



中移智库

数智人行业发展概述

中国移动研究院 业务研究所
2024年3月

摘要

数智人，业界也将其称为虚拟人、数字人，是具有数字化外形的虚拟人物，借助计算机图形学、动作捕捉、人工智能等计算机手段创造，依赖显示设备存在，拥有人的相貌、行为及思维[1]。作为一种多技术融合的虚实交互媒介，数智人与人工智能、虚拟现实、元宇宙等新型产业发展密切相关，并持续受到各方关注，而相关政策的出台更为行业快速发展注入动力。2022年1月国务院发布的《“十四五”数字经济发展规划》中指出要加强类人智能自然交互与虚拟现实等技术研究。2022年8月，北京市经信局发布国内首个数字人产业专项支持政策——《北京市促进数字人产业创新发展行动计划(2022-2025年)》，并提出目标到2025年，北京市数字人产业规模突破500亿元。

本文通过阐述数智人主流产品类别、商业模式及行业

发展进程，分析了数智人技术发展的六大最新趋势，并同时对行业发展现状进行总结与分析。

一、 主流产品类型和商业模式分析

业内主流数智人产品按照功能应用维度可划分为 IP 型、服务型、身份型三种，分别对应 IP 运营、客户服务和流量变现三种商业模式[2]。

（一）IP 型数智人

IP 型数智人产品定位为具有情感导向、价值内涵和圈层影响力的虚拟偶像和品牌形象大使，第一需求方是影视游戏、文娱公司等内容生产 B 端厂商。目前相关技术基本成熟，已有较多成功商用案例，如洛天依、AYAYI、柳叶熙、翎 Ling 等虚拟偶像，和以 5G 冰雪推广大使 Meet Gu 为代表的中国移动数智人家族，其商业价值及社会影响力已得到验证。

由于用户对 IP 型数智人外观的精细度和写实度要求较

高，这类产品的形象风格从最初的二次元，逐渐发展为美型偏写实、超写实风格。生产技术路线早期以手工原画设计为主，之后主要是基于相机阵列扫描等专业采集设备的高精度 3D 建模配合 CG 模型精修，形象制作成本从 50 到 100 万元不等，制作周期需 2-3 个月，人工依赖程度高、成本高、耗时长、产能低。面向影视、营销场景的 IP 型数智人对交互实时性要求较低，且通常依赖美术特效团队进行后期加工，其驱动方式以采用专业动捕场地及设备的真人驱动为主。

IP 型数智人实现商业变现的核心在于 IP 运营，包括完善人设、提升建模精美程度、持续产出多元化内容等。其主要收入渠道有：（1）品牌推广、直播带货等营销收入；（2）商演、直播、影视参演、专辑发售等形象授权和内容创作收入；（3）游戏、动漫、手办模型等周边衍生收入。

（二）服务型数智人

服务型数智人通过交付给企业客户实现变现，帮助企业对外提供自动化、智能化、标准化且支持 7*24 小时的行业服务，起到对人工的增强和辅助作用，助力企业降本增效。相关技术已部分成熟，但由于这类产品大多不直接创造收入，商业价值仍处于待验证阶段。

相比 IP 型数智人，服务型数智人的外形要求相对较低，但对交互的流畅性、行业知识的丰富程度要求较高，因此其生产不仅包含外形建模，还涉及行业知识库及知识图谱的构建，对技术与产业的整合升级提出了更高要求。为满足企业降低成本的需要，功能服务型数智人的建模方式也更倾向于采用轻量化的 AI 算法进行人物 2D/3D 建模，驱动方式也从真人驱动逐渐转向全自动的 AI 驱动。

服务型数智人已在金融、文旅行业积累较多应用成果，

如浦发银行和百度共同发布数字员工“小浦”，中国文物交流中心与百度智能云打造的虚拟文物解说员“文夭夭”等，这类产品后续将在医疗、教育等具有流程属性的行业有较大应用潜力。例如面向医疗行业，满足就医引导、诊疗辅助、患者情感陪伴等需求，将衍生出虚拟导诊、家庭陪护、健康顾问等专业化数智人产品；面向教育行业，未来可能会出现虚拟教师、虚拟作业陪写员等应用。

（三）身份型数智人

身份型数智人相关技术和产品正在探索中，尚无规模化应用案例，待产业进一步完善，目标是实现用户利用平台工具制作千人千面的虚拟形象并能够在虚拟场景内进行社交、会议、演出观看等活动，平台厂商通过对虚拟空间进行规则设定来实现流量变现。身份型数智人旨在面向广大C端用户提供现实世界与虚拟空间之间全面感知和链接的

