

数智赋能·文娱新生

中国大文娱产业人工智能应用发展研究报告

人民数据管理（北京）有限公司

中国信息通信研究院云计算与数字化研究所

广州趣丸网络科技有限公司

2026 年 7 月

版权声明

本研究报告版权属于人民数据管理（北京）有限公司、中国信息通信研究院云计算与数字化研究所、广州趣丸网络科技有限公司，并受法律保护。转载、摘编或利用其他方式使用本报告内容或观点，请注明：“来源：《数智赋能·文娱新生：中国大文娱产业人工智能应用发展研究报告》”。违反上述声明者，编者将追究其相关法律责任。

前 言

当前，数字中国战略、文化强国战略和“十五五”规划等一系列政策正加速落地实施，新一轮科技革命与产业变革也在深入推进，以生成式大模型为代表的人工智能技术正深刻影响文化领域，成为推动文化创新和产业转型升级的重要驱动力量。这为包括大文娱行业在内的新兴文化创意产业提供了宝贵的发展机遇。在此背景下，人工智能技术与大文娱产业的深度融合，正在积极释放多重价值，促使“AI+大文娱”产业链不断创新发展。

主管部门目前尚未对“大文娱产业”概念及边界进行权威界定。综合产业趋势、技术演变与主流共识，本报告将“AI+大文娱”产业的范畴界定为：以满足大众精神文化需求为根本目的，以内容创作与体验消费为核心驱动力，以人工智能为核心的数智技术为关键创新载体，横跨多种内容介质与消费场景，全面涵盖写作、语音、音乐、视频和数字人等领域的内容孵化、生产制作、分发传播、互动消费及IP衍生运营全链条的产业形态集合体。

从应用实践来看，人工智能技术正加速向写作、语音、音乐、视频和数字人等细分业态纵深渗透，有效降低了内容生产的边际成本，显著提升了创意实现的效率，推动大文娱产业向数字化、智能化方向加速转型。与此同时，AI多模态技术在文化出海领域的应用持续深化，为传播中国文化提供了重要的技术支撑。

为全面呈现上述发展变化，系统梳理人工智能技术在大文娱产业的应用现状与发展趋势，人民数据、中国信息通信研究院云计算与数字化研究所（以下简称“云大所”）与趣丸科技组织开展了专项研究，编制形成本报告。

本报告的调研与撰写工作，得到了相关行业专家、智库学者及前沿科技企业的大力支持与积极配合。各参与企业在调研阶段提供的丰富产业实践案例、技术洞察与一线数据，为本报告的落地分析提供了重要的事实依据，在此一并致以诚挚谢意。主要参与撰写人员：杨森滢、栗蔚、宋克、刘畅、陈丽、窦佳丽、龙玲、任少峰、刘聪、张媛媛、李杉、王姚、李家希、顾雨薇、荣思嘉、陈明勋、景晨龙、李朔。

受人工智能技术快速迭代及大文娱产业形态持续演变等因素影响，本报告的部分观察与结论难免存在局限，敬请各界同仁批评指正。我们将持续跟踪研究，不断更新完善，以期为行业发展提供更为准确、深入的参考支撑。

目 录

一、 政策背景：“AI+文化”的发展机遇	1
（一） “十五五”时期：新大众文艺和文化“新三样”带来发展红利	1
（二） 文化数字化和“人工智能+”：大文娱产业的数智化转型路线图 ..	4
二、 产业全景：“AI+大文娱”产业发展现状	8
（一） 需求驱动下的产业发展特征	8
（二） 中国大文娱产业 AI 应用渗透全景	9
（三） 产业生态图谱与竞争格局分析	11
三、 技术全景：驱动大文娱产业变革的核心 AI 技术	15
（一） 多模态大模型：打破内容载体壁垒的核心基座	15
（二） 智能语音与 AI 音频技术：促进听觉交互与内容生产的底层引擎 ..	17
（三） 数字人与空间计算技术：拓展虚实融合内容边界	20
（四） 智能体应用：打造沉浸式文娱互动生态	21
四、 应用场景：AI 赋能大文娱产业的典型场景与模式创新	23
（一） “AI+语音”：打造沉浸式听觉新体验	23
（二） “AI+视频”：影视工业与微短剧的“视觉升级”	29
（三） “AI+音乐”：以技术演进推动音乐创作的大众化探索	33
（四） “AI+写作”：文本创作从人工主导向人机协同演进	39
（五） “AI+数字人”：加快多模态交互应用拓展	41
五、 全球化传播：AI 驱动中国文化出海的新路径	46
（一） 发展机遇：政策环境与技术支撑持续推动文化出海	46
（二） 关键路径：多模态技术提升内容本地化生产效率	47
（三） 典型实践案例	48

