



2020 年虚拟数字人发展 白皮书

中国人工智能产业发展联盟总体组

中关村数智人工智能产业联盟数字人工作委员会

2020 年 12 月

版权声明

本白皮书版权属于中国人工智能产业发展联盟总体组和中关村数智人工智能产业联盟数字人工作委员会，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源：中国人工智能产业发展联盟总体组和中关村数智人工智能产业联盟数字人工作委员会”。违反上述声明者，将追究其相关法律责任。

致 谢

本白皮书是在中国人工智能产业发展联盟总体组和中关村数智人工智能产业联盟数字人工作委员会指导和组织下，由中国信息通信研究院云计算与大数据研究所牵头，中国工商银行股份有限公司、上海浦东发展银行、凌云光技术股份有限公司、南京新一代人工智能研究院联合参与起草，有关内容得到了网易（杭州）网络有限公司、深圳市腾讯计算机系统有限公司、华为技术有限公司、魔珐(上海)信息科技有限公司、量子动力（深圳）计算机科技有限公司、北京聚力维度科技有限公司等单位的支持。限于编写时间、起草人员知识积累与产业尚未完全定型等方面的因素，内容恐有疏漏，烦请不吝指正。

主要起草人：颜媚、石霖、张丹、孙明俊、杨艺、熊伟、曾义、张孟熙、齐凡、刘博文、罗涛、施佳子、李锋、李旭佳

前言

在人工智能、虚拟现实等新技术浪潮的带动下，虚拟数字人制作过程得到有效简化、各方面性能获得飞跃式提升，开始从外观的数字化逐渐深入到行为的交互化、思想的智能化。以虚拟主播、虚拟员工等为代表的数字人成功进入大众视野，并以多元的姿态在影视、游戏、传媒、文旅、金融等众多领域大放异彩。

为加快推动虚拟数字人技术和产业创新发展，中国人工智能产业发展联盟（以下简称“**AIIA** 联盟”）总体组协同中关村数智人工智能产业联盟（以下简称“**ZAI** 联盟”）数字人工作委员会于 2020 年发起虚拟数字人推进计划，致力于开展虚拟数字人技术研究、标准制定、评估测试、合作交流、成果发布及生态建设工作。本白皮书是 **AIIA** 联盟总体组和 **ZAI** 联盟数字人工作委员会首次发布。白皮书回顾了虚拟数字人发展历程，重点分析了虚拟数字人关键技术和产业发展现状，对虚拟数字人的未来发展趋势进行了展望，剖析了数字人发展中的制约因素，提出产业发展建议。

目录

前言.....	1
一、 虚拟数字人概述.....	1
(一) 虚拟数字人研究范畴.....	1
(二) 虚拟数字人发展历程.....	1
(三) 当前虚拟数字人通用系统框架及运作流程.....	4
二、 虚拟数字人关键技术趋势.....	7
(一) 虚拟数字人技术架构.....	7
(二) 建模：静态扫描建模仍为主流，包含人行为动作且具有高视觉保真度的动态光场三维重建初见端倪.....	8
(三) 驱动：智能合成、动作捕捉迁移成为数字人目前主流的动作生产方式.....	9
(四) 渲染：随着硬件能力的提升和算法的突破，真实性和实时性均大幅提升.....	11
三、 虚拟数字人产业应用现状.....	14
(一) 虚拟数字人产业视图.....	14
(二) 虚拟数字人重点领域应用环境及典型应用案例分析.....	16
四、 虚拟数字人发展趋势、制约因素及建议.....	23
(一) 虚拟数字人未来发展趋势.....	23
(二) 虚拟数字人发展制约因素.....	23
(三) 虚拟数字人产业发展建议.....	25
附件：虚拟数字人典型企业介绍.....	28

一、虚拟数字人概述

（一）虚拟数字人研究范畴

“虚拟数字人”一词最早源于 1989 年美国国立医学图书馆发起的“可视人计划”(Visible Human Project, YHP)。2001 年,国内以“中国数字化虚拟人体的科技问题”为主题的香山科学会议第 174 次学术讨论会提出了“数字化虚拟人体”的概念。这些“虚拟数字人”主要是指人体结构的可视化,以三维形式显示人体解剖结构的大小、形状、位置及器官间的相互空间关系,即利用人体信息,实现人体解剖结构的数字化。主要应用于医疗领域的人体解剖教学、临床诊疗等。

与上述医疗领域的数字化人体不同,本篇中所分析的虚拟数字人(以下简称“数字人”)是指具有数字化外形的虚拟人物。与具备实体的机器人不同,虚拟数字人依赖显示设备存在。虚拟数字人宜具备以下三方面特征:一是拥有人的外观,具有特定的相貌、性别和性格等人物特征;二是拥有人的行为,具有用语言、面部表情和肢体动作表达的能力;三是拥有人的思想,具有识别外界环境、并能与人交流互动的能力。

（二）虚拟数字人发展历程

虚拟数字人的发展与其制作技术的进步密不可分,从最早的手工绘制到现在的 CG(Computer Graphics, 电脑绘图)、人工智能合成,虚拟数字人大致经历了萌芽、探索、初级和成长四个阶段,详见图 1。



图 1 虚拟数字人发展历程

20 世纪 80 年代，人们开始尝试将虚拟人物引入到现实世界中，虚拟数字人步入萌芽阶段。该时期虚拟数字人的制作技术以手工绘制为主，应用极为有限。1982 年，日本动画《超时空要塞》播出后，制作方将女主角林明美包装成演唱动画插曲的歌手，并制作了音乐专辑，该专辑成功打入当时日本知名的音乐排行榜 Oricon，林明美也成为了世界上第一位虚拟歌姬。1984 年，英国人 George Stone 创作出一个名为 Max Headroom 的虚拟人物，MAX 拥有人类的样貌和表情动作，身穿西装，佩戴墨镜，曾参演了一部电影，拍摄了数支广告，一度成为英国家喻户晓的虚拟演员。由于技术的限制，其虚拟形象是由真人演员通过特效化妆和手绘实现。

