh Mungoli. Scalable, Distributed AI Frameworks: Leveraging Cloud Computing for Enhanced Deep Learning Performance and Efficiency. arXiv preprint arXiv: 2304.13738, 2023.
- [10] Yu Wei Tan, Siang Ern Low, Jonas Chow, Javon Teo, Anand Bhojan. Distributed Hybrid Rendering with Realistic Real-time Shadows for Interactive Thin Client Metaverse and Game Applications. The Visual Computer 2024.
- [11] Thien Huynh-The, Quoc-Viet Pham, Xuan-Quy Pham, Thanh Thi Nguyen, Zhu Han, Dong-Seong Kim. Artificial Intelligence for the Metaverse: A Survey. Engineering Applications of Artificial Intelligence 117(Part): 105581 (2023).

- [12] 曹焯. Mars 说光场(3)—光场采集.
<https://www.leiphone.com/category/arvr/YX8QJhR2pn2hDtK4.html>.
- [13] VOMMA 视觉. 国内外光场相机应用案例对比.
<https://baijiahao.baidu.com/s?id=1770302644213465960&wfr=spider&for=pc>.
- [14] David J. Brady, Michael E. Gehm, Ronald A. Stack, Daniel L. Marks, David S. Kittle, Dathon R. Golish, Esteban Vera, Steven D. Feller. Multiscale gigapixel photography. *Nature* 486(7403): 386-389 (2012).
- [15] Xiaoyun Yuan, Lu Fang, Qionghai Dai, David J. Brady, Yebin Liu. Multiscale gigapixel video: a cross resolution image matching and warping approach. *IEEE International Conference on Computational Photography* 2017: 33-41.
- [16] Zhihao Xu, Xiaoyun Yuan, Tiankuang Zhou, Lu Fang. A multichannel optical computing architecture for advanced machine vision. *Light: Science & Applications* 11: 255 (2022)
- [17] 张茂军, 刘煜, 王伟等. 《计算摄影学基础》. 科学出版社, 2014
- [18] 郑太雄, 黄帅, 李永福等. 基于视觉的三维重建关键技术研究综述. *自动化学报* 46(4): 631-652 (2020).
- [19] Nikolay Savinov, Lubor Ladicky, Christian Hane, Marc Pollefeys. Discrete optimization of ray potentials for semantic 3D reconstruction. *IEEE Conference on Computer Vision and Pattern Recognition* 2015: 5511-5518.
- [20] Maros Blaha, Christoph Vogel, Audrey Richard, Jan Dirk Wegner, Thomas Pock, Konrad Schindler. Large-scale semantic 3D reconstruction: an adaptive multi-resolution model for multi-class volumetric labeling. *IEEE Conference on Computer Vision and Pattern Recognition* 2016: 3176-3184.
- [21] Niko Sünderhauf, Trung T. Pham, Yasir Latif, Michael Milford, Ian D. Reid. Meaningful Maps With Object-Oriented Semantic Mapping. *IEEE/RSJ International Conference on Intelligent Robots and Systems* 2017: 5079-5085.
- [22] Guangyu Wang, Jinzhi Zhang, Fan Wang, Ruqi Huang, Lu Fang. XScale-NVS: Cross-Scale Novel View Synthesis with Hash Featurized Manifold. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2024: 21029-21039.
- [23] Jianyuan Wang, et al. VGGT: Visual Geometry Grounded Transformer. *CVPR*, 2025.

- [24] 3DCAT. Light stage 技术发展介绍. <https://www.3dcats.live/share/light-stage/>
- [25] Najim Dehak, Patrick J. Kenny, Réda Dehak, Pierre Dumouchel, Pierre Ouellet. Front-End Factor Analysis for Speaker Verification. *IEEE Transaction on Audio, Speech and Language Processing* 19(4): 788-798 (2011).
- [26] Ehsan Variani, Xin Lei, Erik McDermott, Ignacio López-Moreno, Javier Gonzalez-Dominguez. Deep neural networks for small footprint text-dependent speaker verification. *IEEE International Conference on Acoustics, Speech and Signal Processing 2014*: 4052-4056.
- [27] Sun Y, Zhang S, Chen Z, et al. Soft contact simulation and manipulation learning of deformable objects with vision-based tactile sensor. *IEEE Transactions on Automation Science and Engineering* 22: 17618-17630 (2025).
- [28] Shen Z, Sun Y, Zhang S, et al. Simulation of optical tactile sensors supporting slip and rotation using path tracing and IMPM. *IEEE Robotics and Automation Letters* 9(12): 11218-11225 (2024).
- [29] Sun Y, Cheng N, Zhang S, et al. Tactile data generation and applications based on visuo-tactile sensors: A review. *Information Fusion*, 2025: 103162 (2025).
- [30] Gomes D F, Paoletti P, Luo S. Generation of gelsight tactile images for sim2real learning. *IEEE Robotics and Automation Letters* 6(2): 4177-4184 (2021).
- [31] Bi T, Sferrazza C, D'Andrea R. Zero-shot sim-to-real transfer of tactile control policies for aggressive swing-up manipulation. *IEEE Robotics and Automation Letters* 6(3): 5761-5768 (2021).
- [32] Chen Z, Zhang S, Luo S, et al. Tacchi: A pluggable and low computational cost elastomer deformation simulator for optical tactile sensors. *IEEE Robotics and Automation Letters* 8(3): 1239-1246 (2023).
- [33] Sun Y, Zhang S, Shan J, et al. Simulation of vision-based tactile sensors with efficiency-tunable rendering. *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO) 2023*: 1-6.
- [34] Zhao Y, Qian K, Duan B, et al. Fots: A fast optical tactile simulator for sim2real learning of tactile-motor robot manipulation skills. *IEEE Robotics and Automation Letters* 9(6): 5647-5654 (2024).

- [35] Chen W, Xu J, Xiang F, et al. General-purpose sim2real protocol for learning contact-rich manipulation with marker-based visuotactile sensors. *IEEE Transactions on Robotics* 40: 1509-1526 (2024).
- [36] Zhao Y, Qian K, Duan B, et al. Fots: A fast optical tactile simulator for sim2real learning of tactile-motor robot manipulation skills. *IEEE Robotics and Automation Letters* 9(6): 5647-5654 (2024).
- [37] Chen W, Xu Y, Chen Z, et al. Bidirectional sim-to-real transfer for gelsight tactile sensors with cyclegan. *IEEE Robotics and Automation Letters* 7(3): 6187-6194 (2022).
- [38] Cheng N, Xu J, Guan C, et al. Touch100k: A large-scale touch-language-vision dataset for touch-centric multimodal representation. *Information Fusion* 2025: 103305 (2025).
- [39] Yang F, Feng C, Chen Z, et al. Binding touch to everything: Learning unified multimodal tactile representations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024*: 26340-26353.
- [40] Feng R, Hu J, Xia W, et al. Anytouch: Learning unified static-dynamic representation across multiple visuo-tactile sensors. *The Thirteenth International Conference on Learning Representations 2025*: 1-21.
- [41] Li M, Zhou Y H, Zhang L, et al. OneTip: A soft tactile interface for 6-D fingertip pose acquisition in human-computer interaction. *Sensors and Actuators A: Physical* 379: 115896 (2024).
- [42] Suresh S, Qi H, Wu T, et al. NeuralFeels with neural fields: Visuotactile perception for in-hand manipulation. *Science Robotics* 9(96): eadl0628 (2024).
- [43] Zhang S, Du J, Sun Y, et al. A braille detection system based on visuo-tactile sensing. *Measurement* 247: 116827 (2025).
- [44] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, Jerome Revaud, DUST3R: Geometric 3D Vision Made Easy. *IEEE CVPR*, 20697-20709, 2024.
- [45] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, David Novotny, VGGT: Visual Geometry Grounded Transformer, *IEEE CVPR*, 5294-5306, 2025.
- [46] Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan,

Lin Gao. Recent advances in 3d gaussian splatting. arXiv preprint arXiv: 2403.11134, 2024.