六、 挑战与治理：“AI+大文娱”的风险识别与规范发展	50
（一） 风险透视：“AI+大文娱”发展面临的四重挑战	50
（二） 多方共治：产业规范发展建议	52
七、 趋势观察：“AI+大文娱”未来发展新风向	53
（一） 生产范式演进：高水平人机协同赋能高效率内容生产	53
（二） 产业空间拓展：“AI+大文娱”出海促进数字 IP 跨文化传播	54
（三） 媒介形态演进：“AIGC+空间计算”推动沉浸式文娱场景拓展 ..	54
（四） 价值闭环探索：向“个性体验服务”与“数据资产运营”延伸 ..	55
（五） 人才迭代升级：新职业逐步形成	55

图目录

图 1	国内“AI+大文娱”产业链上下游协同全景图	13
图 2	国内“AI+大文娱”技术演进路径图	23

表目录

表 1	文化数字化和“AI+文化”相关政策梳理	5
-----	---------------------------	---

一、政策背景：“AI+文化”的发展机遇

在建设文化强国的进程中，文化产业高质量发展始终引发多方关切。人工智能作为新一轮科技革命和产业变革的重要驱动力量，正在深刻影响着文化产业发展。

（一）“十五五”时期：新大众文艺和文化“新三样”带来发展红利

1. “新大众文艺”正积极推动文化普惠

2026年3月正式发布的《中华人民共和国国民经济和社会发展第十五个五年规划纲要》（以下简称“十五五”规划），明确提出了“广泛开展群众性文化活动，繁荣互联网条件下新大众文艺，扎实推进文化强国建设”的重大战略部署。

新大众文艺是指在党的领导下，依托中华优秀传统文化的深厚滋养，契合数字化、网络化、智能化深度融合趋势，人民群众以前所未有的规模和深度参与文艺生产传播，运用新技术、新工具突破传统文艺生产传播流程，呈现共创共享的一种新型文艺形态。理解其“新”之所在，需聚焦以下三个转变：

其一，创作主体的结构性扩容。传统文艺的创作生产长期依赖少数专业机构与职业创作者，供给高度集中。而在“新大众文艺”的发展格局下，随着人民文化素养的提升与数字工具的普及，文艺创作实现了从“为大众”“写大众”向“大众写”“大众享”的转变。各行各业的普通群众正深度参与内容共创，文化创造的

主体边界得到拓展。

其二，**内容形态的多元化与交互化**。区别于传统文艺以“静态、单向”阅听为主的消费模式，“新大众文艺”形态天然具备动态性、沉浸性与强交互性。网络文学、短视频、微短剧等新业态依托新媒体平台蓬勃生长，弹幕互动、UGC 二次创作等形式深刻重构了文艺内容的表现边界与消费逻辑。

其三，**传播媒介的平台化与泛在化**。内容分发从依赖传统媒介渠道转向以算法推荐与社交裂变为核心的平台生态。内容的触达范围与传播速度实现了数量级的跃升，碎片化、即时性、跨设备的消费模式成为主流。传播方式正从“中心广播”向“网络化、多向度的分布式传播”转变。

“新大众文艺”带来的变化与影响，正在积极推动文化普惠，丰富百姓的精神文化生活。“十五五”规划强调“丰富人民精神文化生活”，并将“社会文明程度明显提升”作为“十五五”时期经济社会发展主要目标之一。乘“新大众文艺”之势，借大文娱产业数智化转型之机，优质数字内容的供给规模正在扩容，多元化的文化资源正加速普及下沉，全面融入大众的日常消费场景。这种规模庞大、触手可及的全民共创与普惠共享，为构建覆盖更广泛人群的精神富裕图景提供了坚实支撑。

2. 新技术赋能文化“新三样”讲好中国故事

随着数字技术的飞速发展与全球文化交流的日益深入，近年

来中国文化“走出去”的步伐不断加快，形态愈发丰富。网络文学、网络影视剧、网络游戏作为中国数字文化产业的重要成果，被业界称为文化“新三样”，已逐步成为对外文化交流的重要载体。较之于此前，文化“新三样”根植于互联网原生土壤，天然具备数字化、可复制、易传播的内容属性，能够以相对较低的边际成本实现跨地域、跨语言传播。