21 世纪初，传统手绘逐渐被 CG、动作捕捉等技术取代，虚拟数字人步入探索阶段。该阶段的虚拟数字人开始达到实用水平，但造价不菲，主要出现在影视娱乐行业，如数字替身、虚拟偶像等。电影制作中的数字替身一般利用动作捕捉

技术，真人演员穿着动作捕捉服装，脸上点上表情捕捉点，通过摄像机、动作捕捉设备将真人演员的动作、表情采集处理，经计算机处理后赋予给虚拟角色。2001 年，《指环王》中的角色咕噜就是由 CG 技术和动作捕捉技术产生，这些技术后续还在《加勒比海盗》、《猩球崛起》等电影制作中使用。2007 年，日本制作了第一个被广泛认可的虚拟数字人“初音未来”，初音未来是二次元风格的少女偶像，早期的人物形象主要利用 CG 技术合成，人物声音采用雅马哈的 VOCALOID¹ 系列语音合成，呈现形式还相对粗糙。

近五年，得益于深度学习算法的突破，数字人的制作过程得到有效简化，虚拟数字人开始步入正轨，进入初级阶段。该时期人工智能成为虚拟数字人不可分割的工具，智能驱动的数字人开始崭露头角。2018 年，新华社与搜狗联合发布的“AI 合成主播”，可在用户输入新闻文本后，在屏幕展现虚拟数字人形象并进行新闻播报，且唇形动作能与播报声音实时同步。2019 年，浦发银行和百度共同发布的数字员工“小浦”，也是利用自然语言处理、语音识别、计算机视觉等人工智能技术制作的虚拟数字人，可通过移动设备为用户提供“面对面”的银行业务服务。

当前，虚拟数字人正朝着智能化、便捷化、精细化、多样化发展，步入成长期。2019 年，美国影视特效公司数字王

¹ VOCALOID（日语：ボーカロイド）是日本乐器制造商雅马哈公司开发的电子音乐制作语音合成软件。在软件中输入音调和歌词，就可以合成出原为人类声音的歌声。

国软件研发部负责人 Doug Roble 在 TED 演讲时展示了自己的虚拟数字人“DigiDoug”，可在照片写实级逼真程度的前提下，进行实时的表情动作捕捉及展现。今年，三星旗下的 STAR Labs 在 CES 国际消费电子展上展出了其虚拟数字人项目 NEON，NEON 是一种由人工智能所驱动的虚拟人物，拥有近似真人的形象及逼真的表情动作，具备表达情感和沟通交流的能力。

（三）当前虚拟数字人通用系统框架及运作流程

当前虚拟数字人作为新一代人机交互平台，仍处于发展期，还未有统一的通用系统框架。本白皮书根据虚拟数字人的制作技术以及目前市场上提供的数字人服务和产品结构，总结出虚拟数字人通用系统框架，如图 2 所示。虚拟数字人系统一般情况下由人物形象、语音生成、动画生成、音视频合成显示、交互等 5 个模块构成。人物形象根据人物图形资源的维度，可分为 2D 和 3D 两大类，从外形上又可分为卡通、拟人、写实、超写实等风格；语音生成模块和动画生成模块可分别基于文本生成对应的人物语音以及与之相匹配的人物动画；音视频合成显示模块将语音和动画合成视频，再显示给用户。交互模块使数字人具备交互功能，即通过语音语义识别等智能技术识别用户的意图，并根据用户当前意图决定数字人后续的语音和动作，驱动人物开启下一轮交互。

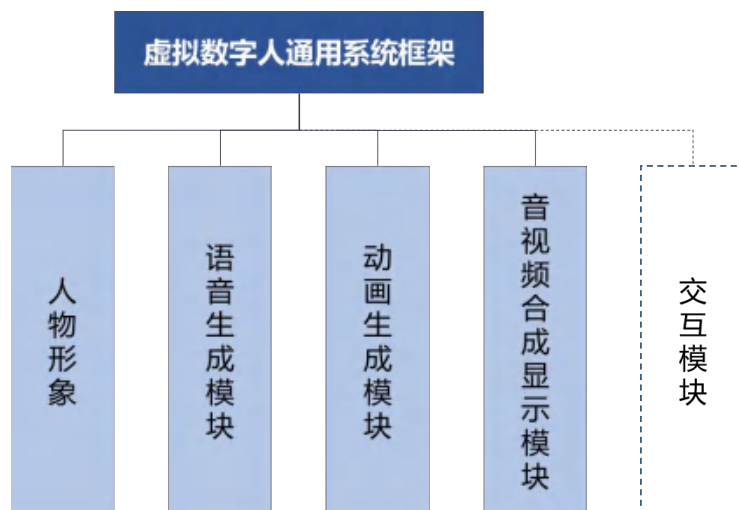


图2 虚拟数字人通用系统框架

交互模块为扩展项，根据其有无，可将数字人分为交互型数字人和非交互型数字人。非交互型数字人系统的运作流程如图3所示。系统依据目标文本生成对应的人物语音及动画，并合成音视频呈现给用户。