- [47] Wojciech Zielonka, Timur M. Bagautdinov, Shunsuke Saito, Michael Zollhöfer, Justus Thies, Javier Romero. Drivable 3d gaussian avatars. arXiv preprint arXiv: 2311.08581, 2023.
- [48] Jena R, Iyer GS, Choudhary S, Smith B, Chaudhari P, Gee J, Splatarmor: Articulated gaussian splatting for animatable humans from monocular rgb videos. arXiv preprint arXiv: 2311.10812, 2023.
- [49] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, Anurag Ranjan. Hugs: Human gaussian splats. arXiv preprint arXiv: 2311.17910, 2023.
- [50] Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, Lan Xu. Hifi4g: High-fidelity human performance rendering via compact gaussian splatting. arXiv preprint arXiv: 2312.03461, 2024.
- [51] Shunyuan Zheng, Boyao Zhou, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang Nie, Yebin Liu. Gps-gaussian: Generalizable pixel-wise 3d gaussian splatting for real-time human novel view synthesis. arXiv preprint arXiv: 2312.02155, 2024.
- [52] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, Matthias Nießner. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. arXiv preprint arXiv: 2312.02069, 2023.
- [53] Helisa Dhamo, Yinyu Nie, Arthur Moreau, Jifei Song, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero. Headgas: Real-time animatable head avatars via 3d gaussian splatting. arXiv preprint arXiv: 2312.02902, 2023.
- [54] Jun Xiang, Xuan Gao, Yudong Guo, Juyong Zhang. FlashAvatar: High-Fidelity Digital Avatar Rendering at 300FPS. arXiv preprint arXiv: 2312.02214, 2023.
- [55] Zheheng Jiang, Hossein Rahmani, Sue Black, Bryan M. Williams. 3D Points Splatting for Real-Time Dynamic Hand Reconstruction. arXiv preprint arXiv: 2312.13770, 2023.
- [56] Haimin Luo, Min Ouyang, Zijun Zhao, Suyi Jiang, Longwen Zhang, Qixuan Zhang, Wei Yang, Lan Xu, Jingyi Yu. GaussianHair: Hair Modeling and Rendering with Light-aware Gaussians. arXiv preprint arXiv: 2402.10483, 2024.
- [57] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser

- Kehtarnavaz, Mubarak Shah. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys* 56(1): 11:1-11:37 (2024).
- [58] Mir Rayat Imtiaz Hossain, James J. Little. Exploiting temporal information for 3D human pose estimation. *European Conference on Computer Vision* (10) 2018: 69-86.
- [59] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, Michael J. Black. SMPL: A Skinned Multi-Person Linear Model. *ACM Transactions on Graphics* 34(6): 1-16 (2015).
- [60] Muhammed Kocabas, Nikos Athanasiou, Michael J Black. VIBE: Video inference for human body pose and shape estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2020: 5252-5262.
- [61] Helge Rhodin, Jörg Spörri, Isinsu Katircioglu, Victor Constantin, Frédéric Meyer, Erich Müller, Mathieu Salzmann, Pascal Fua. Learning Monocular 3D Human Pose Estimation From Multi-View Images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2018: 8437-8446.
- [62] Helge Rhodin, Mathieu Salzmann, and Pascal Fua. Unsupervised Geometry-Aware Representation for 3D Human Pose Estimation. *European Conference on Computer Vision* (10) 2018: 765-782.
- [63] Georgios Pavlakos, Xiaowei Zhou, Konstantinos G. Derpanis, Kostas Daniilidis. Harvesting Multiple Views for Marker-Less 3D Human Pose Annotations. *IEEE Conference on Computer Vision and Pattern Recognition Conference* 2017: 1253-1262.
- [64] Tao Wang, Jianfeng Zhang, Yujun Cai, Shuicheng Yan, Jiashi Feng. Direct Multi-view Multi-person 3D Pose Estimation. *Advances in Neural Information Processing Systems* 2021: 13153-13164.
- [65] 张子瑞, AIGC 赋能科幻影视技术观察: 动作捕捉篇, 中国科普作家网.
<https://www.kpcswa.org.cn/web/press/members/works/0FAB22023.html>, 2023.
- [66] Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, Tao Chen. MotionGPT: Human Motion as Foreign Language. *Advances in Neural Information Processing Systems* 2023: 20067-20079.
- [67] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative adversarial nets. *Advances*

in Neural Information Processing Systems 2014: 2672-2680.

- [68] Ho, Jonathan, Ajay Jain, Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 2020: 6840-6851.
- [69] Karras, Tero, Samuli Laine, Timo Aila. A style-based generator architecture for generative adversarial networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2019: 4401-4410.
- [70] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J. Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, Gordon Wetzstein. Efficient geometry-aware 3d generative adversarial networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2022: 16102-16112.
- [71] Yiqian Wu, Hao Xu, Xiangjun Tang, Xien Chen, Siyu Tang, Zhebin Zhang, Chen Li, Xiaogang Jin. Portrait3D: Text-Guided High-Quality 3D Portrait Generation Using Pyramid Representation and GANs Prior. *arXiv preprint arXiv: 2404.10394*, 2024.
- [72] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, Mark Chen. Point-e: A system for generating 3d point clouds from complex prompts. *arXiv preprint arXiv: 2212.08751*, 2022.
- [73] Heewoo Jun, Alex Nichol. Shap-e: Generating conditional 3d implicit functions. *arXiv preprint arXiv: 2305.02463*, 2023.
- [74] Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, Xinggang Wang. Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2024: 6796-6807.
- [75] Devikalyan Das, Christopher Wewer, Raza Yunus, Eddy Ilg, Jan Eric Lenssen. Neural parametric gaussians for monocular non-rigid object reconstruction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference* 2024: 10715-10725.
- [76] Agelos Kratimenos, Jiahui Lei, Kostas Daniilidis. Dynmf: Neural motion factorization for real-time dynamic view synthesis with 3d gaussian splatting. *arXiv preprint arXiv: 2312.00112*, 2023.
- [77] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, Jing Liao. Clip-nerf:

Text-and-image driven manipulation of neural radiance fields. IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference 2022: 3825-3834.

- [78] Alexander Vilesov , Pradyumna Chari , Achuta Kadambi. Cg3d: Compositional generation for text-to-3d via gaussian splatting. arXiv preprint arXiv: 2311.17907, 2023.
- [79] Xinhai Li, Huaibin Wang, Kuo-Kun Tseng. Gaussiandiffusion: 3d gaussian splatting for denoising diffusion probabilistic models with structured noise. arXiv preprint arXiv: 2311.11221, 2023.
- [80] <https://www.tripo3d.ai/app>
- [81] <https://hyperhuman.deemos.com/rodin>
- [82] Ha D, Schmidhuber J. World models. arXiv preprint arXiv:1803.10122, 2018, 2(3).
- [83] Assran M, Duval Q, Misra I, et al. Self-supervised learning from images with a joint-embedding predictive architecture. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 15619-15629.
- [84] Liu Y, Zhang K, Li Y, et al. Sora: A review on background, technology, limitations, and opportunities of large vision models. arXiv preprint arXiv:2402.17177, 2024.
- [85] Shang Y, Lin Y, Zheng Y, et al. Urbanworld: An urban world model for 3d city generation. arXiv preprint arXiv:2407.11965, 2024.
- [86] Bruce J, Dennis M D, Edwards A, et al. Genie: Generative interactive environments. Forty-first International Conference on Machine Learning. 2024.
- [87] Mao X, Lin S, Li Z, et al. Yume: An interactive world generation model. arXiv preprint arXiv:2507.17744, 2025.
- [88] He X, Peng C, Liu Z, et al. Matrix-game 2.0: An open-source, real-time, and streaming interactive world model. arXiv preprint arXiv:2508.13009, 2025.
- [89] Song K, Chen B, Simchowitz M, et al. History-guided video diffusion. arXiv preprint arXiv:2502.06764, 2025.
- [90] Teng H, Jia H, Sun L, et al. MAGI-1: Autoregressive Video Generation at Scale. arXiv preprint arXiv:2505.13211, 2025.
- [91] Wu T, Yang S, Po R, et al. Video World Models with Long-term Spatial Memory. arXiv preprint arXiv:2506.05284, 2025.
- [92] Li R, Torr P, Vedaldi A, et al. VMem: Consistent Interactive Video Scene Generation with

- Surfel-Indexed View Memory. arXiv preprint arXiv:2506.18903, 2025.
- [93] Li J, Tang J, Xu Z, et al. Hunyuan-GameCraft: High-dynamic Interactive Game Video Generation with Hybrid History Condition. arXiv preprint arXiv:2506.17201, 2025.
- [94] Wan T, Wang A, Ai B, et al. Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314, 2025.
- [95] Li Z, Zhang J, Lin Q, et al. Hunyuan-dit: A powerful multi-resolution diffusion transformer with fine-grained chinese understanding. arXiv preprint arXiv:2405.08748, 2024.
- [96] Yang Z, Teng J, Zheng W, et al. Cogvideox: Text-to-video diffusion models with an expert transformer. arXiv preprint arXiv:2408.06072, 2024.
- [97] Stackpole Thomas. What Is Web3? Harward Business Review, 2022
<https://hbr.org/2022/05/what-is-web3>
- [98] 中金公司, 元宇宙: 空间升维、时间延展、社会重构, 2022.
<https://research.cicc.com/frontend/recommend/detail?id=2976>
- [99] Nakamoto Satoshi. A peer-to-peer electronic cash system.
<https://bitcoin.org/bitcoin.pdf>, 2008.
- [100] What is the Ethereum Cancun-Deneb (Dencun) Update?
<https://medium.com/buildbear/what-is-the-ethereum-cancun-deneb-dencun-update-dde8bd9943ae>
- [101] Buterin Vitalik. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform.
https://ethereum.org/content/whitepaper/whitepaper-pdf/Ethereum_Whitepaper_-_Buterin_2014.pdf, 2014
- [102] Coinbase. What is a smart contract?
<https://www.coinbase.com/learn/crypto-basics/what-is-a-smart-contract>
- [103] <https://ethereum.org/zh/developers/docs/intro-to-ethereum/>
- [104] Casey M, Crane J, Gensler G, Johnson S, Narula N. The impact of blockchain technology on finance: A catalyst for change. 2018.
- [105] <https://ethereum.org/zh/developers/docs/web2-vs-web3/>
- [106] Kyle Croman, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, Emin Gün Sirer, Dawn Song,