入口与载体，其核心属性是身份和交互。这要求在技术上实现通过简便快捷的数据采集，结合低成本、全自动的建模/生成技术来高效制作形象，并利用实时视觉动捕等技术使用户能够自然地驱动形象在虚拟空间中活动并与其他人实时互动，这进一步提高了对算法、算力、网络传输、软硬件基础和技术集成能力的要求。目前，大规模多人互动及虚拟空间相关技术和产品还在发展中，百度希壤等元宇宙应用中已经出现了个性化形象生成功能，ZEPETO、SOUL等虚拟社交产品中也有相关应用，市场潜力较大但商业模式仍不明确。

二、 行业应用发展进程

业界将数智人应用发展按时间维度大致划分为三个阶段[3]:

（一）前期探索阶段（2010-2015）

在行业发展初期，IP型数智人在泛娱乐场景率先落地，

拉动技术进步的同时，向社会各界普及虚拟人的概念，持续形成宣传与示范效应。

2007年日本推出了史上第一个被广泛认可的虚拟歌手初音未来，经过10余年运营已在全球70个国家拥有6亿粉丝，代言上百家品牌，其商业价值已超百亿日元。国内第一虚拟歌手洛天依从2012年开始活跃在动漫“二次元”圈层，于2021年登上了央视春晚，标志着虚拟偶像逐渐在主流媒体中争得一席之地。洛天依“出圈”的背后是因其开创了一个基于同人创作圈的运营模式，其推出的上万首作品，绝大多数由其粉丝兼音乐人投稿完成。基于“VOCALOID”歌声合成引擎，每个人都可以通过为洛天依赋予形象或故事来实现自我表达，这使粉丝与洛天依之间建立起一种超越真人偶像的情感于价值连接。

（二）产业融合发展阶段（2015-2021）

跟随娱乐行业的步伐，金融、文旅、快消等其他垂直产业应用跟进，开始出现为政府、企业、品牌对外提供公共服务的数智人解决方案和平台工具，且产品核心价值逐渐

向两个方向分化：一是专业化，体现为细分领域的专业虚拟顾问；二是情感化，强调与人之间的情感互动，满足陪护、督促等需求[4]。

腾讯智影提供了低成本数字人播报 SaaS 工具，以虚拟数字人为核心打造了文本配音、文章转视频、模板创作、在线视频剪辑等功能。平台收费模式包括付费会员和定制付费两种，会员可在 2D/3D 数字人模板库中选择形象制作播报内容；如果用户需要生成和特定真人类似的数字人或数字分身，可以采购其定制服务，上传照片、视频、音频即可实现定制。百度智能云曦灵面向金融、媒体、互娱等行业，提供服务型、演艺型数字人解决方案，可应用在手机 APP 端、云屏端，扮演智能客服、数字理财经理、数字商品导购、数字培训师、数字讲解员等角色，实现人机可视化语音交互服务。此外，搜狗、魔法科技、相芯科技均有提供相应数智人技术服务的平台。

提升数智人情感互动的典型案例是 OPPO 发布的手机端智能服务助手小布 4.0。当用户与小布进行语音对话时，小布

能创造真实自然的聊天环境，响应用户情感并反馈体现在动作、表情、音调中；当用户点击它身体不同的部位，例如摸头、摸手、摸肚子等交互动作时，小布也会做出对应的身体动作反应。

另外，在这一阶段，**IP 型数智人的应用领域不断拓宽，运营方式及商业模式趋向多元化，传播力和社会影响力进一步扩大[5]**。2021 年，中国移动推出的超写实格斗数智达人尤子希亮相咪咕演播室，为观众带来了精彩的 UFC 赛事解说首秀。2022 年，尤子希以元宇宙 UFC 的特派员身份开启首场 MMA 健身直播，与粉丝实时互动聊天。凭借精心制作的 3A 级逼真形象和性格爱好、形式逻辑等完整人物设定，尤子希作为虚拟 IP 的可塑性和虚实互动的空间进一步提升，在体育、时尚领域影响力持续增，入选了清华大学新闻与传播学院新媒体研究中心发布的“2022 年虚拟数字人综合指数排名 TOP10”榜单[6]。

（三）多模态智能发展阶段（2021-至今）

伴随着以 ChatGPT 为代表的 AIGC、大模型技术的引入，

数智人技术进入了高速发展期。语音、图片、视频等多模态元素的融合扩展了数智人感知理解的维度，提升了其交互内容和交互形式的丰富度，因而数智人的创造力、自动化程度和真人替代能力也进一步提升，极大地拓宽了数智人在多个领域的应用潜力。

