

## 本章导语

本章简要介绍数字人的发展历史，不同的分类方式、产业应用情况及基本的制作流程。通过这些内容，有助于学习者了解数字人相关的发展历程和理论知识，从而能够更好地理解计算机生成的数字人与现实世界的人类之间的相似之处和相关性。这些计算机生成的人物在一定程度上也反映了各个技术在不同时期的发展进程。随着技术的不断发展，数字人正不断突破自身的局限性，并在一些新兴技术的帮助下不断拓展其应用领域。

## 学习目标

- 了解数字人的发展历史及基本分类。
- 了解学习者在数字人产业应用中所扮演的角色，并思考学习者应具备什么样的知识技能。

## 1.1 数字人的发展历史

数字人是一种通过计算机生成的虚拟人物，是基于真实人物的模型，可以具备人类般的外貌、语言和行为特征。数字人技术的发展使得我们能够创造逼真、可交互的虚拟人物，使其在虚拟现实、游戏、直播等领域中扮演角色或提供服务。通过数字人，我们可以实现增强现实、沉浸式体验的目的，获得更多互动的机会。随着技术的更新迭代，数字人也经历了快速的发展过程。

### 1.1.1 数字人发展历程

数字人以计算机图形学算法为支撑，随着计算机图形学的发展而发展。早在 20 世纪 60 年代，数字人的雏形就开始出现，当时的数字人通常被描述为计算机生成的人类形象。1964 年，诞生于波音公司设计实验室的波音人（Boeing Man）（图 1-1），是第一个计算机生成的具有完整形象的数字人，由美国计算机图形艺术家、计算机图形学应用先驱威廉·费特（William Fetter）创造。作为第一个人体数字模型，这个由简单线条呈现的形象，用于研究飞行员和飞行舱的空间关系、评估人体工程学质量及探索视野问题，以提高飞机的安全性。在计算机还未普及的时代，大部分参与计算机图形实践的人都是工程师和

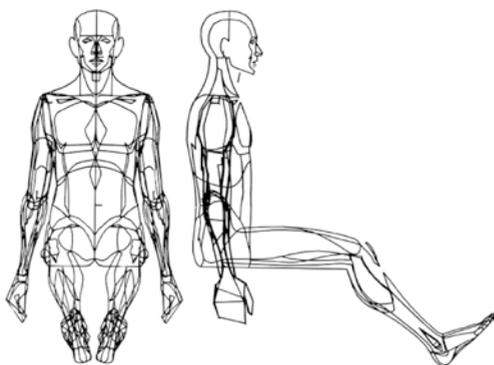


图 1-1 波音人

科学家，很少有艺术背景的人。对于艺术家来说，要想使用计算机进行创作需要学习编程技巧等，这种冷漠、机械的新创作方式和艺术形式使他们普遍难以接受。因此，20 世纪 60 年代的数字人仅仅停留在图像的形式上，对人类形象的模拟写实性还不够强。

在 20 世纪 70 年代，研究者们开始探索模拟人类行走、说话等行为方式的解决方案，并在计算机图形学领域取得了突破性的成就。与此同时，他们也开始追求计算机生成人物的“真实性”，这成了研究者们的新目标。他们开始

研究如何将三维物体的网格模型、明暗关系和色彩等信息输入计算机中。例如，1976 年，理查德·赫夫伦（Richard T. Heffron）执导的电影《未来世界》（*Future World*）中就尝试用计算机生成人物角色（图 1-2），旨在创造逼真、具有真实感的数字人物，使其在电影制作中扮演角色。

在三维建模软件出现之前，人物面部网格的生成首先需要将真人的面部涂成白色，并投影上规则的网格，然后拍摄照片，使用计算机计算照片的网格交叉点位置，从而获取面部数据。这个过程为计算机生成逼真三维人物模型奠定了基础，并为后续基于三维扫描技术构建人物模型提供了基本思路。