三大赛道之间已形成较为成熟的IP联动机制：网络文学为影视和游戏提供故事原型等内容设定，影视与游戏的热度反哺原著关注度，形成多向度的内容传播协同效应。据公开数据，2025年我国自主研发游戏海外年销售收入突破200亿美元，网络文学海外活跃用户突破2亿，覆盖全球200多个国家和地区，国产网络影视剧、微短剧在东南亚、中东、欧美等市场热度持续提升。国内已初步构建起从原创创作、IP改编、多语种本土化适配到全球发行与衍生开发的产业链，逐渐具备了持续输出优质出海内容的产业基础。

当前，文化“新三样”的海外传播正产生多层面的积极效应：其一，有助于打破刻板印象，构建真实立体的中国形象。“新三样”将中华优秀传统文化元素融入通俗叙事与沉浸式场景，以轻量化、具象化的表达方式降低文化解码门槛；其二，有助于带来双向互动的文化对话格局，增强海外传播亲和力。海外受众可通过评论、弹幕、游戏互动等方式深度参与文化“新三样”传播，

传播主体的多元化使文化叙事更具亲和力与渗透力，有效减少了文化传播中的心理隔阂；其三，有助于延伸文化影响半径，带动多元领域协同发展。“新三样”出海的溢出效应正在向文旅、电竞、衍生品等领域延展，提升“入境游”对海外受众的吸引力。

（二）文化数字化和“人工智能+”：大文娱产业的数智化转型路线图

1. 文化数字化和“人工智能+”提升文化产业新质生产力

以文化数字化和人工智能驱动文化产业提升新质生产力，将数智变革的动能系统性注入文化产业全链条，既是大势所趋，又是行业机遇。从政策发展脉络来看，我国文化数字化和“AI+文化”领域有关的政策部署经历了如下演变：

2017年，《文化部“十三五”时期文化产业发展规划》与《文化部“十三五”时期文化科技创新规划》相继落地，加速文化与科技融合发展的步伐。2019年，《关于促进文化和科技深度融合的指导意见》明确提出“打通文化和科技融合的最后最后一公里”，政策导向开始从“理念倡导”转向“路径攻坚”。2022年，《关于推进实施国家文化数字化战略的意见》明确提出“建成物理分布、逻辑关联、快速链接、高效搜索、全面共享、重点集成的国家文化大数据体系”。2023年，《数字中国建设整体布局规划》进一步将数字文化纳入数字中国建设的整体框架，提出“打造自信繁荣的数字文化”。有关政策的相继发布与实施，

为后续数字文化领域的深度探索和应用提供了重要参考依据。

2025年8月，国务院《关于深入实施“人工智能+”行动的意见》以专门文件部署“人工智能+”行动，推动人工智能在繁荣文化生产、增强文化传播、促进文化交流中展现更大作为，利用人工智能辅助创作更多具有中华文化元素和标识的文化内容，壮大文化产业。进入2026年，政策部署进一步向纵深化、下沉化发展。

“十五五”规划首次将“新大众文艺”写入五年规划，要求繁荣文化创作生产；《智能体规范应用与创新实施意见》鼓励“研发文学、音乐、绘画、视听、演艺等内容创作智能体，促进优秀文化传播推广”；《2026年数字社会发展工作要点》将“深化人工智能在‘影视+文旅’模式中的创新应用”纳入数字社会建设重点任务。

表1 文化数字化和“AI+文化”相关政策梳理

发布日期	政策名称	发布主体	核心要求
2017年4月	《文化部“十三五”时期文化产业发展规划》	文化部	推进“文化+”和“互联网+”战略
2017年4月	《文化部“十三五”时期文化科技创新规划》	文化部	全面推进科技融入文化领域
2019年8月	《关于促进文化和科技深度融合的指导意见》	科技部等六部委	打通文化和科技融合“最后一公里”
2021年5月	《“十四五”文化产业发展规划》	文化和旅游部	“上云用数赋智”，推动文化产业全面转型升级
2022年5月	《关于推进实施国家文化数字化战略的意见》	中共中央办公厅、国务院办公厅	到2035年，建成物理分布、逻辑关联、快速链接、高效搜索、全面共享、重点集成的国家文化大数据体系，中华文化全景呈现，中华文化数字化成果全民共享。
2022年8月	《“十四五”文化发展规划》	中共中央办公厅、国务院办公厅	打造公共文化数字资源库群，建设国家文化大数据体系
2023年2月	《数字中国建设整体布局规划》	中共中央办公厅、国务院办公厅	推进文化数字化发展，深入实施国家文化数字化战略，建设国家文化大数据体系，形成中华文化数据库。提升数字文化服务能力，打造若干综合性数字文化展示平