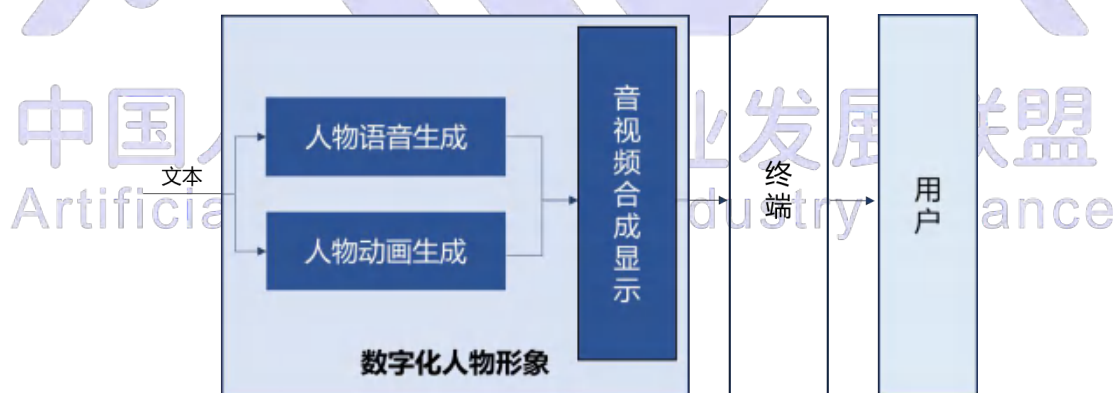


图3 非交互类虚拟数字人系统运作流程

交互型数字人根据驱动方式的不同可分为智能驱动型和真人驱动型。智能驱动型数字人可通过智能系统自动读取并解析识别外界输入信息，根据解析结果决策数字人后续的输出文本，然后驱动人物模型生成相应的语音与动作来使数字人跟用户互动。该人物模型是预先通过AI技术训练得到

二、虚拟数字人关键技术趋势

（一）虚拟数字人技术架构

当前，虚拟数字人的制作涉及众多技术领域，且制作方式尚未完全定型，通过对现有虚拟数字人制作中涉及的常用技术进行调研，本白皮书在虚拟数字人通用系统框架的基础上提炼出五横两纵的技术架构，如图 6 所示。“五横”是指用于数字人制作、交互的五大技术模块，即人物生成、人物表达、合成显示、识别感知、分析决策等模块。其中，人物表达包括语音生成和动画生成。动画生成则包含驱动（动作生成）和渲染两大部分。“两纵”是指 2D、3D 数字人，3D 数字人需要额外使用三维建模技术生成数字形象，信息维度增加，所需的计算量更大。



图 6 虚拟数字人技术架构

本白皮书重点关注虚拟数字人制作过程涉及的建模、驱动、渲染三大关键技术。

（二）建模：静态扫描建模仍为主流，包含人行为动作且具有高视觉保真度的动态光场三维重建初见端倪

相机阵列扫描重建替代结构光扫描重建成为人物建模主流方式。早期的静态建模技术以结构光扫描重建为主。结构光扫描重建可以实现 0.1 毫米级的扫描重建精度，但其扫描时间长，一般在 1 秒以上，甚至达到分钟级，对于人体这类运动目标在友好度和适应性方面都差强人意，因此被更多的应用于工业生产、检测领域。近年来，拍照式相机阵列扫描重建得到飞速发展，目前可实现毫秒级高速拍照扫描（高性能的相机阵列精度可达到亚毫米级），满足数字人扫描重建需求，成为当前人物建模主流方式。国际上 IR、Ten24 等公司已经将静态重建技术完全商业化，服务于好莱坞大型影视数字人制作，国内凌云光等公司制作的拍照式人体扫描系统也已经在电影、游戏、虚拟主播项目中成功应用。

相比静态重建技术，动态光场重建不仅可以重建人物的几何模型，还可一次性获取动态的人物模型数据，并高品质重现不同视角下观看人体的光影效果，成为数字人建模重点发展方向。动态光场重建是目前世界上最新的深度扫描技术，此技术可忽略材质，直接采集三维世界的光线，然后实时渲染出真实的动态表演者模型，它主要包含人体动态三维重建和光场成像两部分。人体动态三维重建一直是计算机视觉、计算机图形学等领域研究的重点，主要采用摄像机阵列

采集动态数据，可重建高低频几何、纹理、材质、三维运动信息。光场成像是计算摄像学领域一项新兴技术，它不同于现有仅展示物体表面光照情况的 2D 光线地图，光场可以存储空间中所有光线的方向和角度，从而产出场景中所有表面的反射和阴影，这为人体三维重建提供了更加丰富的图像信息。近年来 Microsoft、Google、Intel、Facebook 等公司都在积极展开相关研究，其中 Microsoft 的 108 摄像机 MRstudio 已经在全球各大洲均有建设；Google 的 Relightable 系统将结构光、动态建模、重光照技术集成到一起，在一套系统中包含模型重建、动作重建、光照重建的全部功能；国内清华大学、商汤科技、华为等也展开了相关研究，并取得国际水平的同步进展。

（三）驱动：智能合成、动作捕捉迁移成为数字人目前主流的动作生产方式

2D、3D 数字人均已实现嘴型动作的智能合成，其他身体部位的动作目前还只支持录播。2D、3D 数字人嘴型动作智能合成的底层逻辑是类似的，都是建立输入文本到输出音频与输出视觉信息的关联映射，主要是对已采集到的文本到语音和嘴型视频（2D）/嘴型动画（3D）的数据进行模型训练，得到一个输入任意文本都可以驱动嘴型的模型，再通过模型智能合成。然而，2D 视频和 3D 嘴型动画底层的数学表达不一样，2D 视频是像素表达；3D 嘴型动画是 3D 模型

对应的 BlendShape 的向量表达。除了嘴型之外的动作，包含眨眼、微点头、挑眉等动画目前都是通过采用一种随机策略或某个脚本策略将预录好的视频/3D 动作进行循环播放来实现。例如 3D 肢体动作目前就是通过某个位置触发这个预录好的肢体动作数据得到。触发策略是通过人手动配置得到的，未来希望通过智能分析文本，学习人类的表达，实现自动配置。截至目前，国内外科技企业在数字人动作智能合成方面都有一定进展，国际上如 Reallusion 公司研究的利用语音生成面部表情的 Craytalk 技术已在动画制作中被成功商用，国内搜狗、相芯科技等公司也有部分项目落地应用。