20 世纪 80 年代，计算机在性能上有了显著的提高。计算机生成的图像在电视、电影等大众传媒中广泛传播，数字人开始以三维的形式出现，使用计算机生成数字人的初步发展阶段已经到来。1981 年，理查德·泰勒（Richard Taylor）和加里·迪莫斯（Gary Demos）制作的计算机动画作品在 Information International Inc. (Triple I) 发布，其中诞生了第一个完全彩色的三维立体计算机生成人物亚当·派沃斯（Adam Powers，抛接杂技演员）（图 1-3），它也是最早的计算机生成图像（computer generated imagery, CGI）拟人角



图 1-2 《未来世界》电影海报



图 1-3 亚当·派沃斯

色之一。亚当·派沃斯是一个身着黑色燕尾晚礼服、头戴黑色高顶帽的绅士。他站立在黑白相间的棋盘格地面上，巧妙地玩弄着一个立方体、一个圆锥体和一个球体。这一时期的数字人在外观呈现上取得了明显的进步。

1982年，日本动画《超时空要塞》播出，其中的女主角林明美（图1-4）被包装成演唱动画插曲的歌手，并发行了专辑，林明美也因此成为世界上第一个虚拟歌手。由于技术的限制，这一时期的虚拟形象多由手绘或真人演员进行特效化妆实现。在三维数字人的研究上，进一步改善建模和着色技术是当时重点发展方向。



图 1-4 《超时空要塞》中的林明美

另外，为了追求更加逼真的效果，早在20世纪70年代，研究者们就开始探索对数字人面部表情和身体动作的模拟。除了参数化建模方法外，研究者们还运用雕塑和摄影等辅助手段来构建模型，积极探索创造更加流畅真实的动画效果。在身体动作的模拟方面，20世纪70年代末出现了动作捕捉技术，通过软硬件系统配合采集表演者的运动信息，并将这些信息与三维数字人体相关联，完成各式的运动模拟。到现在，动作捕捉技术已相当普及，不仅可以捕捉身体数据，还可以捕捉面部表情和复杂的肌肉形变。然而，研究者们发现，在数字人运动时，计算机生成的面部表情仍然显得僵硬，与真实人类表情相比存在明显差距。从20世纪80年代末期开始，研究者对人类皮肤及服装的模拟进行了研究，旨在解决柔体与刚体间相互作用引起的形变问题，以及其呈现的动态视觉效果。

进入20世纪90年代，伴随着计算机软硬件技术的迅猛发展，运算速度不断提升，价格日趋低廉，存储容量也日益扩大。计算机生成的图像分辨率大幅提高，细节呈现更加精致。于是，研究的重点逐渐转向渲染技术，计算机生成的图像实现了质的飞跃。在数字人的表现上，重点体现在对人物细节的模拟，诸如肌肤、服饰、头发等方面。由于这些模拟对计算机的运算能力要求较高，因此寻找快速有效的方法进行模拟成为当时研究的重点，例如，将皮肤的纹理细节和毛孔使用贴图技术进行处理，衣服尽量使用紧身贴合的衣服，减少服装飘动的动力学模拟，使用网格面片结合贴图进行毛发制作，使用较少的网格数量实现对头发的模拟，减轻计算机的运算压力。20世纪90年代中后期，随着计算机的普及，逐渐有许多家庭和个人拥有自己的计算机，开发者与艺术家对创作平台的需求日