数字助手微软小冰从第五代产品开始已具备了多模态生成交互能力，融合了全双工语音交互、实时视觉与核心对话引擎，实现用户与人工智能同时边听边说边看的交互体验。基于歌声合成、文生图等艺术家模型，小冰集诗人、歌手、主持人、画家和设计师多种功能于一身，通过不断学习迭代来实现与人类艺术家同等质量水准的创作。

另外，在此阶段，数智人生产能力也朝着智能化、便捷化、精细化、多样化发展。2021年，Epic发布Metahuman Creator，基于云端的混合型变计算和渲染，支持用户在网页端创建高保真的3D虚拟形象，并结合苹果ArKit等实时3D面部关键点捕捉工具生成面部动画。

（四）未来发展趋势

根据市场分析机构预测[7]，2025年我国虚拟人带动产业市场规模和核心市场规模将分别达到6402.7亿元和480.6亿元。从当前市场构成方面，数智人生产和运营的B端市场占比最高，达到了79%，并且规模不断扩大。其中，2023年企业用户购买最多的数智人产品类型是数字员工，占比60.2%，其次是自定义数字人，占比为53.3%，为明星等名人定制数字形象的需求占比为38.6%。

随着技术进步和市场认知提升，数智人的技术门槛和软硬件成本进一步下降，数智人服务逐渐向工具化和系统化演进。面向广大C端用户的身份型数智人的应用潜力巨大，与元宇宙概念结合，将形成更开放的创作生态和全新商业模式。在可预见的未来，更多支持制作二次元、三维写实超写实等风格数智人的开放工具和平台将充分满足C端用户生产、管理、运营形象资产和虚拟身份的需求；借助成熟

的视频、语音等轻量级多模态驱动技术，只需手机等移动端的单摄像头和算力就能实时获得驱动数智人所需的高精度动作和表情数据；再将虚拟化身与各类虚拟空间平台相结合，实现高自由度的、深度沉浸的 6DoF 空间交互通信，人们互联互通方式将再次升级。

三、 行业技术发展方向

第一，产学研共同推动低成本轻量化算法覆盖数智人全流程，促进 C 端应用成熟。近年来涌现出一系列轻量化计算机图形学技术，如隐式三维重建、实时神经渲染等，推动三维人物建模的实时性、自动化、精细度提升。仅需单目/多目便携式设备采集的视频数据进行训练就能全自动创建人物模型，无需依赖特殊采集环境及设备 and 手工建模，大幅降低了数智人生产成本。2023 年 8 月，学术界提出了一种新方法实现 3D 辐射场的实时渲染，即 3D Gaussian

Splatting (3DGS) [8]，该方法相比已有的隐式神经辐射场技术，进一步突破了三维人物/场景的渲染速度和渲染质量的边界。苹果团队 2023 年 12 月发表的最新工作 [9] 引入 3DGS 来表示可驱动的人体及场景，只需要 2 到 4 秒的视频（50-100 帧），就能生成完整的人体模型，并对皮肤和衣服、头发等细节进行精细建模，速度比当前其他同类技术（NeuMan[10]、Vid2Avatar[11]等）快约 100 倍。

另外，2022 年兴起的 text-to-3D 生成式 AI 技术 [12]，通过收集大规模 3D 数据库和对应的文本描述训练扩散模型，只需输入文本提示词，完全避免了用户侧的数据采集步骤，高效率、低门槛地生成 3D 模型。影眸科技推出的 ChatAvatar [13] 和微软 RODIN [14] 就基于 text-to-3D 技术只需要简单的文本提示词描述外貌特征，就能在快速生成 3D 数字人资产，并且模型带有规整的拓扑、材质贴图 and 绑

定数据，可以直接接入 Unity、Unreal Engine 和 Maya 等制作引擎的生产管线中。

在驱动方面，视觉、语音驱动算法的优化推动专业动捕设备趋向便捷化，面部捕捉技术向更简单的硬件、更细腻的表情、更自动化的流程方向发展。已有学术及产业界团队提出采用 AI 生成式模型[15]，基于语音生成节奏一致性的人体面部口型和肢体动作序列来驱动人物三维模型。这类技术将逐步替代需要穿戴动捕设备的真人驱动方式，只需要音频或视觉特征点数据输入，对驱动端的硬件要求很低，有效降低了数智人的使用门槛。