通过将捕捉采集的动作迁移至数字人是目前 3D 数字人动作生成的主要方式，核心技术是动作捕捉。动作捕捉技术按照实现方式的不同，可分为光学式、惯性式、电磁式及基于计算机视觉的动作捕捉等。现阶段，光学式和惯性式动作捕捉占据主导地位，基于计算机视觉的动作捕捉成为聚焦热点。**光学动作捕捉**通过对目标上特定光点的监视和跟踪来完成运动捕捉的任务。最常用的是基于 **Marker**（马克点）的光学动作捕捉，即在演员身上粘贴能够反射红外光的马克点，通过摄像头对反光马克点的追踪，来对演员的动作进行捕捉。这种方式对动作的捕捉精度高，但对环境要求也高，并且造价高昂。光学式解决方案比较出名的企业有英国的 **Vicon**，美国的 **OptiTrack** 和魔神（**Motion Analysis**），国内

的 Nokov、uSens、青瞳视觉等。惯性动作捕捉主要是基于惯性测量单元（Inertial Measurement Unit, IMU）来完成对人体动作的捕捉，即把集成了加速度计、陀螺仪和磁力计的 IMU 绑在人体的特定骨骼节点上，通过算法对测量数值进行计算，从而完成动作捕捉。这种惯性动作捕捉方案价格相对低廉，但精度较低，会随着连续使用时间的增加产生累积误差，发生位置漂移。惯性式动捕方案的代表性企业有荷兰的 Xsens，以及国内的诺亦腾（Noitom）、幻境、国承万通等。基于计算机视觉的动作捕捉主要是通过采集及计算深度信息来完成对动作的捕捉，是近些年才兴起的技术。这种视觉动捕方式因其简单、易用、低价，已成为目前使用的频率较高的动作捕捉方案，代表性产品有 Leap Motion、微软 Kinect 等。以上动捕方案的性能对比如表 1 所示。

表 1 主流动作捕捉方案性能对比

方案	精度	抗遮挡	对环境要求	硬件成本	算法开发难度	应用公司
光学捕捉	高	低	高	高	中	Vicon、OptiTrack
惯性捕捉	低	高	低	中	低	Xsens、Noitom
视觉捕捉	低	低	低	低	高	Leap Motion、微软 Kinect

（四）渲染：随着硬件能力的提升和算法的突破，真实性和实时性均大幅提升

PBR (Physically Based Rendering, 基于物理的渲染技术)

渲染技术的进步以及重光照等新型渲染技术的出现使数字人皮肤纹理变得真实，突破了恐怖谷效应。恐怖谷效应由日本机器人专家森政弘提出，认为人们对机器人的亲和度随着其仿真程度增加而增高，但当达到一个较高的临界点时，亲和度会突然跌入谷底，产生排斥、恐惧和困惑等负面心理。数字人恐怖谷效应主要由数字人外表、表情动作上与真人的差异带来，其中外表真实感的关键就是皮肤材质的真实感，无论是塑料感还是蜡像感都会给人类带来不适。在 PBR 技术出现之前，限于相关软硬件的发展程度，所有的 3D 渲染引擎，更多的着重在于实现 3D 效果，在真实感体现方面差强人意。PBR 是基于真实物理世界的成像规律模拟的一类渲染技术的集合，它的关键在于微表面模型和能量守恒计算，通过更真实的反映模型表面反射光线和折射光线的强弱，使得渲染效果突破了塑料感。目前常见的几款 3D 引擎，如 UnrealEngine 4, CryEngine 3, Unity 3D 5，均有了各自的 PBR 实现。重光照技术通过采集模拟多种光照条件的图像数据，测算数字人表面光照反射特性，并合成出数字人模型在新的光照下的渲染结果，使计算机中的虚拟数字人在任意虚拟环境下都可以呈现近乎真实的效果，它彻底改变了传统渲染方式通过模拟皮肤复杂的透射反射来计算渲染总会带来误差的局面。该技术在 2000 年初由南加州大学实验室创建 LightStage 平台时提出，并开始了相关研究，目前已经经过 7 代的迭代发展，已被成功应用到《阿凡达》、《复仇者联盟》等众多经典影片的角色制作中。国内清华大学、浙江大学也

都建设了重光照系统，可以实现高精度人体光照采集与重建。

实时渲染技术的突破助力写实类数字人实现实时交互，应用范围快速扩大。实时渲染指图形数据的实时计算与输出，其每一帧都是针对当时实际的环境光源、相机位置和材质参数计算出来的图像。与离线渲染相比，实时渲染面临较大挑战。一是渲染时长短，实时渲染每秒至少要渲染 30 帧，即在 33 毫秒内完成一帧画面渲染，离线渲染则可以花费数小时甚至更长时间渲染一帧画面；二是计算资源有限，实时渲染受限于时效要求，计算资源一般是不能及时调整，而离线渲染受时效限制较低，可临时调配更多的计算资源。早期的实时渲染只能选择高度抽象和简化过的渲染算法，牺牲了画面质量。随着硬件能力的提升和算法的突破，渲染速度、渲染效果的真实度、渲染画面的分辨率均大幅提升，在虚拟人物实时渲染方面，已经能做到以假乱真。2016 年，Epic Games 联合 3Lateral、Cubic Motion、Ninja Theory 等公司联合开发的可实时驱动虚拟人物在当年的 Siggraph (Special Interest Group for Computer GRAPHICS，计算机图形图像特别兴趣小组，致力于推广和发展计算机绘图和动画制作的软硬件技术) 会议中做了演示，成功在消费级的硬件环境下实时渲染了高质量的虚拟角色。2018 年 5 月，腾讯发布虚拟人 Siren，也是一个支持实时渲染的虚拟人物。

电视、投影、LED 显示等 2D 显示设备，也包括裸眼立体、AR、VR 等 3D 显示设备。**光学器件**用于视觉传感器、用户显示器的制作。**传感器**用于数字人原始数据及用户数据的采集。**芯片**用于传感器数据预处理和数字人模型渲染、AI 计算。**建模软件**能够对虚拟数字人的人体、衣物进行三维建模。**渲染引擎**能够对灯光、毛发、衣物等进行渲染，主流引擎包括 Unity Technologies 公司的 Unity 3D、Epic Games 公司的 Unreal Engine 等。总体来看，处于基础层的厂商已经深耕行业多年，已经形成了较为深厚的技术壁垒。