图 1-5 初音未来



图 1-6 洛天依

益迫切，因此，各类 3D 制作软件如雨后春笋般出现，如 Maya、Cinema 4D 和 3ds Max 等。

20 世纪 60—90 年代，这些计算机生成的数字人通常作为虚拟演员及歌手出现，这些数字人与现实世界中的人类之间是一种被看与看的关系，没有任何肢体间的互动。20 世纪 90 年代末到 21 世纪初期，随着网络和三维技术的成熟，数字人朝着同时满足视觉逼真性、交互性和开放性方面发展，数字人步入快速发展阶段。2001 年，科学家雷·库兹韦尔（Ray Kurzweil）在 TED 大会上展示了计算机生成的具有学习能力的聊天机器人——雷蒙娜，这引发了研究者们对于智能数字人的关注。计算机生成的数字人开始尝试走出游戏、电视，融入更广泛的媒体中。不仅可以在杂志上见到它们的身影，还能在周边海报上欣赏到它们的形象。它们不再只是简单的虚拟形象，而是具备了详细的个人信息，以自身的魅力开始反向影响真实世界人类的审美和行为，无论是在娱乐界还是其他领域，数字人都有了更多的发展空间。2007 年，日本制作了虚拟偶像歌手初音未来（图 1-5），这一绑着碧绿色长马尾的二次元少女形象一经问世，就引来了人们的广泛关注和认可，成为世界上第一个使用全息投影技术举行演唱会的虚拟人类。初音未来的出现使电子音乐再度引起热潮，也刺激了电子音乐创作之外的一系列创作，如绘画、动画音乐短片（music video, MV）、角色扮演、主题餐厅等。2012 年，国内首位虚拟偶像歌手洛天依诞生（图 1-6），也同样吸引了大批忠实粉丝，发挥着自身的 IP 品牌价值。这一时期的数字人，

不再以逼真的模拟现实人为主要目标，而是通过融合计算机图形学（computer graphics, CG）技术、动作捕捉技术、语音合成技术等追求更加真实、立体的形象。

近年来，随着深度学习算法的突破，数字人创作开始步入快速发展轨道，人工智能为数字人发展提供重要支撑。1950 年，阿兰·图灵提出的著名问题：计算机能够思考吗？这个关于计算机能否像人类一样思考的假设，经过一代代科学家深入探索，到现在已经逐渐变为现实。数字人开始能够实现更加自然和智能的交互。例如，通过语音识别技术和自然语言处理技术，用户可以与数字人进行对话。以 2019 年浦发银行数字员工“小浦”为例（图 1-7），其外貌、动作和面部表情已经接近真人，与用户交流时语调和说话风格自然，能够实时感知和领会用户情绪，并及时作出相应反应。此外，“小浦”还具备主动学习能力，涵盖了 134 项常见零售业务和 1150 个常见问题，以及 72 项对公业务和 126 个常见问题，同时还提供了 809 个闲聊场景。此外，“小浦”还具备其他技能，如查询天气、汇率，进行抽奖，以及提供周边餐饮、交通和购物信息等。这进一步提升了网点人工智能 AI 服务的智能水平，有助于提高服务效率和质量。

当前，数字人的发展进入了快速增长的阶段。随着 2021 年“元宇宙”概念进入大众



图 1-7 浦发银行数字员工“小浦”

视野，人们对数字人的研究热情更加高涨。3D 数字人建模技术不断突破，人物形象呈现更加精细，制作周期更短，制作和使用成本也随之降低，数字人呈现出更高的智能化和多样化发展趋势。全面实现基于人工智能的驱动或许会成为数字人发展的最终形态。但到目前为止，AI 数字人在制作门槛和成本上仍然高于传统的数字人，全面普及还存在一定难度，在呈现效果上也有待提升。2022 年年底，ChatGPT（chat generative pre-trained transformer，人工智能研究实验室 OpenAI 研发的聊天机器人模型）的火热为交互数字人的发展方向增添了更多可能性，ChatGPT 作为目前世界上参数规模最大的语言模型，能够回答连续的问题，处理更加复杂和抽象的语言文本，会承认错误，所表现出的高度灵活性刷新了人们对 AI 的认知。将 ChatGPT 与数字人相结合，能够使数字人与真人的交互能力和表现力更加接近于理想状态。