- Roger Wattenhofer. On Scaling Decentralized Blockchains. International Conference on Financial Cryptography and Data Security 2016: 106-125.
- [107] Dencun Mainnet Announcement.
<https://blog.ethereum.org/zh/2024/02/27/dencun-mainnet-announcement>
- [108] <https://l2beat.com>.
- [109] Binance Research. Half-Year Report 2025.
<https://public.bnbstatic.com/static/files/research/half-year-report-2025.pdf>
- [110] Alex De Vries. Bitcoin's growing energy problem. Joule2(5): 801-5 (2018).
- [111] Ethereum's energy expenditure.
<https://ethereum.org/en/energy-consumption>
- [112] ERC-4337: Account Abstraction Using Alt Mempool, Account abstraction without consensus-layer protocol changes, instead relying on higher-layer infrastructure.
<https://eips.ethereum.org/EIPS/eip-4337>
- [113] What are zk-snarks? <https://z.cash/learn/what-are-zk-snarks/>
- [114] Szabo Nick. Smart Contracts: Building Blocks for Digital Markets.
- [115] <https://www.certora.com>
- [116] <https://docs.openzeppelin.com/upgrades-plugins>
- [117] SCALING, <https://ethereum.org/en/developers/docs/scaling/>
- [118] Poon Joseph, Dryja Thaddeus. The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments, 2016.
- [119] Adam Back, Matt Corallo, Luke Dashjr, Mark Friedenbach, Gregory Maxwell, Andrew Miller, Andrew Poelstra, Jorge Timón, Pieter Wuille. Enabling Blockchain Innovations with Pegged Sidechains, 2014
- [120] Poon Joseph, Buterin Vitalik. Plasma: Scalable Autonomous Smart Contracts, 2017
- [121] Rollup Protocol Overview.
<https://docs.optimism.io/stack/protocol/rollup/overview>
- [122] Zk Sync Documents, URL: <https://docs.zksync.io/build>
- [123] Dimitar Bogdanov. Optimistic Rollups vs ZK Rollups: Examining Six of the Most Exciting Layer 2 Scaling Projects for Ethereum, URL: <https://limechain.tech/blog/optimistic-rollups-vs-zk-rollups/>, 2021

- [124] <https://ethereum.org/zh/developers/docs/scaling/validium/>
- [125] <https://celestia.org>
- [126] <https://docs.eigenlayer.xyz>
- [127] <https://ethereum.org/zh/developers/docs/bridges/>
- [128] What Are Blockchain Bridges and How Can We Classify Them?
<https://li.fi/knowledge-hub/blockchain-bridges-and-classification/>
- [129] Arjun Bhuptani. The Interoperability Trilemma, URL: <https://medium.com/connext/the-interoperability-trilemma-657c2cf69f17>, 2021
- [130] <https://rekt.news/leaderboard>
- [131] <https://cosmos.network/ibc>
- [132] <https://layerzero.network>
- [133] <https://ethereum.stackexchange.com/questions/301/why-cant-contracts-make-api-calls>
- [134] <https://ethereum.org/zh/developers/docs/oracles/>
- [135] Steve Ellis, Ari Juels, Sergey Nazarov. ChainLink: A Decentralized Oracle Network, URL: <https://research.chain.link/whitepaper-v1.pdf>, 2017
- [136] Lorenz Breidenbach, Christian Cachin, Benedict Chan, Alex Coventry, Steve Ellis, Ari Juels, Farinaz Koushanfar, Andrew Miller, Brendan Magauran, Daniel Moroz, Sergey Nazarov, Alexandru Topliceanu, Florian Tram`er, Fan Zhang. Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks, 2021.
<https://research.chain.link/whitepaper-v2.pdf>
- [137] Ryan Sean Adams. Why we need decentralized oracles, 2021.
<https://www.bankless.com/why-we-need-decentralized-oracles>, 2021
- [138] Hayden Adams, Noah Zinsmeister, Dan Robinson. Uniswap v2 Core, URL: <https://uniswap.org/whitepaper.pdf>
- [139] <https://pyth.network>
- [140] <https://bandprotocol.com>
- [141] <https://api3.org>
- [142] ERC-721 NON-FUNGIBLE TOKEN STANDARD
<https://ethereum.org/en/developers/docs/standards/tokens/erc-721/>
- [143] ERC-1155 MULTI-TOKEN STANDARD

- <https://ethereum.org/en/developers/docs/standards/tokens/erc-1155/>
- [144] Logan Kugler. Non-fungible tokens and the future of art. *Communications of the ACM* 64(9): 19-20 (2021).
 - [145] ERC-1948: Non-fungible Data Token,
<https://eips.ethereum.org/EIPS/eip-1948>
 - [146] ERC-2981: NFT Royalty Standard
<https://eips.ethereum.org/EIPS/eip-2981>
 - [147] ERC-1523: Standard for Insurance Policies as ERC-721 Non Fungible Tokens
<https://eips.ethereum.org/EIPS/eip-1523>
 - [148] Chainlink. What Is a Dynamic NFT (dNFT)?
<https://chain.link/education-hub/what-is-dynamic-nft>
 - [149] Coinbase. What are Real-World Assets (RWA)?
<https://www.coinbase.com/learn/crypto-glossary/what-are-real-world-assets-rwa>
 - [150] Ohlhaver, Puja and Weyl, Eric Glen and Buterin, Vitalik. Decentralized Society: Finding Web3's Soul (May 10, 2022).
<https://ssrn.com/abstract=4105763> or <http://dx.doi.org/10.2139/ssrn.4105763>
 - [151] Ralph C Merkle. DAOs, Democracy and Governance, 2016.
 - [152] Samer Hassan, Primavera De Filippi. Decentralized Autonomous Organization. *Internet Policy Review* 10(2) (2021).
 - [153] Mitchell Goldberg, Fabian Schär. Metaverse governance: An empirical analysis of voting within Decentralized Autonomous Organizations. *Journal of Business Research* 5(160) (2023).
 - [154] <https://aragon.org>
 - [155] QUADRATIC VOTING.
<https://www.radicalxchange.org/concepts/quadratic-voting>
 - [156] Optimism. How Retro Funding Works.
<https://community.optimism.io/citizens-house/how-retro-funding-works>
 - [157] Agostino Capponi, Alfio Gliozzo, Chunghyun Han, Junkyu Lee. DAO-AI: Evaluating Collective Decision-Making through Agentic AI in Decentralized Governance.
<https://arxiv.org/abs/2510.21117>

- [158] Marc Jansen, Christophe Verdot. QOC DAO -- Stepwise Development Towards an AI Driven Decentralized Autonomous Organization.
<https://arxiv.org/abs/2511.08641>
- [159] Solana Foundation, Proof of History: How Solana brings time to crypto
<https://solana.com/news/proof-of-history>, 2021
- [160] <https://docs.eigenlayer.xyz/eigenlayer/overview/key-terms>
- [161] Drummond Reed, Manu Sporny, Markus Sabadello. Decentralized identifiers (dids) v1. 0. Draft Community Group Report, 2020
- [162] Mao, Y., et al. (2025). Mapping the metaverse minefield: A TIPS framework for security-conscious business adoption. Computers & Security, 149, 104710.
<https://doi.org/10.1016/j.cose.2025.104710>
- [163] Towards a secure Metaverse: Leveraging hybrid model for IoT anomaly detection. (2025). PLOS ONE. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11996210/>
- [164] Yvo Desmedt. Threshold cryptosystems. International Workshop on the Theory and Application of Cryptographic Techniques 1992: 1-14.
- [165] Rosario Gennaro, Stanislaw Jarecki, Hugo Krawczyk, Tal Rabin. Secure distributed key generation for discrete-log based cryptosystems. Journal of Cryptology 20(1): 51-83 (2007).
- [166] Quantum-Safe Threshold Cryptography for Decentralized Group Key Management via Dealerless DKG (CRYSTALS - Kyber). (2025). Mathematics, 13(21), 3429.
<https://www.mdpi.com/2227-7390/13/21/3429>
- [167] Weiqin Zou, David Lo, Pavneet Singh Kochhar, Xuan-Bach Dinh Le, Xin Xia, Yang Feng, Zhenyu Chen, Baowen Xu. Smart contract development: Challenges and opportunities. IEEE Transactions on Software Engineering 47(10): 2084-2106 (2021).
- [168] 10 Metaverse and NFT Trends 2025-2026. (2025). <https://beinsure.com/metaverse-nfts-trends-outlook/>
- [169] 7 Top NFT Trends in 2025. (2025). CoinDCX.
<https://coindcx.com/blog/cryptocurrency/top-nft-trends/>
- [170] James Joshi, Elisa Bertino, Usman Latif, Arif Ghafoor. A generalized temporal role-based access control model. IEEE Transactions on Knowledge and Data Engineering 17(1):

4-23 (2005).