(2) 平台层

平台层包括软硬件系统、生产技术服务平台、AI 能力平台，为虚拟数字人的制作及开发提供技术能力。建模系统和动作捕捉系统通过产业链上游的传感器、光学器件等硬件获取真人/实物的各类信息，利用软件算法实现对人物的建模、动作的重现；渲染平台用于模型的云端渲染。解决方案平台基于自身技术能力为广大客户提供数字人解决方案。AI 能力平台提供计算机视觉、智能语音、自然语言处理技术能力。平台层汇聚的企业较多，腾讯、百度、搜狗、魔法科技、相芯科技均有提供相应数字人技术服务平台。

(3) 应用层

应用层是指虚拟数字人技术结合实际应用场景领域，切入各类，形成行业应用解决方案，赋能行业领域。按照应用

场景或行业的不同，已经出现了娱乐型数字人（如虚拟主播、虚拟偶像）、教育型数字人（如虚拟教师）、助手型数字人（如虚拟客服、虚拟导游、智能助手）、影视数字人（如替身演员或虚拟演员）等。不同外形、不同功能的虚拟数字人赋能影视、传媒、游戏、金融、文旅等领域，根据需求为用户提供定制化服务。

（二）虚拟数字人重点领域应用环境及典型应用案例分析

虚拟数字人应用给传统领域带来变革。通过虚拟数字人产品与生产生活相融合，其规模化、可定制化、可复制化的能力能够推动改善传统环节流程、提升效能、降低成本等，大幅提升业务体验，给传统领域带来变革。典型的应用领域、场景及充当的角色如表 3 所示。

表 3 虚拟数字人应用领域、场景及角色

领域	场景	角色	
泛娱乐	影视	数字替身特效可以帮助导演实现现实拍摄中无法表现的内容和效果，已成为特效商业大片拍摄中的重要技术手段和卖点。	数字替身
	传媒	定制化虚拟主持人/主播/偶像，支持从音频/文本内容一键生成视频，实现节目内容快速、自动化生产，打造品牌特有 IP 形象，实现观众互动，优化观看体验。	虚拟主持人 虚拟主播 虚拟偶像
	游戏	越来越真实的数字人游戏角色使游戏者有了更强的代入感，可玩性变得更强。	数字角色
金融	通过智能理财顾问、智能客服等角色，实现以客户为中心的、智能高效的人性化服务。	智能客服 智能理财顾问	
文旅	博物馆、科技馆、主题乐园、名人故居等虚拟小剧场、虚拟导游、虚拟讲解员。	虚拟导游 虚拟讲解员	
教育	基于 VR/AR 的场景式教育，虚拟导师帮助构	虚拟导师	

	建自适应/个性化学习环境。	
医疗	以数字人实现家庭陪护/家庭医生/心理咨询，实时关注家庭成员身心健康，并及时提供应对建议。	心理医生 家庭医生
零售	从大屏到机器人到全息空间，从数据分析、个性营销、智能货架、无人商店四大应用场景切入构建线下零售服务新流程。电商直播在虚拟数字人与真人主播和观众互动,介绍商品。	顾客服务数字人 商家管理数字人 虚拟主播

(1) 虚拟数字人+影视

特效电影广受市场认可，扶持政策的密集出台，以数字替身为代表的虚拟数字人+影视类的产品应用正在展现广阔的市场前景。影视是对视觉效果的要求最高且最大程度影响社会对数字人形象认知、品牌的一个领域。近年来中国影视数字人特效取得快速发展，部分特效大片获得市场认可。2019 年中国影视票房收入超过 640 亿级别，连续多年增长 8%左右，其中特效电影约占 10%。国家对于影视特效的发展十分重视，先后出台了一系列相关扶持政策。2019 年，科技部、文化部联合六部委发布《关于促进文化和科技深度融合的指导意见》，提出加强激光放映、虚拟现实、光学捕捉、影视摄录、高清制播、图像编辑等高端文化装备自主研发及产业化。2020 年，国家电影局《关于促进科幻电影发展的若干意见》，提出以科幻电影特效技术发展引领带动电影特效水平整体提升，并要落实财税支持政策，同时对入驻影视文化基地的科技企业进行租金减免。于 2019 年 2 月 22 日在中国内地上映的《阿丽塔：战斗天使》是虚拟数字人技术与影

视相结合的典型应用案例之一。剧中的女主角阿丽塔是一位完全采用数字人技术制作的角色。该电影通过采用特殊的面部捕捉仪器对真人演员人脸细节进行精准捕捉，然后将其作为电脑中虚拟角色的运动依据，使虚拟角色的动作和表情能像真人一样自然逼真。