第二，在线可交互式数智人将替代离线非交互式数智人。

在行业技术早期发展阶段，基于离线渲染、后期制作的非交互类型数智人是主流，互联网内容分发平台将离线制作的包含虚拟偶像的数字内容融入到现有的业务逻辑中，在

数字化营销、文娱等领域广泛应用。随着数智人概念的普及和 AI 驱动技术快速发展，用户很快进入体验突破阶段，在电商直播、数字服务领域已有使用实时交互和实时渲染的数智人作为新型交互方式的应用案例。未来，以元宇宙发展为契机，实时 AI 驱动、渲染传输等技术门槛进一步降低，社交、电商、娱乐、VR 应用等会出现以在线可交互式数智人为入口，与数字场景深度耦合的全新的业务逻辑，并结合更深的行业理解，带来更真实的沉浸感和更丰富的商业价值。

第三，数智人在不断发展“形似人”的基础上更加“神似人”。“神似人”是要求数智人具备类似人的分析思考能力，包括学习、记忆、推理、决策，并具备理解反馈和传达情感的能力。目前大多数数智人具备的初级沟通功能本质是以 NLP 为核心的文本驱动，通过 ASR-NLP-TTS 的 AI

管线实现数智人感知、理解、决策、表达。其中，支撑数智人分析思考的NLP技术也将向专业化和情感化两个方向进化：向专业化方向发展，将多行业知识图谱与数智人对话系统对接，应用于特定场景；向情感化方向发展，将从人物性格、语言风格、知识储备等方面还原真人级互动体验。未来ASR-NLP-TTS技术将在文本语义交互的基础上，强化对文本中情感/意图的理解，并实现与语义和情感一致的语音情感、微表情和人物姿态的智能合成。另外，多模态AIGC技术的出现将成为数智人拥有思想和自主创造能力的最大推动力，在输入端使数智人具备多模态输入的感知理解能力，在输出端使数智人能够进行多模态交互和自动化内容生产。

第四，数智人与空间计算加速融合。随着数智人的交互维度升级，交互信息媒介将从一维的语音数据到二维的图

像、视频数据，再拓展到虚实空间的位置关系、路径、姿态等三维数据。数智人作为一种虚拟的三维存在形式，势必将与其他三维数字空间技术如 SLAM（即时定位与地图构建）、体积视频、空间音频等融合，以人为原点构建完整的虚实交互技术体系，实现真人在虚拟空间中的多方互动漫游，虚拟/真实空间中真人和数字人的交互，数字人对真实空间的感知等。

从 2021 年起，百度希壤等沉浸式虚拟社交应用初步搭建了中心化分发的多人虚拟漫游产品框架。但这类应用目前仅支持对固定虚拟场景的多视角观看，难以实现对空间内的虚拟物体进行操控或进行个性化空间定制，单人活动和多人互动的模式单一；由于 VR 头显、手机等端侧算力有限，部分场景采用贴图等 2.5D 技术构建，人物建模也采用低精度的卡通捏脸方式，以减轻端侧渲染压力。立体呈现效果

的粗糙和操控方式的受限导致了沉浸感降低。未来要实现数智人在虚拟空间内进行多点 6DoF 同步观看和对空间内物体的任意操控，人/物/空间关系与行为分析、编码传输、端边云协同渲染等关键能力亟待攻关，空间感知、空间交互处理、立体呈现技术栈仍需逐渐完善。

第五，数智人显示载体仍以 2D 显示设备为主，3D 显示与数智人将形成相互助推，并出现多设备协同联动的数智人产品。目前，PC、手机、智慧大屏、银行 VTM 及等 2D 主流显示设备仍是数智人显示的主要载体。数智人和 3D 数字内容的应用发展，与裸眼 3D、VR/AR、全息等 3D 显示设备的普及可形成相互助推。一方面，依托 3D 显示更能突显数智人沉浸逼真的特点，为数智人在文娱社交等特定领域带来新的增长空间。例如，湖南卫视元宇宙产品“芒果幻城”计划在 PICO、NOLO 等 VR 平台上构建虚拟秀场、虚拟演唱会，

使用户以数字形态空降节目现场，实现与虚拟明星零距离接触的沉浸式“追星”体验。苹果 Vision Pro 头显搭载的 Persona 应用，可基于高逼真度虚拟化身进行视频通话，这些探索拓展了数智人的更多“新玩法”。另一方面，借助已有数智人 IP/产品的流量和用户粘性，针对性地提供更多基于 3D 显示的创意数字内容和解决方案，也为 3D 显示带来探索多元化业务和产业链延伸的契机。2022 年，中国移动 Meet Gu 亮相北京西单商圈、王府井步行街和上海静安寺的户外裸眼 3D 大屏，吸引广大体育爱好者争相“打卡”，在超写实高精度人物建模与裸眼 3D 技术双重加持下，为大众带来了耳目一新的沉浸互体验，促进了裸眼 3D 概念普及和“文化+科技”全新业务形态探索。