发布日期	政策名称	发布主体	核心要求
			台，加快发展新型文化企业、文化业态、文化消费模式。
2023 年 7 月	《生成式人工智能服务管理暂行办法》	国家互联网信息办公室等七部委	对生成式人工智能服务实行包容审慎和分类分级监管
2023 年 9 月	《元宇宙产业创新发展三年行动计划（2023—2025 年）》	工业和信息化部办公厅等五部委	建设文旅元宇宙；加快元宇宙在广播电视听场景的应用
2024 年 11 月	《数字虚拟人技术要求》	国家广播电视总局	规定了广播电视和网络视听行业数字虚拟人的技术要求
2025 年 1 月	《关于推动文化高质量发展的若干经济政策》	国务院办公厅	探索文化和科技融合的有效机制……支持文化领域大模型建设
2025 年 3 月	《人工智能生成合成内容标识办法》	国家互联网信息办公室等四部委	促进人工智能健康发展，规范人工智能生成合成内容标识
2025 年 4 月	《网络出版科技创新引领计划》	国家新闻出版署等十部门	支持网络出版企业围绕人工智能、大数据、云计算、区块链等网络出版相关技术建立企业实验室、企业技术中心等研发机构
2025 年 8 月	《关于深入实施“人工智能+”行动的意见》	国务院	推动人工智能在繁荣文化生产、增强文化传播、促进文化交流中展现更大作为，利用人工智能辅助创作更多具有中华文化元素和标识的文化内容，壮大文化产业。
2026 年 3 月	《中华人民共和国国民经济和社会发展第十五个五年规划纲要》	国务院	全面实施“人工智能+”行动，加强人工智能同文化建设相结合
2026 年 3 月	《文化和旅游部办公厅关于申报“人工智能+文化和旅游”应用试点的通知》	文化和旅游部办公厅	六大重点方向：辅助文化艺术创作生产、提升公共文化服务水平、助力文化遗产保护传承、培育文化和旅游新业态、提升旅游服务和治理效能、加强文化和旅游市场监管
2026 年 5 月	《智能体规范应用与创新发展实施意见》	国家互联网信息办公室等三部委	研发文学、音乐、绘画、视听、演艺等内容创作智能体，促进优秀文化传播推广
2026 年 5 月	《2026 年数字社会发展工作要点》	国家数据局	深化人工智能在“影视+文旅”模式中的创新应用

数据截至：2026-6-20|中国信通院基于公开政策文本整理

2. AI 驱动全民创作热潮

“十五五”规划提出“坚持文化惠民，实施公共文化服务提质增效行动，促进优质文化资源直达基层。广泛开展群众性文化活动，加强群众文艺团队建设”。这一愿景的实现，有赖于一个

前提：普通用户能够真正跨越内容创作的专业门槛。在这一语境下，AIGC（人工智能生成内容）技术不仅是生产效率的提升工具，更是激发全民创作热潮、改变内容供给生态的底层技术引擎。

技术普惠有助于打破高度专业化的技能壁垒。传统文娱内容的生产创作属于典型的高门槛劳动。例如，网络文学需要深厚的文学素养与持续的更新产能；音乐创作横跨乐理、编曲、演唱、混音等多个专业维度；视频制作涉及拍摄、剪辑、特效、配音等复杂环节；数字人开发更需要昂贵的动态捕捉、3D建模。这导致优质内容的供给长期依赖于少数专业团队。而人工智能技术的突破，正在打破这些壁垒：一个不懂乐理的普通用户可以通过自然语言对话生成一首词曲编唱俱全的歌曲；一个从未接触过视频剪辑的创作者，凭脚本便能生成影视级画面；任何人都能以照片为基础创建具有实时交互能力的专属数字形象。这种“创意即创作”的技术赋能，极大地拓宽了文化表达的主体边界，使普罗大众得以广泛参与到高质量的内容生产中。

技术普惠点燃了全民共创的文化热潮。技术普惠以生成内容标识、版权合规、内容安全和未成年人保护等为前提。技术的普惠下沉，带来的不仅是微观层面生产效率提升，更是全社会文化创新活力的迸发。当数以亿计的普通用户不再受限于专业技能的羁绊，能够借助可用、好用的AI创作工具将天马行空的创意随时转化为多姿多彩的数字内容时，大众参与文化建设的热情将被