### 1.1.2 数字人概念的变迁

回顾计算机生成数字人的发展历程可以发现，在不同的时期，对于这些数字人的命名也有所不同。比如早期活跃在影视和动画领域的计算机生成人类被称为“人造演员”“合成演员”或者“代理演员”等。在 3D 动画领域，为了与传统手绘二维角色区别开来，这些数字人被称为“3D 角色”“3D 人类”……这些数字人各种不同的命名，往往也代表着各个领域对数字人不同的研究重点和应用方向。

数字人的思想起源于赛博格（Cyborg），1985 年，哈拉维在其赛博格宣言中将赛博格定义为无机物机器与生物体的结合体，如安装了假牙、假肢、心脏起搏器等的身体，这些身体模糊了人类与动物、有机体与机器、物质与非物质的界限。

另外，现在人们经常讨论的“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”。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。这里对数字人的研究重点主要指医疗领域的人体结构可视化，这与当今的数字人研究重点不同。

在名称定义上，人们对计算机生成的人类类别持有不同的观点。许多研究者认为，从



图 1-8 数字人、虚拟人和虚拟数字人之间的关系

广义来看，“数字人”“虚拟人”“虚拟数字人”三者可以看作等同的关系，即运用数字技术创造出来的、与人类形象接近的数字化人物形象（digital human, meta human）。从狭义上看，这三者也存在一定的区别，“数字人”包含“虚拟人”和“虚拟数字人”。“数字人”指运用数字技术创造出来的、与人类形象接近的人物形象，与物理世界的人物形象一致地被称为“数字孪生”。“虚拟人”是“数字人”的组成部分，包含“虚拟数字人”，是指存在于虚拟世界中，虚构的人物身份。“虚拟数字人”是最小的概念范畴，指存在于虚拟世界，具有人类特征和人类能力的数字化形象。这里更加强

调通过 AI 技术“一站式”完成虚拟人的创建、驱动和内容生成，并具备感知、表达等无须人工干预的自动交互能力的数字人（图 1-8）。《2020 年虚拟数字人发展白皮书》中指出，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。

随着数字人生成技术的深入研究和应用场景的不断拓展，人们逐渐意识到它需要多个学科的支持和交叉融合。从 20 世纪五六十年代的计算机图形学开始，人们就研究如何用计算机生成和处理各种图像和动画。数字人的三维建模、渲染、动画等技术都是基于计算机图形学的研究成果。随着时间的推移，数字人的研究领域逐渐扩展到其他学科领域。例如，人工智能技术、机器学习、知识图谱、自然语言处理等技术可以使数字人更加智能、灵活，从而更好地与用户进行交互；社交心理学研究了人类社交行为的基本机制，数字人技术中的情感计算、情感表达等技术依赖于社交心理学的研究成果；而人机交互则是数字人技术的另一个重要基础学科，它研究人类与计算机之间的交互方式和技术；数字人中的语音识别、手势识别、视觉跟踪等技术为数字人的交互质量提供了强有力的支持。

## 1.2 数字人的分类

### 1.2.1 按表征人物数据资源和人物形象风格分类

计算机生成数字人形象，与计算机图形学领域的技术发展密切相关，呈现的形式经历了从简单的线条勾勒到现在与照片一样真实的形象，数据的表示方式也由二维空间的点、线、面，发展到由三维顶点坐标、三角面片顶点索引坐标、法向信息、纹理贴图坐标等构成的三维图形资源数据，制作方式也变得日益多样化，尤其是摄影摄像设备性能的提升和应用的普及，使纹理图片资源获取极为便捷，在数字人创作设计中，纹理图片得到了广泛的应用，极大提升了数字人形象的视觉效果。