- [171] Mitigating Risks in the Cloud-Based Metaverse Access Control Strategies. (2025). International Journal of Cloud Applications and Computing. <https://www.sciencedirect.com/org/science/article/pii/S2156183424000020>
- [172] Virtual Reality and Identity: Securing the Metaverse Frontier. (2025). Avatier. <https://www.avatier.com/blog/virtual-reality-and-identity/>
- [173] Denis Diemert, Tibor Jager. On the tight security of TLS 1.3: Theoretically sound cryptographic parameters for real-world deployments. Journal of Cryptology 34(3): 30 (2021).
- [174] René Schwonnek, Koon Tong Goh, Ignatius W. Primaatmaja, Ernest Y.-Z. Tan, Ramona Wolf, Valerio Scarani & Charles C.-W. Lim. Device-independent quantum key distribution with random key basis. Nature communications 12(1): 2880 (2021).
- [175] Efficient post-quantum attribute-based access control scheme for blockchain-empowered metaverse data management. (2025). Computer Communications, 236, 490. <https://www.sciencedirect.com/science/article/pii/S0140366425000490>
- [176] Michael Rosenberg, Jacob D. White, Christina Garman, Ian Miers. zk-creds: Flexible anonymous credentials from zksnarks and existing identity infrastructure. IEEE Symposium on Security and Privacy 2023: 790-808.
- [177] Chao Lin, Debiao He, Xinyi Huang, Muhammad Khurram Khan, Kim-Kwang Raymond Choo. DCAP: A secure and efficient decentralized conditional anonymous payment system based on blockchain. IEEE Transactions on Information Forensics and Security 15: 2440-2452 (2020).
- [178] Wanxin Li, Hao Guo, Mark M. Nejad, Chien-Chung Shen. Privacy-preserving traffic management: A blockchain and zero-knowledge proof inspired approach. IEEE access, 8: 181733-181743 (2020).
- [179] Advanced Metaverse Wireless Communication Systems. (2025). IET Digital Library. <https://digital-library.theiet.org/doi/book/10.1049/PBTE112E>
- [180] Vu Tuan Truong, Long Bao Le. MetaCIDS: A metaverse collaborative intrusion detection system based on blockchain and federated learning[J]. Authorea Preprints, 2023.
- [181] Vu Tuan Truong, Long Bao Le. MetaCIDS: Privacy-preserving collaborative intrusion detection for metaverse based on blockchain and online federated learning. IEEE Open

Journal of the Computer Society 4: 253-266 (2023).

- [182] Vu Tuan Truong, Long Bao Le. Security for the Metaverse: Blockchain and Machine Learning Techniques for Intrusion Detection. IEEE Network, 2024.
- [183] Peng Sun, Zhibo Wang, Liantao Wu, Yunhe Feng, Xiaoyi Pang, Hairong Qi, Zhi Wang. Towards personalized privacy-preserving incentive for truth discovery in mobile crowdsensing systems. IEEE Transactions on Mobile Computing 21(1): 352-365 (2022).
- [184] Abhinav Jain, Hima Patel, Lokesh Nagalapatti, Nitin Gupta, Sameep Mehta, Shanmukha C. Guttula, Shashank Mujumdar, Shazia Afzal, Ruhi Sharma Mittal, Vitobha Munigala. Overview and importance of data quality for machine learning tasks. ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2020: 3561-3562.
- [185] L. Lyu et al., "Privacy and Robustness in Federated Learning: Attacks and Defenses," in IEEE Transactions on Neural Networks and Learning Systems (2024).
- [186] Erik Daniel, Florian Tschorsch. IPFS and friends: A qualitative comparison of next generation peer-to-peer data networks. IEEE Communications Surveys & Tutorials 24(1): 31-52 (2022).
- [187] David Butler, Andreas Lochbihler, David Aspinall, Adrià Gascón. Formalising Σ -Protocols and commitment schemes using CryptHOL. Journal of Automated Reasoning 65(4): 521-567 (2021)
- [188] Juha Partala, Tri Hong Nguyen, Susanna Pirttikangas. Non-interactive zero-knowledge for blockchain: A survey. IEEE Access 8: 227945-227961 (2020).
- [189] Xiaoqiang Sun, F. Richard Yu, Peng Zhang, Zhiwei Sun, Weixin Xie, Xiang Peng. A survey on zero-knowledge proof in blockchain. IEEE Network 35(4): 198-205 (2021).
- [190] P. Duan, Z. Ma, H. Gao, T. Tian and Y. Zhang, "Multi-Authority Attribute-Based Encryption Scheme With Access Delegation for Cross Blockchain Data Sharing," in IEEE Transactions on Information Forensics and Security (2025).
- [191] Geoffroy Couteau, Michael Klooß, Huang Lin, Michael Reichle. Efficient range proofs with transparent setup from bounded integer commitments. Annual International Conference on the Theory and Applications of Cryptographic Techniques (3) 2021: 247-277.
- [192] Zhu D, Zhou Z, Li Y, Zhang H, Chen Y, Zhao Z, Zheng J. A Survey of Data Security Sharing.

Symmetry (2025).

- [193] Lucjan Hanzlik, Daniel Slamanig. With a little help from my friends: Constructing practical anonymous credentials. ACM SIGSAC Conference on Computer and Communications Security 2021: 2004-2023.
- [194] Shivansh Kumar, Aman Kumar Bharti, Ruhul Amin. Decentralized secure storage of medical records using Blockchain and IPFS: A comparative analysis with future directions. Security and Privacy 4(5) (2021).
- [195] M. Tian, C. Gao, L. Chen, H. Zhong and J. Chen, "Proofs of Retrievability With Public Verifiability From Lattices," in IEEE Transactions on Information Forensics and Security, vol. 20, pp. 3925-3935 (2025).
- [196] David Joseph, Rafael Misoczki, Marc Manzano, Joe Tricot, Fernando Dominguez Pinuaga, Olivier Lacombe, Stefan Leichenauer, Jack Hidary, Phil Venables, Royal Hansen. Transitioning organizations to post-quantum cryptography. Nature 605(7909): 237-243 (2022).
- [197] Qian Wei, Bingzhe Li, Wanli Chang, Zhiping Jia, Zhaoyan Shen, Zili Shao. A survey of blockchain data management systems. ACM Transactions on Embedded Computing Systems 21(3): 25:1-25:28 (2022).
- [198] Rafael Belchior, André Vasconcelos, Sérgio Guerreiro, Miguel Correia. A survey on blockchain interoperability: Past, present, and future trends. ACM Computing Surveys 54(8): 168:1-168:41 (2022).
- [199] H. Zhang, X. Qi, H. Tang, Z. Zhang, C. Jin and A. Zhou, "Loom: A Deterministic Execution Framework Towards Nested Contract Transactions," 2025 IEEE 41st International Conference on Data Engineering (ICDE), Hong Kong, Hong Kong, 2025, pp. 2408-2421.
- [200] Gilad Asharov, Abhishek Jain, Adriana López-Alt, Eran Tromer, Vinod Vaikuntanathan, Daniel Wichs. Multiparty computation with low communication, computation and interaction via threshold FHE. Annual International Conference on the Theory and Applications of Cryptographic Techniques 2012: 483-501.
- [201] Martin R. Albrecht, Christian Rechberger, Thomas Schneider, Tyge Tiessen, Michael Zohner. Ciphers for MPC and FHE. Annual International Conference on the Theory and Applications of Cryptographic Techniques 2015: 430-454.