第六，数智人可信保护技术及行业合规体系亟待完善。

数智人行业总体在快速发展阶段，随着产业规模扩大、数

智人资产数量增多、生产门槛降低，数智人可信安全及版权保护的问题日益凸显。2022年7月，杭州某网络公司在网络上发布两段视频，居中位置使用了某数字人厂商于2019年发布的视频内容，并在片头片尾替换有关标识，添加了该公司的注册商标添及虚拟数字人课程的营销信息。2023年，杭州互联网法院就首例涉“虚拟数字人”侵权案作出一审判决，认定被告杭州某网络公司构成著作权侵权及不正当竞争，判决其承担消除影响并赔偿损失12万元的法律责任。

数智人在数据表示层面具有多模态和异构化特点，其外形存在点云，网格，体素、隐式表达等表征形式，其声音包括音色、语调、口音等信息，其运动包括骨骼绑定参数等驱动信息，甚至可能具有面容、指纹等生物特征/身份信息。这导致了针对数智人的数字侵权形式更加多样化和复

杂化，已有的传统鉴权技术难以满足数智人资产可信保护的需求。目前，业界尚未形成公认统一技术和标准体系来保障数智人资产的安全流通与共享，亟需进一步研究数智人数据安全、版权的确权鉴权技术，涉及自然人肖像、声音等生物特征/数字身份的保护策略，AI 驱动数智人生成虚拟内容的可控鉴别措施等。

四、 总结与思考

综合来看，数智人技术正在向轻量化、智能化、多元化发展，但数智人资产的生产效率仍待提高。2D 数智人由于建模数据采集便捷，技术成熟度、稳定性高，相关商用实例较多，而具备成熟的写实 3D 数智人技术的厂商仍然较少，且大部分仍需要 CG 美术团队加持，面向 To B 的少量定制化需求，暂时难以满足全自动化、高机动性、高频次的规模应用需求。目前，行业中大多数公司聚焦于数智人全流程

上的一环或几环，如资产制作和策划运营，面向直播营销内容生产等单一领域应用，门槛较低、可替代性较高。具备技术和场景的头部厂商的优势在于综合技术服务能力和分布式算力等基础设施，但这类公司基本处于试验产品可行性、依托自身优势场景打造标杆应用的阶段。如何促进数智人产业链各个节点的密切协同，构建统一完善的行业标准，是进一步释放产能、打破技术壁垒的关键。

[参考文献]

[1] 《2020年虚拟数字人发展白皮书》，中国人工智能产业发展联盟。

[2] “Digital Avatar Market Report 2023 - 2030”，Grand View Research, Report ID: GVR-4-68040-083-2, June 6, 2023, <https://www.grandviewresearch.com/industry-analysis/digital-avatar-market-report>.

[3] 《虚拟数字人研究报告 2.0 版》，清华大学新闻与传播学院新媒体研究中心，2022。

[4] 《虚拟数字人研究报告：溯源、应用、发展》，清华大学新闻与传播学院新媒体研究中心，2022。

[5] 《中国虚拟数字人影响力指数报告(2022年度)》，中国传媒大学媒体融合与传播国家重点实验室媒体大数据研究中心。

[6] 《2022虚拟数字人综合评估指数报告》，清华大学新闻与传播学院新媒体研究中心。

- [7] 《2023 年中国虚拟人产业发展与商业趋势研究报告》，艾媒咨询。
- [8] Kerbl, Bernhard, et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." *ACM Transactions on Graphics* 42.4, 2023.
- [9] Kocabas, Muhammed, et al. "Hugs: Human gaussian splats." *arXiv preprint arXiv:2311.17910*, 2023.
- [10] Jiang, Wei, et al. "Neuman: Neural human radiance field from a single video." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
- [11] Guo, Chen, et al. "Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition." *Proceedings of the*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

[12] Poole, Ben, et al. "Dreamfusion: Text-to-3d using 2d diffusion." arXiv preprint arXiv:2209.14988, 2022.

[13] Zhang, Longwen, et al. "DreamFace: Progressive Generation of Animatable 3D Faces under Text Guidance." arXiv preprint arXiv:2304.03117, 2023.

[14] Wang, Tengfei, et al. "Rodin: A generative model for sculpting 3d digital avatars using diffusion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

[15] Zhang, Fan, et al. "DiffMotion: Speech-driven gesture synthesis using denoising diffusion model." International Conference on Multimedia Modeling. Cham: Springer International Publishing, 2023.

审稿：某某 | 部所

本文作者



中移智库

谢柠蔚，就职于中国移动研究院，主要从事多媒体处理、视觉 AI、元宇宙等领域研究工作。