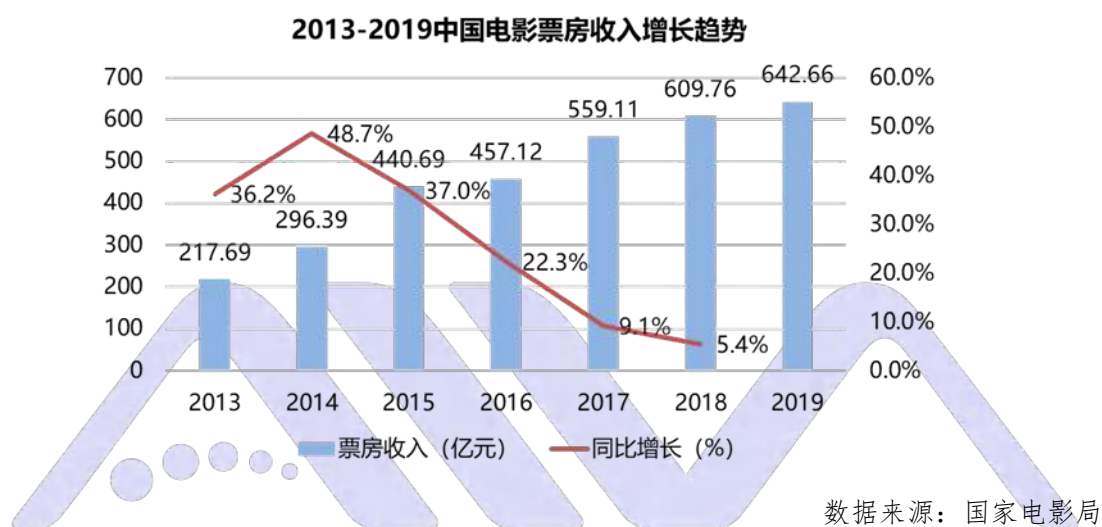


图 8 2013-2019 中国电影票房收入增长趋势



数据来源：国家电影事业发展专项资金管理委员会办公室

图 9 中国内地电影票房总收入 TOP10

(2) 虚拟数字人+传媒

以虚拟主播为代表传媒行业应用很好的满足了媒体传播领域对内容生成方面的业务需求，成为了融媒体时代的传媒利器。据前瞻产业研究院统计，2019 年中国视频直播行业收入达到 1082 亿人民币，全国有 3.9 亿人在关注虚拟偶像，其中最大的二次元活动社区B站月活用户达 1140 万，虚拟主播占直播营收 40%。同期，B站、虎牙等大型直播平台也都开启了线上、线下虚拟主播活动。如Bilibili Macro Link系列活动VR直播观看如数达到 660 万，10 万级的“up主”（指在视频网站、论坛、FTP站点上传视频音频文件的人）就可以达到 100 万级真人主播的营收。此外，在传统媒体领域，以虚拟主持人为代表的数字人应用也开始进入公众视野。在 2019 年央视网络春晚舞台上，以撒贝宁为原型制作的AI虚拟主持人“小小撒”首次上岗与原型同台秀技。这也是国内第一次人类主持人和自己的虚拟孪生数字人共同主持大规模的国家级文化活动，引起了业界及网民的强烈关注。

（3）虚拟数字人+游戏

游戏市场日趋激烈，对于精品的需求日益增长，或是虚拟数字人技术在该领域落地的福音。2019 年游戏市场规模达到 2300 亿，其中与数字人相关度较大的RPG类游戏约占 30% 的份额。我国国内游戏监管环境和游戏版号审批政策仍旧趋严，游戏用户数相比 2018 年仅增加 0.1 亿，市场竞争日趋激烈，对于精品的需求越来越高。虚拟数字人技术则能够有效

简化和加快游戏动画制作过程，可以在有限成本的情况下让游戏中更多的虚拟角色都具备丰富的肢体动作和精细的面部表情，给玩家带来更沉浸的游戏体验。例如网易伏羲实验室成功将虚拟数字人技术应用到《逆水寒》等多个游戏剧情动画场景制作中，在无手工参与下，快速生成动画，这使得大量的虚拟角色都能富有视觉的表达能力。

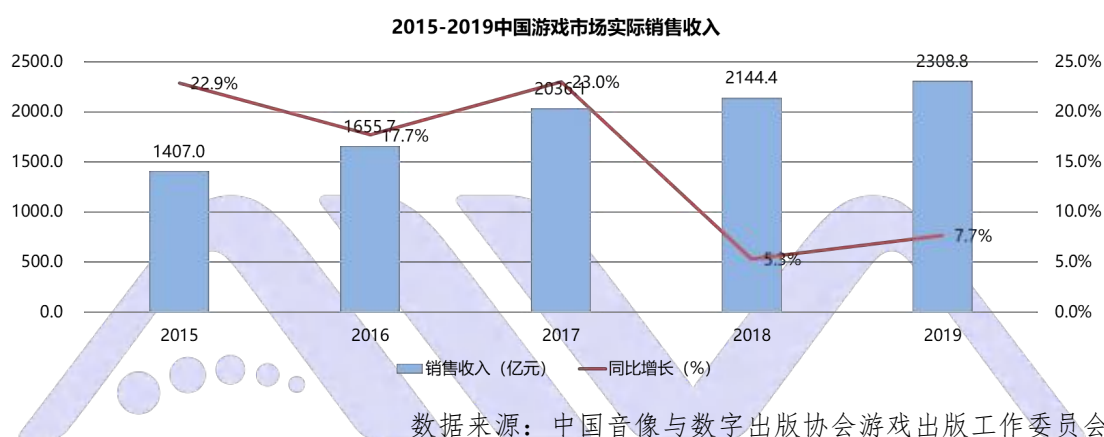


图 10 2015-2019 中国游戏市场实际销售收入

(4) 虚拟数字人+金融

虚拟数字人能够有效助力金融机构实现“降本增效”，各大金融巨头争相布局“数字员工”。运用金融科技实现“降本增效”应对市场竞争成为金融行业发展共识，以科技为核心的竞争导向，带来了金融机构信息科技投入的逐年增加。据轻金融统计，2019年，16家全国性银行科技总投入超千亿，总计1034.1亿元。其中工农中建四大行投入均超过100亿元，建行投入最高，达到176.33亿元。虚拟数字人拥有拟人的表情动作，可进行智能对话，能够与服务类场景较多的金融行业天然结合。多家金融机构正利用虚拟数字人技术打造“数

字员工”，成为了科技创新、降本增效的重要方向。以中国工商银行为例，其推出的数字人银行员工，在移动终端和大屏等媒介上展现数字拟态形象，能够在产品营销讲解、金融业务办理、资讯播报、咨询问答等多个业务场景，实现与用户的可视化交互，为用户带来个性化服务，有效解除用户顾虑，提升用户体验和驻留时间，真正让数字化服务“听得见”的同时也“看得见”。