- [202] Assaf Ben-David, Noam Nisan, Benny Pinkas. FairplayMP: a system for secure multi-party computation. ACM Conference on Computer and Communications Security 2008: 257-266.
- [203] G. Lin et al., "Is MPC Secure? Leveraging Neural Network Classifiers to Detect Data Leakage Vulnerabilities in MPC Implementations," 2025 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2025, pp. 2809-2826.
- [204] Dayeol Lee, David Kohlbrenner, Shweta Shinde, Krste Asanovic, Dawn Song. Keystone: An open framework for architecting trusted execution environments. European Conference on Computer Systems 2020: 38:1-38:16.
- [205] Bernard Ngabonziza, Daniel Martin, Anna Bailey, Haehyun Cho, Sarah Martin. Trustzone explained: Architectural features and use cases. IEEE International Conference on Collaboration and Internet Computing 2016: 445-451.
- [206] S. Kumar, A. Panda, A. Nerlikar and S. R. Sarangi, "A Tug-of-War Between Static and Dynamic Memory in Intel SGX," 2025 38th International Conference on VLSI Design and 2024 23rd International Conference on Embedded Systems (VLSID), Bangalore, India, 2025, pp. 272-277.
- [207] Qinbin Li, Bingsheng He, Dawn Song. Model-contrastive federated learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021: 10713-10722.
- [208] Elabd, E. Dynamic differential privacy technique for deep learning models. Sci Rep 15, 39353 (2025).
- [209] Reeba Zahid, Ayesha Altaf, Tauqir Ahmad, Faiza Iqbal, Yini Airt Miró Vera, Miguel Angel López Flores, Imran Ashraf. Secure data management life cycle for government big-data ecosystem: Design and development perspective. Systems 11(8): 380 (2023).
- [210] Mosharraf, Maedeh. "Data governance in metaverse: Addressing security threats and countermeasures across the data lifecycle." Technology in Society 82 (2025): 102910.
- [211] Rakesh Agrawal, Peter J. Haas, Jerry Kiernan. Watermarking relational data: framework, algorithms and analysis. The VLDB Journal 12(2): 157-169 (2003).
- [212] Guohua Tian, Yunhan Hu, Jianghong Wei, Zheli Liu, Xinyi Huang, Xiaofeng Chen, Willy Susilo. Blockchain-based secure deduplication and shared auditing in

- decentralized storage. *IEEE Transactions on Dependable and Secure Computing* 19(6): 3941-3954 (2022).
- [213] Oded Goldreich, Yair Oren. Definitions and properties of zero-knowledge proof systems. *Journal of Cryptology* 7(1): 1-32 (1994).
- [214] 魏文君, 徐亨, 刘学清, 尤庆亮, 邹琳玲, 刘继延, 曹元成. 现代显示技术发展展望. *功能材料与器件学报* 21: 99-106 (2015).
- [215] Bruce Lane. Stereoscopic displays. *Proceedings of SPIE* 0367 (1982).
- [216] Jiang-Hao Xiong, En-Lin Hsiang, Zi-Qian He, Tao Zhan, Shin-Tson Wu. Augmented reality and virtual reality displays: emerging technologies and future perspectives, *Light: Science & Applications* 10:216 (2021).
- [217] De-Wen Cheng, Yong-Tian Wang, Hong Hua, M. M. Talha. Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism. *Applied Optics* 48:2655-2668 (2009).
- [218] Bernard C. Kress, Maria Pace. Holographic optics in planar optical systems for next generation small form factor mixed reality headsets. *Light: Advanced Manufacturing* 3: 55 (2022).
- [219] Suyeon Choi, Changwon Jang, Douglas Lanman, Gordon Wetzstein. Synthetic aperture waveguide holography for compact mixed-reality displays with large étendue. *Nature Photonics* 19: 854-863 (2025).
- [220] Hui Ren, Li-Xia Ni, Hai-Feng Li, Xin-Zhu Sang, Xin Gao, Qiong-Hua Wang. Review on tabletop true 3D display. *Journal of Society for Information Display* 28: 75-91 (2020).
- [221] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, Jitendra Malik. Reconstructing hands in 3d with transformers. *arXiv preprint arXiv: 2104.12167*, 2023.
- [222] Dixuan Lin, Yuxiang Zhang, Mengcheng Li, Yebin Liu, Wei Jing, Qi Yan, Qianying Wang, Hongwen Zhang. 4DHands: Reconstructing Interactive Hands in 4D with Transformers. *arXiv preprint arXiv: 2405.20330*, 2024.
- [223] Yumeng Liu, Xiaoxiao Long, Zemin Yang, Yuan Liu, Marc Habermann, Christian Theobalt, Yuexin Ma, Wenping Wang. EasyHOI: Unleashing the Power of Large Models for

Reconstructing Hand–Object Interactions in the Wild. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.

- [224] Apple Inc. Apple Vision Pro – Interaction. <https://www.apple.com/apple-vision-pro/>
- [225] Sizhen Bian, Mengxi Liu, Bo Zhou, Paul Lukowicz, Michele Magno. Body-Area Capacitive or Electric Field Sensing for Human Activity Recognition and Human-Computer Interaction: A Comprehensive Survey. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 8(1): 4:1-4:49 (2024).
- [226] Yufei Xu, Jing Zhang, Qiming Zhang, Dacheng Tao. ViTPose++: Vision Transformer for Generic Body Pose Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2024.
- [227] Nuri Murat Arar, Hua Gao, Jean-Philippe Thiran. Robust gaze estimation based on adaptive fusion of multiple cameras. IEEE International Conference and Workshops on Automatic Face and Gesture Recognition 2015: 1-7.
- [228] Jinglin Sun, Zhipeng Wu, Han Wang, Peiguang Jing, Yu Liu. A novel binocular eye-tracking system with stereo stimuli for 3d gaze estimation. arXiv preprint arXiv: 2104.12167, 2021.
- [229] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, Otmar Hilliges. Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. European Conference on Computer Vision (5) 2020: 365-381.
- [230] Yihua Cheng, Feng Lu. Dvgaze: Dual-view gaze estimation. IEEE/CVF International Conference on Computer Vision 2023: 20575-20584.
- [231] Jun Wang, Hao Ruan, Mingjie Wang, Chuanghui Zhang, Huachun Li, Jun Zhou. GazeCLIP: Towards Enhancing Gaze Estimation via Text-Guided Multimodal Learning. arXiv:2401.00260, <https://arxiv.org/abs/2401.00260>
- [232] Xiaochi Gu, Yifei Zhang, Weize Sun, Yuanzhe Bian, Dao Zhou, Per Ola Kristensson. Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR. CHI Conference on Human Factors in Computing Systems 2016: 1991-1995.
- [233] Haptx. <https://haptx.com>
- [234] SenseGlove. <https://www.senseglove.com/>

- [235] Ziqi Wang, Xiangjie Zhou, Zejian Zhou, Yan Zhang, Yuru Zhang, Dangxiao Wang. MateJam: Multi-Material Teeth-Clutching Layer Jamming Actuation for Soft Haptic Glove. *IEEE Transactions on Haptics* 16(2): 276-286 (2023).
- [236] Alexa F. Siu, Eric J. Gonzalez, Shenli Yuan, Jason B. Ginsberg, Sean Follmer. Shapeshift: 2D spatial manipulation and self-actuation of tabletop shape displays for tangible and haptic interaction. *CHI Conference on Human Factors in Computing Systems* 2018: 291.
- [237] Bruno Araújo, Ricardo Jota, Varun Perumal, JiaXian Yao, Karan Singh, Daniel Wigdor. Snake Charmer: Physically Enabling Virtual Objects. *International Conference on Tangible, Embedded, and Embodied Interaction* 2016: 218-226.
- [238] Victor Mercado, Maud Marchal, Anatole Lécuyer. ENTROPiA: Towards Infinite Surface Haptic Displays in Virtual Reality Using Encountered-Type Rotating Props. *IEEE Transactions on Visualization and Computer Graphics* 27(3): 2237-2243 (2021).
- [239] Kotaro Yamaguchi, Ginga Kato, Yoshihiro Kuroda, Kiyoshi Kiyokawa, Haruo Takemura. A Non-grounded and Encountered-type Haptic Display Using a Drone. *Symposium on Spatial User Interaction* 2016: 43-46.
- [240] Zhuang Zhang, Zhenghao Xu, Luoqian Emu, Pingdong Wei, Sentao Chen, Zirui Zhai, Lingyu Kong, Yong Wang, Hanqing Jiang. Active mechanical haptics with high-fidelity perceptions for immersive virtual reality. *Nature Machine Intelligence* 5(6): 643-655 (2023).
- [241] Margaret Koehler, Nathan S. Usevitch, Allison M. Okamura. Model-Based Design of a Soft 3-D Haptic Shape Display. *IEEE Transactions on Robotics* 36(3): 613-628 (2020).
- [242] Anthony Steed, Eyal Ofek, Mike Sinclair, Mar Gonzalez-Franco. A mechatronic shape display based on auxetic materials. *Nature Communications* 12(1): 4758 (2021).
- [243] Zemin Wang, Yan Zhang, Dongjie Zhao, Ruibo He, Yuru Zhang, Dangxiao Wang. Perceptually Inspired C0-Continuity Haptic Shape Display with Trichamber Soft Actuators. *Soft Robotics*, 2024.
- [244] Tadas Baltrušaitis, Chaitanya Ahuja, Louis-Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 41(2), 423-443, 2019.