根据表征人物数据资源的特征，可分为 2D 和 3D 两大类；在人物形象风格，又可以分为卡通、拟人、写实、超写实等风格（图 1-9）。

（1）卡通风格通常具有明亮的颜色、简化的线条和夸张的表情。角色通常具有大眼睛、夸张的表情和姿势，并且身体的比例可能会被扭曲或夸大。这种风格常常用于娱乐和幽默



图 1-9 数字人按形象风格分类

的目的，能够吸引观众的注意力并传达情感或故事。它可以让角色和场景更加生动有趣，同时也可以传达一些深层次的主题和情感。

(2) 拟人风格的数字人是指以人类为基础，但在外貌和特征上进行一定程度的抽象和变形的数字人。这种风格的数字人通常也有特征夸张，但夸张程度弱于卡通风格，使其看起来像人但又比真人更可爱、有趣、独特。

(3) 写实风格的数字人是指以逼真的方式呈现人物形象，超写实风格的数字人通常具有高度逼真的外貌和细节，使观众很难分辨其是否为真人，需要强大的技术作为支撑。

写实和超写实风格的数字人在许多领域都有广泛的应用。例如，在电影和游戏制作中，可以使用写实风格的数字人来创建逼真的角色，使故事更加引人入胜。此外，写实风格的数字人也可以用于虚拟现实和增强现实应用中，使用户可以与逼真的虚拟人物进行互动。

### 1.2.2 按应用分类

数字人有广泛的应用领域，如娱乐和游戏、教育和培训、营销和品牌形象、医疗和健康等，按照数字人在应用领域中特征可分为以下三类（图 1-10）。

(1) 偶像型应用：主要应用于娱乐影视行业，如电影、电视剧、动画短片、游戏和虚拟演唱会等，扮演特定角色。在故事情节中，它们能够与其他角色进行互动，甚至与真实演员实现无缝融合。角色数字人的外貌逼真、个性鲜明，能够通过精细的脸部动画和肢体语言表达人物性格和情感。通过衍生周边和商业宣传活动，打造独特的 IP 形象或社会身份，以获得大众的认同，以致发展为大众的偶像。制作这种类型的数字人需要高昂的成本。

(2) 服务型应用：广泛应用于智能手机、智能音箱、智能家居等设备中，为用户提供智能化的交互式服务。它们以人类形象的虚拟角色形式出现，通过语音、文字或图像界面与用户进行交流和互动。它们不仅提供实用的服务，还提升用户体验和便利性。这种类型的数字人注重服务的质量和智能性，而并非过于关注外观造型。随着人工智能和自然语言处理技术的不断进步，虚拟助手数字人的功能和表现形式也在不断发展和改进。

(3) 身份型应用：通常应用于元宇宙或虚拟现实世界之中，是真实世界中的人在虚拟世界中的化身，类似于线上会议或聊天室中的头像，除了外观形象外，数字人与真人具有同一性，代表相同的身份，它们是一种数字孪生式存在。这种类型的数字人为用户提供了完全改变自身形象的机会，用户能够根据自己的偏好来设计和装扮数字人的外貌，以全新的形象出现在虚拟世界中，展示个性和风格的独特之处。此外，身份型数字人还能实现在现实世界中难以实现的愿望和想法，为用户带来崭新的体验。



图 1-10 数字人应用分类

### 1.2.3 按交互模块分类

根据数字人有无交互模块，可以分为非交互型数字人和交互型数字人。

(1) 非交互型数字人包括静态和动态形式，静态是不带动画的角色形象。动态形式的内容是预先制作完成的，并在呈现时以固定的方式播放，类似于视频。其中一种动态形式是基于图像空间生成的视频内容，如基于视频或图像序列创建的虚拟讲解员，这种方式不需要三维建模。另一种方式是通过三维模型生成动画视频内容，这需要进行角色设计、三维模型创建、材质和纹理制作、骨骼绑定、权重设置、动画调整、渲染和光照等环节，最后进行音视频合成完成内容设计制作。非交互型数字人是通过让观众被动接收视听觉信息的方式，达到信息传播目的。