部分国际领先金融机构信息技术投入金额与增速



图11 部分国际领先金融机构信息技术投入金额与增速

全国性科技总投入与占营收比重



数据来源：轻金融

图12 全国性科技总投入与占营收比重

（5）虚拟数字人+文旅

数字文旅产业表现突出，以虚拟讲解员为代表的虚拟数字人应用或为其贡献蓬勃力量。据中国旅游研究院统计，2019 年前三季度，文旅营业收入 62187 亿元，比上年同期增长 7.6%，特别是数字文旅产业表现突出，成为文旅产业转型升级的重要引擎。数字文化内容与互联网旅游、智慧旅游、虚拟旅游等新模式联动发展，特别是疫情期间得到不断增长。目前，虚拟数字人在文旅领域暂时还没有落地产品，但相关概念产品已经开始映入公众眼帘。例如商汤科技在 2020 世界人工智能大会（WAIC）上展示的 AI 数字人“小糖”。“小糖”可依托于展台前的滑动屏幕，为观众介绍和讲述预先设定好的讲解内容。虚拟数字人带来的全新展示方式和讲解模式，在帮助文旅场馆解决人力不足问题的同时，更凭借其智能化、电子化和展示效果多样化的特性，为文旅行业的未来带来更多可能性。



数据来源：中国旅游研究院

图 13 2012-2019 全国国内旅游人数

四、虚拟数字人发展趋势、制约因素及建议

（一）虚拟数字人未来发展趋势

虚拟数字人的采集、制作流程逐步简单化、一体化，迭代式提升真实感。未来数字人的制作技术将会变得更加简单，会有更加一体化、自动化的设备同步获取模型、身体、表情、手指运动、声音等所有数据，无需穿戴专业传感设备。在特定的环境下人们甚至无法分辨数字人的真假，可以更加自然与数字人交流互动。

虚拟数字人的交互方式衍变，智能化程度不断加深。全双工技术将推进数字人的交互方式衍变，让数字人拥有一次唤醒、多次交互的能力，具备实时智能响应、智能打断、智能纠错、多轮对话等功能。另外，随着计算机视觉、语音和自然语言处理等人工智能技术的不断进步，虚拟数字人将逐渐具备“看”、“听”、“说”、“懂”的能力。

虚拟数字人逐渐实现在多场景、多领域的融合、应用、落地。尽管目前数字人的发展环境、整体情况还处于尚未成熟的起步阶段，但随着虚拟数字人技术的精进、市场价值的释放，其将更全面、更深入地融入影视、金融、文旅等各个领域，充分发挥应用价值，迸发巨大的潜力。

（二）虚拟数字人发展制约因素

目前，制约虚拟数字人产业发展的因素主要有以下几个方面。

政策环境层面，行业依然处于培育期，缺乏政策与资本扶持。一是缺少上下游协作交流平台，制约整体产业化发展。虚拟数字人产业链长，参与主体多，既有大型平台企业、也有中小单位和初创公司，行业沟通成本高。二是行业投资回报周期较长，为数字人产业投资竖立了“无形壁垒”。数字人产业还不完全成熟，投资回报周期较传统技术应用更长，缺乏长期耐心资本来扶持企业成长。由于缺乏资本力量的推动，各产业规模扩张与资源整合的进程相对滞后。

技术应用层面，制作方式自动化程度低、生产门槛高，关键技术还不够完全成熟。目前 3D 数字人建模依然需要大量的人工制作参与，整体制作效率相对较低，与部分应用场景快速、批量制作的需求并不匹配。此外，在虚拟数字人的关键技术方面，还存在实时面部表情捕捉与还原精准度不足、语音识别在强噪声干扰及远场识别等方面还很难达到实用化要求等瓶颈，需要加大研究力度以提高算法精度。

人才培养层面，技术人才、综合人才极度缺乏，人才供应体系不完善。数字人的技术跨度大，从制作端的美术师、模型师、绑定师、材质师、动作捕捉师、灯光师、特效师、摄像师、导演、特效演员，到技术端的光学工程师、硬件工程师、软件工程师、算法工程师、系统工程师，以及运营端的产品经理、经纪人等每一个环节的人才都很重要，急需一

套完善的人才供应体系，特别是跨界人才体系，才能保障产业的良性运转。

标准体系层面，行业内缺乏统一技术标准和体系，产品质量良莠不齐。数字人的产业处于发展初期，制作型公司、技术型公司、运营型公司、应用型公司均已入场，但各企业的技术、产品质量差异较大。通过建立虚拟数字人的技术、产品标准体系，保护优质数字人厂商，促进行业健康发展。

安全伦理层面，相关法律法规和伦理规范尚待完善，存在潜在风险。虚拟数字人技术在发展中会引发一些法律问题和伦理风险，需要提前建立相关制度进行防范。虚拟世界里面的虚拟人物所有权、犯罪行为尚没有法律上的界定。虚拟数字人带来的沉浸式体验可能会对用户的精神、心理状态带来影响，改变用户在脱离虚拟世界后的行为，比如可能会增加暴力倾向，存在一定伦理风险。

（三）虚拟数字人产业发展建议

优化产业政策，推动产业健康发展。为进一步推动虚拟数字人产业发展，建议从以下几个方面优化产业政策。一是引导长线资本，加大对我国虚拟数字人发展的基础理论、实现方法、研发工具、技术系统等方面的资金支持力度，重点推进计算机图形学等基础理论研究，加快三维重建、动作捕捉、渲染合成、智能交互等核心关键技术研究，开展虚拟数字人制作流程智能升级研究，简化制作难度，减少人工作业，

提高生产效率。二是搭建数字人产学研一体化生态平台，建立健全高校、研究机构、行业协会、智库等的协同推进机制，增进企业技术协同、产品互联互通、上下游积极联动，加强在技术攻关、瓶颈突破、标准制定等方面的协调配合。三是加强行业对接，积极鼓励数字人技术与金融、文旅、教育等传统行业融合发展，鼓励政府部门率先运用虚拟数字人提升业务效率和管理服务水平，为行业发展起到率先垂范作用，营造良好的产业氛围。

加快人才培养，鼓励合作共赢。要加强人才队伍建设，加强后备人才培养力度，为产业发展提供更加充分的人才支撑。搭建基础研究和交叉学科研究的创新平台，培养学科交叉、知识融合、技术集成的复合型人才。鼓励科研机构、高等院校和相关企业联合开展虚拟数字人相关基础知识和应用培训。鼓励高校、科研院所与企业间开展合作，建设一批虚拟数字人领域实训基地。同时，鼓励国内外科研院所、企业、行业组织拓宽交流渠道，广泛开展合作，实现优势互补、合作共赢。