- [245] Songning Lai, Xifeng Hu, Haoxuan Xu, Zhaoxia Ren, Zhi Liu. Multimodal sentiment analysis: A survey. *Displays*, 80, 102563, 2023.
- [246] Manju Priya Arthanarisamy Ramaswamy, Suja Palaniswamy. Multimodal emotion recognition: A comprehensive review, trends, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 14(6): e1563, 2024.
- [247] Rosalind W. Picard. *Affective Computing*. MIT Press, 1997.
- [248] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, Thomas S. Huang. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *Proceedings of the IEEE* 97(6): 101–123, 2009.
- [249] Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, Susanne Boll. Let's face it: Influence of facial expressions on social presence in collaborative virtual reality. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*: 1-16, 2023.
- [250] Vincent van Brakel, Miguel Barreda-Ángeles, Tilo Hartmann. Feelings of presence and perceived social support in social virtual reality platforms. *Computers in Human Behavior* 139: 107523, 2023.
- [251] Hui Ma, Jian Wang, Hongfei Lin, Bo Zhang, Yijia Zhang, Bo Xu. A transformer-based model with self-distillation for multimodal emotion recognition in conversations. *IEEE Transactions on Multimedia* 26: 776-788, 2023.
- [252] Dou Hu, Yinan Bao, Lingwei Wei, Wei Zhou, Songlin Hu. Supervised Adversarial Contrastive Learning for Emotion Recognition in Conversations. *The 61st Annual Meeting of the Association for Computational Linguistics*, 2023.
- [253] Geng Tu, Tian Xie, Bin Liang, Hongpeng Wang, Ruifeng Xu. Adaptive graph learning for multimodal conversational emotion detection. *Proceedings of the AAAI Conference on Artificial Intelligence* 38(17): 19089-19097, 2024.
- [254] Tao Wang, Shuang Liu, Feng He, Weina Dai, Minghao Du, Yufeng Ke, Dong Ming. Emotion recognition from full-body motion using multiscale spatio-temporal network. *IEEE Transactions on Affective Computing* 14(4): 2582-2594, 2023.
- [255] Rongyang Li, Jianguo Ding, Huansheng Ning. Emotion arousal assessment based on multimodal physiological signals for game users. *IEEE Transactions on Affective*

Computing 14(4): 2582-2594, 2023.

- [256] Gopendra Vikram Singh, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya. Emoint-trans: A multimodal transformer for identifying emotions and intents in social conversations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31: 290-300, 2022.
- [257] Rui Liu, Haolin Zuo, Zheng Lian, Xiaofen Xing, Björn W. Schuller, Haizhou Li. Emotion and intent joint understanding in multimodal conversation: A benchmarking dataset. *arXiv preprint arXiv: 2407.02751*, 2024.
- [258] Yousif Khairuddin, Zhuofa Chen. Facial emotion recognition: State of the art performance on FER2013. *arXiv preprint arXiv: 2105.03588*, 2021.
- [259] Siddique Latif, Rajib Rana, Sara Khalifa, Raja Jurdak, Junaid Qadir, Björn Schuller. Survey of deep representation learning for speech emotion recognition. *IEEE Transactions on Affective Computing* 14(2): 1634-1654, 2021.
- [260] Yuanyuan Lei, Houwei Cao. Audio-visual emotion recognition with preference learning based on intended and multi-modal perceived labels. *IEEE Transactions on Affective Computing* 14(4): 2954-2969, 2023.
- [261] Jiawen Deng, Fuji Ren. A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing* 14(1): 49-67, 2021.
- [262] Sicheng Zhao, Xingxu Yao, Jufeng Yang, Guoli Jia, Guiguang Ding, Tat-Seng Chua, Bjoern W. Schuller, Kurt Keutzer. Affective image content analysis: Two decades review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(10): 6729-6751, 2021.
- [263] Sana Ullah, Jie Ou, Yuanlun Xie, Wenhong Tian. Facial expression recognition (FER) survey: a vision, architectural elements, and future directions. *PeerJ Computer Science* 10: e2024, 2024.
- [264] Linan Zhu, Zhechao Zhu, Chenwei Zhang, Yifei Xu, Xiangjie Kong. Multimodal sentiment analysis based on fusion methods: A survey. *Information Fusion* 95: 306-325, 2023.
- [265] Ahmed Almohanadi, Shohei Yokoyama. Emotion mapping: sentiment analysis using emoji in twitter data from Japan in the COVID-19 Era. *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*: 666-669, 2023.

- [266] Toshiki Tomihira, Atsushi Otsuka, Akihiro Yamashita, Tetsuji Satoh. What does your tweet emotion mean? Neural emoji prediction for sentiment analysis. *Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services*: 289-296, 2018.
- [267] Shafkat Farabi, Tharindu Ranasinghe, Diptesh Kanojia, Yu Kong, Marcos Zampieri. A Survey of Multimodal Sarcasm Detection. *IJCAI* 2024.
- [268] Tyler Loakman, Aaron Maladry, Chenghua Lin. The Iron (ic) Melting Pot: Reviewing Human Evaluation in Humour, Irony and Sarcasm Generation. *Findings of the Association for Computational Linguistics: EMNLP*: 6676-6689, 2023.
- [269] Md Kaykobad Reza, Ashley Prater-Bennette, M. Salman Asif. Robust multimodal learning with missing modalities via parameter-efficient adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47(2):742-754, 2024.
- [270] Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, Gustavo Carneiro. Multi-modal learning with missing modality via shared-specific feature modelling. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 15878-15887, 2023.
- [271] Juhjung Sun, Sarah Lasser, Sun Kyong Lee. Understanding emojis: Cultural influences in interpretation and choice of emojis. *Journal of International and Intercultural Communication* 16(3): 242-261, 2023.
- [272] Isabelle Boutet, Megan LeBlanc, Justin A. Chamberland, Charles A. Collin. Emojis influence emotional communication, social attributions, and information processing. *Computers in Human Behavior* 119: 106722, 2021.
- [273] Peng Xu, Xiatian Zhu, David A. Clifton. Multimodal learning with transformers: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(10): 12113-12132, 2023.
- [274] Yuan Yuan, Zhaojian Li, Bin Zhao. A survey of multimodal learning: Methods, applications, and future. *ACM Computing Surveys* 57(7): 1-34, 2025.
- [275] Ankita Gandhi, Kinjal Adhvaryu, Soujanya Poria, Erik Cambria, Amir Hussain. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion* 91: 424-444,

2023.

- [276] Devamanyu Hazarika, Roger Zimmermann, Soujanya Poria. MISA: Modality-invariant and -specific representations for multimodal sentiment analysis. *Proceedings of the 28th ACM International Conference on Multimedia*: 1122-1131, 2020.
- [277] Changqin Huang, Jili Chen, Qionghao Huang, Shijin Wang, Yaxin Tu, Xiaodi Huang. AtCAF: Attention-based causality-aware fusion network for multimodal sentiment analysis. *Information Fusion* 114: 102725, 2025.
- [278] Jian Huang, Yanli Ji, Zhen Qin, Yang Yang, Heng Tao Shen. Dominant single-modal supplementary fusion (SIMSUF) for multimodal sentiment analysis. *IEEE Transactions on Multimedia* 26: 8383-8394, 2023.
- [279] Chul Min Lee, Shrikanth S. Narayanan. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing* 13(2): 293-303, 2005.
- [280] Arijit Nandi, Fatos Xhafa, Laia Subirats, Santi Fort. Real-time multimodal emotion classification system in e-learning context. *International Conference on Engineering Applications of Neural Networks*: 423-435, 2021.
- [281] Uzair Nadeem, Mohammed Bennamoun, Ferdous Sohel, Roberto Togneri. Learning-based confidence estimation for multi-modal classifier fusion. *International Conference on Neural Information Processing*: 299-312, 2019.
- [282] Johannes Wagner, Elisabeth Andre, Florian Lingenfelser, Jonghwa Kim. Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Transactions on Affective Computing* 2(4): 206-218, 2011.
- [283] Chengyan Wu, Yiqiang Cai, Yang Liu, Pengxu Zhu, Yun Xue, Ziwei Gong, Julia Hirschberg, Bolei Ma. Multimodal emotion recognition in conversations: A survey of methods, trends, challenges and prospects. *arXiv preprint arXiv: 2505.20511*, 2025.
- [284] Chuanguang Yang, Zhulin An, Linhang Cai, Yongjun Xu. Hierarchical self-supervised augmented knowledge distillation. *arXiv preprint arXiv: 2107.13715*, 2021.
- [285] ShiHao Zou, Xianying Huang, XuDong Shen, Hankai Liu. Improving multimodal fusion with main modal transformer for emotion recognition in conversation. *Knowledge-Based Systems* 258: 109978, 2022.
- [286] Shihao Zou, Xianying Huang, Xudong Shen. Multimodal prompt transformer with hybrid

- contrastive learning for emotion recognition in conversation. Proceedings of the 31st ACM International Conference on Multimedia: 5994-6003, 2023.
- [287] Zilong Huang, Man-Wai Mak, Kong Aik Lee. MM-NodeFormer: Node Transformer Multimodal Fusion for Emotion Recognition in Conversation. Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech: 4069-4073, 2024.
- [288] Xiaofei Zhu, Jiawei Cheng, Zhou Yang, Zhuo Chen, Qingyang Wang, Jianfeng Yao. CMATH: Cross-Modality Augmented Transformer with Hierarchical Variational Distillation for Multimodal Emotion Recognition in Conversation. arXiv preprint arXiv: 2411.10060, 2024.
- [289] Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, Alexander Gelbukh. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): 154-164, 2019.
- [290] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, Erik Cambria. DialogueRNN: An attentive RNN for emotion detection in conversations. Proceedings of the AAAI Conference on Artificial Intelligence 33(1): 6818-6825, 2019.
- [291] Jingwen Hu, Yuchen Liu, Jinming Zhao, Qin Jin. MMGCN: Multimodal Fusion via Deep Graph Convolution Network for Emotion Recognition in Conversation. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers): 5666-5675, 2021.
- [292] Jiaze Li, Hongyan Mei, Liyun Jia, Xing Zhang. Multimodal emotion recognition in conversation based on hypergraphs. Electronics 12(22): 4703, 2023.
- [293] Cam Van Thi Nguyen, Tuan Anh Mai, Dang Hai Kieu, Duc-Trong Le. Conversation Understanding using Relational Temporal Graph Neural Networks with Auxiliary Cross-Modality Interaction. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing 2023.