(2) 交互型数字人具备感知用户行为及产生类人动作的能力，可实现与真实世界的交互。根据驱动方式的不同，交互型数字人可分为真人驱动型数字人和智能驱动型数字人。真人驱动型数字人是通过真人来驱动数字人，其主要原理是在完成建模和关节绑定之后，利用动作捕捉采集系统或摄像头将真实人类的表情、动作等数据实时映射到数字人身上，从而产生相应的动作和表情，实现与用户的交互。此外，也可以通过语音来驱动数字人，实现实时语音对话互动。与真人驱动不同，智能驱动型数字人是利用深度学习技术，学习现实中真人的动作、表情、口型、语气等信息，通过系统自动读取并解析这些信息，根据解析结果决策数字人后续的行为，驱动人物模型生成相应的语音和动作，从而实现数字人与用户的互动。在最终的呈现过程中，需要考虑动作解算和模型渲染的实时性，以确保交互的低延迟，以满足用户对流畅体验的要求。

## 1.3 数字人的产业应用

在国家大力发展数字经济的背景下，各行各业都积极推进产业的数字化进程。数字人作为一项重要的数字化工具，展现出广泛的应用前景，并在多个行业中得到应用，形成有竞争力的数字人产业集群，实现技术、内容和应用的协同发展，创造全新的商业模式。未来，数字人将全面渗透到传统产业，提升各行各业的效率。数字人的产业链可划分为基础层、平台层和应用层（图 1-11）。



图 1-11 数字人产业链格局（摘自《2020 年虚拟数字人发展白皮书》）

### 1. 基础层

数字人产业链中的基础层，主要为技术提供方，是数字人技术发展的基石。基础层为数字人发展提供软硬件支持，硬件包括数据采集、模型构建、感知输入、显示输出等方面的设备、传感器、芯片等。软件包括基础建模软件、驱动插件、渲染与开发引擎等。在现阶段，数字人最终会呈现在显示设备上，因此显示设备是数字人的基本载体，包括电视、手机、投影、LED 屏、VR、裸眼 3D 屏幕等。传感器用于采集用户原始数据，如面部、肢体、手部的运动数据，常见的传感器包括 LeapMotion、Kinect 体感器、数据手套。建模软件进行数字人的身体、毛发、服装等基本的外观制作。主流渲染与开发引擎包括 Unity、Unreal Engine 等。由此可见，技术的迭代是数字人产业革命的根本原因，上游基础层为全产业链注入核心驱动力。

### 2. 平台层

数字人产业链中的平台层，主要为运营商，包括企业与个人，是数字人技术场景的开拓者。一方面，平台层需要与上游完成技术的融合。其中，平台层中的软硬件系统企业

通过基础层获取软硬件技术，获取各类数据信息后通过算法实现人物的外观建模，动作模拟。生产技术服务平台提供一站式数字人解决方案。AI 能力平台提供交互技术支撑，包括视觉感知与听觉感知等多维度的技术点，如视觉感知中涉及物体识别、表情识别、姿势识别等，听觉感知涉及自动语音识别（automatic speech recognition, ASR）、自然语言处理（natural language processing, NLP）、语音合成（text to speech, TTS）、语音驱动面部动画（audio-driven facial animation, ADFA）等技术。另一方面，平台层需要与下游与内容运营者共同探索更多的应用机会和商业运营。

### 3. 应用层

数字人产业链中的应用层，主要为内容运营商，涉及数字人后续各领域商业变现及落地场景，可根据需求为消费者提供定制化服务。例如，虚拟主播、数字角色、数字员工、虚拟导游、虚拟讲解员、虚拟教师等商业变现路径。以内容运营为主的企业还要负责内容的产出以及 IP 的运营制作，最大限度地满足消费者的功能性和情感性需求并扩大数字人的影响力。



### 课后练习

1. 请结合自己的理解，在网络上找到更多的数字人应用案例。
2. 请将自己找到的案例进行分类解读。