加强统筹协调，加快标准化建设。建立统一的虚拟数字人标准体系，围绕标准体系，立足国内、借鉴国际，加强虚拟数字人标准化工作的总体规划和顶层设计，统筹开展虚拟数字人关键技术、基础应用、服务平台、产品应用及服务和安全等领域的标准研制。以虚拟数字人产业发展需求为引

领，围绕产业发展中存在的共性问题，加快推进重点标准和基础标准的研制，动态更新虚拟数字人综合标准化体系，及时满足产业发展亟需。着重在虚拟数字人的三维建模、动作捕捉、智能交互、2D/3D 基础应用、安全可信等方面，开展相关标准的研制，进一步建立评级体系，以标准化手段提高技术创新的市场化效率。

提前规划布局，开展审慎包容监管。为更好地应对虚拟数字人潜在的风险问题，建议联合产业界各方着重以下两个方面的工作推进。一是共同开展数字人身份管理及内容安全的研究工作，从政策层面做好产业体系化布局，在坚决打击恶意违法行为的同时，给新兴技术一定包容发展的空间。二是加紧研究各类虚拟数字人应用对人的心理影响，以及对人类社会带来的影响，并尽快出台相关约束与法律法规。

中国人工智能产业发展联盟
Artificial Intelligence Industry Alliance

附件：虚拟数字人典型企业介绍

1. 浦发银行-金融数字人行业应用领航者

浦发银行“3D 金融数字人”创新计划自 2018 年启动，2019 年建设并投入金融应用，是业内首个综合运用 3D 人像实时驱动、多模态交互、自然语言处理、情绪识别、智能推荐、多因素生物认证等前沿技术，连接多个金融交易系统及物联网系统、在金融云上运行的规模化数字人金融应用。作为银行员工的数字分身，已正式在线上 APP 渠道以及线下营业网点为 C 端用户提供普惠服务。这也是由该行总行行长直接领导的创新应用项目之一，已成为金融数字化转型成效的一张名片。



附图 1 浦发银行数字员工 2020 年 8 月版

2. 凌云光-虚拟数字人软硬件系统及解决方案厂商

凌云光汇集人体重建、运动采集、特效合成等硬件条件与技术团队，构建 AI+数字媒体技术平台，研究数字人制作技术，以及 VR 交互技术。2019 年虎牙 LiveTech 大会现场采用凌云光数字人制作技术，发布虚拟主播晚玉。湖南省博物馆利用凌云光的人体动作采集技术，在 2019 年上线马王堆 VR 历史穿越体验项目，让观众能直观体验汉代辛追夫人一天的生活，深度了解汉代历史。



附图 2 凌云光高精度人脸 3D 建模系统 Facial 3D+

3. 腾讯-IP 虚拟人解决方案厂商

腾讯提供 IP 角色授权与多模态交互方案，致力于打造“能听会说”的 AI 智能形象。2020 年腾讯数字生态大会，采用了 IP 虚拟人技术，基于镜双城动画番剧女主角那笙的 IP 角色，腾讯构建了“那笙虚拟助手”，使其来到大会上与演讲嘉宾进行互动。2020 年世界 VR 大会，基于著名主持人钟石

的形象，腾讯构建了虚拟主持人“钟小石”，使其与钟石本人共同主持了世界 VR 大会的开幕式。2020 年深圳文博会北京专区，基于著名主持人春妮的形象及音色，腾讯构建了虚拟导览及虚拟客服“小春妮”，使其在线上展厅为用户提供绘声绘色的展厅播报讲解以及问题解答等能力。

4. 网易伏羲实验室-虚拟人科技创新中心

网易伏羲致力于全面推进虚拟人前沿技术研究、工程优化和解决方案，汇集美术制作、形象智能创建、智能表情合成、智能肢体动作合成、智能语音合成、智能对话生成、智能用户感知、运动采集等关键技术。已赋能于多行业多场景，包括游戏角色、虚拟教师、虚拟讲解员、虚拟诗人、虚拟品牌形象代言人、虚拟偶像；创新成果丰富，（截至 2020 年 11 月）共计发表国际顶级会议及期刊论文 58 篇，授权发明专利 34 项。

5. 搜狗-虚拟数字人计算机视觉、自然语言处理技术厂商

搜狗推出了多模态人工智能技术，基于文本、语音、图像、图形等多种模态信息建模生成高逼真度的虚拟数字人。2018 年 11 月 7 日第五届世界互联网大会上，搜狗与新华社联合发布 AI 主持人。通过语音合成、唇形合成、表情合成以及深度学习等技术，克隆出具备和真人主播一样播报能力的“AI 合成主播”。2020 年 5 月 21 日，搜狗联合新华社又推

出了 3D 版 AI 合成主播“新小微”，“新小微”AI 合成主播高度还原真人发肤，在立体感、灵活度、可塑性、交互能力等多个方面，都有了大幅跃升，而且还能根据播报的内容，做出各种更接近于真人的姿势和表情。



附图 3 搜狗多模态识别技术

6. 魔法科技-智能虚拟数字人全栈式解决方案提供商

魔法科技运用自主研发的全栈式 AI 表演动画解决方案、全智能虚拟数字人技术及文本实时合成语音动画（TTSA）技术，打造全智能虚拟数字人能力平台，赋能数字人在各类终端与用户进行三维、实时多模态地智能互动，可提供 7*24 小时的智能化服务。2019 年，光大银行率先落地应用虚拟银行雇员，即采用魔法科技的虚拟数字人技术。2020 年，小米开发者大会采用魔法科技虚拟数字人直播方案，发布全新小爱 5.0。目前，魔法科技 AI 智能交互虚拟数字人已广泛应用于教育、金融、医疗、市政等领域。