- [294] Junwei Feng, Xueyan Fan. Cross-modal Context Fusion and Adaptive Graph Convolutional Network for Multimodal Conversational Emotion Recognition. arXiv preprint arXiv: 2501.15063, 2025.
- [295] Jian Huang, Yuanyuan Pu, Dongming Zhou, Jinde Cao, Jinjing Gu, Zhengpeng Zhao, Dan Xu. Dynamic hypergraph convolutional network for multimodal sentiment analysis. *Neurocomputing* 565: 126992, 2024.
- [296] Songtao Li, Hao Tang. Multimodal alignment and fusion: A survey. arXiv preprint arXiv: 2411.17040, 2024.
- [297] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, Yue Gao. Hypergraph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence* 33(1): 3558-3565, 2019.
- [298] Cuong Tran Van, Thanh V. T. Tran, Van Nguyen, Truong Son Hy. Effective Context Modeling Framework for Emotion Recognition in Conversations. *ICASSP 2025 IEEE International Conference on Acoustics, Speech and Signal Processing*: 1-5, 2025.
- [299] Nian Liu, Xiao Wang, Deyu Bo, Chuan Shi, Jian Pei. Revisiting graph contrastive learning from the perspective of graph spectrum. *Advances in Neural Information Processing Systems* 35: 2972-2983, 2022.
- [300] Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, Zhendong Niu. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in Neural Information Processing Systems* 36: 69638-69660, 2023.
- [301] Wei Ai, Fuchen Zhang, Yuntao Shou, Tao Meng, Haowen Chen, Keqin Li. Revisiting Multimodal Emotion Recognition in Conversation from the Perspective of Graph Spectrum. *Proceedings of the AAAI Conference on Artificial Intelligence* 39(11): 11418-11426, 2025.
- [302] Aaron van den Oord, Yazhe Li, Oriol Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv: 1807.03748, 2018.
- [303] Libo Qin, Qiguang Chen, Xiachong Feng, Yang Wu, Yongheng Zhang, Yinghui Li, Min Li, Wanxiang Che, Philip S. Yu. Large language models meet NLP: A survey. arXiv preprint arXiv: 2405.12819, 2024.
- [304] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad

- Usman, Naveed Akhtar, Nick Barnes, Ajmal Mian. A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology* 16(5): 1-72, 2025.
- [305] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, Philip S. Yu. Multimodal large language models: A survey. 2023 IEEE International Conference on Big Data (BigData): 2247-2256, 2023.
- [306] Wenxuan Zhang, Mahani Aljunied, Chang Gao, Yew Ken Chia, Lidong Bing. M3Exam: A multilingual, multimodal, multilevel benchmark for examining large language models. *Advances in Neural Information Processing Systems* 36: 5484-5505, 2023.
- [307] Jintao Tong, Shiwei Li, Zijian Zhuang, Jinghan Hu, Yixiong Zou. EmoSync: Multi-Stage Reasoning with Multimodal Large Language Models for Fine-Grained Emotion Recognition. *Proceedings of the 3rd International Workshop on Multimodal and Responsible Affective Computing* 2025.
- [308] Zhengxuan Zhang, Jianying Chen, Xuejie Liu, Weixing Mai, Qianhua Cai. ‘What’ and ‘where’ both matter: Dual cross-modal graph convolutional networks for multimodal named entity recognition. *International Journal of Machine Learning and Cybernetics* 15(6): 2399-2409, 2024.
- [309] Soumya Dutta, Sriram Ganapathy. LLM supervised pre-training for multimodal emotion recognition in conversations. *ICASSP 2025 IEEE International Conference on Acoustics, Speech and Signal Processing*: 1-5, 2025.
- [310] Yumeng Fu, Junjie Wu, Zhongjie Wang, Meishan Zhang, Lili Shan, Yulin Wu, Bingquan Liu. LaERC-S: Improving LLM-based emotion recognition in conversation with speaker characteristics. *Proceedings of the 31st International Conference on Computational Linguistics*: 6748-6761, 2025.
- [311] Lin Song, Yukang Chen, Shuai Yang, Xiaohan Ding, Yixiao Ge, Ying-Cong Chen, Ying Shan. Low-rank approximation for sparse attention in multi-modal LLMs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 13763-13773, 2024.
- [312] Guimin Hu, Ting-En Lin, Yi Zhao, Guangming Lu, Yuchuan Wu, Yongbin Li. UniMSE: Towards unified multimodal sentiment analysis and emotion recognition. *arXiv preprint*

arXiv: 2211.11256, 2022.

- [313] Zaijing Li, Ting-En Lin, Yuchuan Wu, Meng Liu, Fengxiao Tang, Ming Zhao, Yongbin Li. UniSA: Unified generative framework for sentiment analysis. Proceedings of the 31st ACM International Conference on Multimedia: 6132-6142, 2023.
- [314] Yang Yang, Xunde Dong, Yupeng Qiang. MSE-Adapter: A Lightweight Plugin Endowing LLMs with the Capability to Perform Multimodal Sentiment Analysis and Emotion Recognition. Proceedings of the AAAI Conference on Artificial Intelligence 39(24): 25642-25650, 2025.
- [315] Zehui Wu, Ziwei Gong, Lin Ai, Pengyuan Shi, Kaan Donbekci, Julia Hirschberg. Beyond silent letters: Amplifying LLMs in emotion recognition with vocal nuances. Findings of the Association for Computational Linguistics: NAACL: 2202-2218, 2025.
- [316] Zheng Lian, Haiyang Sun, Licai Sun, Hao Gu, Zhuofan Wen, Siyuan Zhang, Shun Chen, et al. Explainable multimodal emotion recognition. arXiv preprint arXiv: 2306.15401, 2023.
- [317] Zheng Lian, Haiyang Sun, Licai Sun, Haoyu Chen, Lan Chen, Hao Gu, Zhuofan Wen, et al. OV-MER: Towards Open-Vocabulary Multimodal Emotion Recognition. Proceedings of the 42nd International Conference on Machine Learning 2025.
- [318] Zheng Lian, Haiyang Sun, Licai Sun, Jiangyan Yi, Bin Liu, Jianhua Tao. AffectGPT: Dataset and framework for explainable multimodal emotion recognition. arXiv preprint arXiv: 2407.07653, 2024.
- [319] 世界超高清视频产业联盟（UWA），元宇宙时代超高清视音频技术白皮书，2023.
- [320] 中国网，北京市聚焦高精尖产业需求 布局构建新型共性技术平台支撑体系，2023. http://www.china.com.cn/zhibo/content_85047436.htm.
- [321] 第十一届中国网络视听大会，AIGC 赋能元宇宙数字内容产业变革，2024. <https://zhuanlan.zhihu.com/p/690311031>.
- [322] 陈永伟，《第二人生》：一个早期元宇宙的兴衰，经济观察报，2021. https://m.thepaper.cn/baijiahao_15951912
- [323] Mu'ad Diib, Revised EVE Activities & Careers Chart Project - Work in Progress, <https://forums.eveonline.com/t/revised-eve-activities-careers-chart-project-work-in-progress/298585>
- [324] IT 之家，《EVE Online》与微软合作推出 Excel 插件，轻松管理游戏数据，

<https://www.ithome.com/0/701/127.htm>

[325] <https://eve.huijiwiki.com/wiki/EVE> 国服编年史

[326] EVE Online: A Metaverse Blueprint, <https://vimeo.com/505913778>

[327] 网易伏羲, 揭秘元宇宙: 当下已有的惊艳应用案例.

<https://yaotai.163.com/cs/news/ZLlZ-Y0Bm1p-cBRUK8QP.html>

[328] 范佳来, 元宇宙是个什么宇宙?

https://www.thepaper.cn/newsDetail_forward_15813148, 澎湃新闻, 2021

[329] Rajiv B, Reimagining How People Come Together Through Communication, Connection, and Expression, <https://blog.roblox.com/2023/10/reimagining-bringing-people-together-communication-connection-expression/>

[330] 地平线 HOA, 如何看待刘慈欣说「扎克伯格的元宇宙不是未来, 元宇宙最后将引人类走向死路」?, <https://daily.zhihu.com/story/9742803>

[331] 中金公司, 元宇宙: 空间升维、时间延展、社会重构

<https://research.cicc.com/frontend/recommend/detail?id=2976>

[332] 俞勇, 元宇宙金融提前布局探析, 金融时报-中国金融新闻网, 2023. <https://weibo.com/ttarticle/p/show?id=2309404925545707340325>

[333] 中关村互联网金融研究院, 工商银行推出 VR 元宇宙营业厅, 2022. https://mp.weixin.qq.com/s?__biz=MjM5ODI0MTI2Ng==&mid=2650537364&idx=2&sn=60eb3c8894f5a4f347b0c1acfada7fca

[334] 《加快数智供应链发展专项行动计划》

https://www.gov.cn/zhengce/zhengceku/202505/content_7025257.htm

[335] 京东 11.11 直播技术全面升级 立影 3D 技术、JoyAI 大模型重构沉浸式购物体验, 新华网, 2025.10.25. <http://tibet.news.cn/20230825/302e276a93a74125b6222a8748d72496/c.html>

[336] 《Agricultural Metaverse: Key Technologies, Application Scenarios, Challen 京东物流发布“超脑大模型 2.0”, 新华网, 2025.09.2. <https://www.news.cn/auto/20250926/d37949e7deb94cb18e6c93828560b6e0/c.html>

[337] 京东 11.11 超三万“数字员工”上岗, 超级供应链下 AI 激发创新生产力, 中国经济网, 2025.11.11. http://finance.ce.cn/home/jrzq/dc/202511/t20251111_2573208.shtml

- [338] 京东发布 JoyInside 2.0 贴身智能平台 为重构人机交互体验提供新可能, 新华网, 2025.09.27.
<http://bj.news.cn/20250927/c36034eb4bc8467ab28cd13a433dc949/c.html>
- [339] <https://www.wur.nl/nl/onderzoek-resultaten/kennisonline-onderzoeksprojecten-lvvn/kennisonline/transition-to-a-data-driven-agriculturen-ttadda-for-a-new-dutch-japanese-potato-circular-value-chain.htm>
- [340] <https://www.wur.nl/nl/onderzoek-resultaten/kennisonline-onderzoeksprojecten-lvvn/kennisonline/transition-to-a-data-driven-agriculturen-ttadda-for-a-new-dutch-japanese-potato-circular-value-chain.htm>
- [341] Digital twin technology in smart agriculture: Enhancing productivity and sustainability, <https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9780443288845000154>
- [342] 陈枫,孙传恒,邢斌,等.农业元宇宙:关键技术、应用情景、挑战与展望[J].智慧农业(中英文),2022,4(04):126-137.
- [343] 《Advances in 3D simulation technology for solar greenhouse systems: A review》, <https://www.sciencedirect.com/science/article/abs/pii/S1364032125010603?via%3Dihub>
- [344] <https://www.wur.nl/en/research-results/research-institutes/plant-research/business-units/greenhouse-horticulture/show-greenhouse/agros-ii-project-launched.htm>
- [345] <https://www.wur.nl/en/newsarticle/fewer-humans-more-ai-is-this-the-future-of-the-agricultural-and-horticultural-sectors.htm?sessionid=>
- [346] 《元宇宙安全治理上海倡议》
- [347] 李晟,徐春朋.元宇宙场景中的数据安全与治理:基于数据三维结构范式.金融客 12: 40-44 (2022).
- [348] 傅琳雅, 邬慧.元宇宙视域下个人信息保护的困境及法规重塑.西南石油大学学报(社会科学版) 26(3): 61-68 (2024).
- [349] 袁苗. 元宇宙虚拟现实中的大数据隐私保护与安全防范.信息记录材料 25(04) (2024)
- [350] 高一乘, 杨东.应对元宇宙挑战: 数据安全综合治理三维结构范式.行政管理改革.
- [351] 中国现代国际关系研究院.元宇宙与国家安全.北京:中国现代国际关系研究院, 2021.
- [352] 曼纽尔·卡斯特.千年终结.夏铸九,黄慧琦,译.北京:社会科学文献出版社 416, 2003.

- [353] 严小十. 意识形态安全视角下的元宇宙空间治理初探. 江南社会学院学报 24(2):29-33 (2022).
- [354] 王瞰. 从互联网到元宇宙: 数字空间的意识形态安全审视. 贵州社会科学 2: 35-43 (2023).
- [355] 刘亦石, 周亚建, 崔莹, 等. 人工智能大模型应用中的安全问题与解决策略. 网络空间安全科学学报 2(1): 83-91 (2024).
- [356] 邱心语. 社交型人工智能的网络失范行为及责任承担. 网络空间安全 15(2): 87-93 (2024).
- [357] 袁苗. 元宇宙虚拟现实中的大数据隐私保护与安全防范. 信息记录材料 25(04) (2024).
- [358] 全球数据治理: 跨境数据流动的现状与困境, Dispute Settlement 9(5) (2023).
- [359] 跨境数据流动治理: 框架、实践困境, 《中国信息安全》杂志 2023 年第 10 期
- [360] Hoofnagle, Chris Jay, Bart Van Der Sloot, Frederik Zuiderveen Borgesius. The European Union general data protection regulation: what it is and what it means. Information & Communications Technology Law 28(1): 65-98 (2019).
- [361] Hargittai, Eszter, Jason Schultz, John Palfrey. Why parents help their children lie to Facebook about age: Unintended consequences of the 'Children's Online Privacy Protection Act'. First Monday (2011).

编写人员贡献

白皮书的编写组成员包括：陶建华（清华大学）、杨小康（上海交通大学）、何斌（同济大学）、王茜莺（联想）、鲍虎军（浙江大学）、王琼华（北京航空航天大学）、祝烈煌（北京理工大学）、何晓冬（京东）、李琳（咪咕文化）、杨艺（凌云光）。

在撰写过程中，除编写组成员外，还得到（以下按拼音序）陈春明（咪咕新空）、陈磊（清华大学）、储繁（北京航空航天大学）、邓冰冰（咪咕新空）、邓昊天（北京理工大学）、丁毅（上海健顺投资管理有限公司）、杜华（元客视界）、范存航（安徽大学）、方斌（北京邮电大学）、顾心怡（北京理工大学）、郭宏蕾（清华大学）、赖南华（京东）、李丰（摩尔线程）、李会（清华大学）、林浩（咪咕新空）、廖智勇（咪咕新空）、刘瑞（内蒙古大学）、刘鑫辰（京东）、刘云辉（联想）、吕建阳（咪咕新空）、吕琬军（联想）、盛兴东（联想）、苏铖（咪咕文化）、唐舒梦（咪咕新空）、王党校（北京航空航天大学）、王迪（北京航空航天大学）、王乐（咪咕新空）、温正祺（清华大学）、许江毅（咪咕新空）、许威威（浙江大学）、杨定康（咪咕新空）、杨天威（清华大学）、易江燕（清华大学）、喻纯（清华大学）、于涛（清华大学）、章广涛（清华大学）、张川（北京理工大学）、张洪涛（联想）、张小磊（咪咕新空）、赵佳晨（清华大学）、赵思成（清华大学）、郑四海（咪咕新空）、朱佳伟（咪咕新空）等人帮助。

其中：陶建华、杨小康、何斌负责了白皮书的框架设计、整体撰写和修订工作。陶建华、李琳、苏铖、李会、杨天威、章广涛、郭宏蕾、赵思成、于涛、范存航、丁毅、李丰、吕琬军、杜华参与了第一章、第六章的撰写和

修订；陶建华、鲍虎军、王琼华、祝烈煌、许威威、杜华、丁毅、吕琬军、刘云辉、范存航、盛兴东、张川、王迪、喻纯、易江燕、温正祺、方斌、章广涛、于涛、赵思成、郭宏蕾、李会、王党校、李丰、储繁、刘瑞、邓昊天参与了第二章的撰写和修订；王茜莺、李琳、刘云辉、王乐、林浩、张小磊、杜华、张洪涛、陈春明、许江毅、杨定康、邓冰冰、吕琬军、许威威、丁毅、李丰参与了第三章的撰写和修订；廖智勇、杨艺、何晓冬、朱佳伟、苏铖、郑四海、吕建阳、唐舒梦、邱兴平、刘云辉、杜华、刘鑫辰、张洪涛、赖南华、李丰、丁毅参与了第四章的撰写和修订；顾心怡、李会、赵佳晨参与了第五章的撰写和修订。