彻底激活。这股生生不息的全民共创热潮，不仅丰富了互联网的内容多样性，更在推动“新大众文艺”落地生根，具有坚实的技术动力与群众基础。

二、产业全景：“AI+大文娱”产业发展现状

（一）需求驱动下的产业发展特征

从大文娱行业真实的需求出发，当前 AI 应用正经历深刻的产业范式演进。其并非单纯的技术堆砌，而是由行业对高品质内容、降本增效与高效价值化的渴求所直接驱动，具体呈现出三大核心特征：

其一是核心能力演进，即高品质内容需求驱动大模型从“通用泛化”走向“专业化与垂直领域深耕”。随着文娱内容的消费审美不断升级，产业对 AI 生成内容的标准已从“可用”向“好用”乃至“专业级”转变。传统的通用大模型在面对复杂的文娱创作时，往往存在“通用有余、垂直不足”的固有局限。由此，底层模型能力正加速向垂域深耕演进，也为文娱 AI 应用构建起较高的专业技术壁垒。

其二是生产方式演进，即降本增效需求驱动“单点工具”向“多模态工作流”无缝融合。大文娱产业是高度协同的复杂创意产业。早期离散的、单一模态的 AI 工具极大地增加了创作者的跨平台跳转成本与操作门槛。为了解决这一痛点，当前的 AI 应用正向“一站式、自动化工作流”全面演进。

其三是**商业模式演进**，即需求驱动底层技术与商业生态绑定。AI 赋能带来的效能提升催生出新的商业模式。例如，AI 技术使得每月数万部微短剧的工业化量产与多语种本地化译制成为现实，打破了过往受制于人工产能的瓶颈，提升中国文化及文娱产业的海外影响力。

（二）中国大文娱产业 AI 应用渗透全景

当前，人工智能技术对大文娱产业的影响已进入深水区，呈现出如下四重态势，勾勒出中国大文娱产业数智化转型的全景。

从技术覆盖广度看，**多模态大模型打破介质壁垒**。AI 能力对大文娱产业的技术覆盖，已完成从早期“单一文本生成”到“全域多模态”的跨越。如今，AI 能力横向拓展至文字、图像、音频、音乐、视频及数字人等全域媒介，形成广阔的技术覆盖网。多模态技术的快速发展，意味着不同内容形态之间的介质壁垒正在被打破：一段文字描述可以直接生成视频、配乐与数字主播，一个创意可以无缝流转为多种内容形态，内容的“跨模态生产”正在成为广告、短视频、微短剧等高频内容生产的重要趋势。

从产业链渗透深度看，**从单点辅助走向全链路赋能**。AI 在大文娱产业的介入，已从早期的“后期辅助工具”角色，向上游全面渗透，逐步贯穿内容生产的“IP 孵化—策划—创作—生产—宣发”全生命周期。这实现了从单点效率提升到全链路赋能的转变。

在 IP 孵化与策划阶段，大模型的知识推理与数据分析能力，使得内容趋势预判、受众偏好洞察与题材可行性评估得以自动化实现，显著提升爆款 IP 的商业孵化效率；在创作与生产阶段，AI 正进化为深度嵌入的“创作合伙人”，承担从文字生成、配乐创作、视频制作到数字人驱动的全链条内容生产任务；在宣发阶段，AI 驱动的个性化推荐、智能剪辑与多平台分发策略优化，正在重构内容触达用户的效率边界。

从内容供给扩容看，生产力下沉激活创作者经济。人工智能降低了内容制作门槛，正在激发创作者经济的深层活力，推动行业向“专业机构（PGC）+AI 赋能的超级个体”并存的规模化供给生态演进。借助 AI 的生产力杠杆，市场正催生出大量新型内容品类，改变了传统的劳动密集型分工。

从消费体验升维看，交互模式正发生迭代升级。AI 技术不仅改变了生产侧，也影响了用户的商业消费体验逻辑。具备长期记忆与情感感知能力的 AI 智能体，开创了超越传统的陪伴式消费场景；用户也不再是被动的内容接受者，而是可以影响内容创作的主动参与者。内容消费与创作边界的持续模糊化，正加大“体验经济”在大文娱产业发展中的分量，愿意为 AI 生成的专属体验支付个性化费用的用户正越来越多。

（三）产业生态图谱与竞争格局分析

1. 产业链上下游协同全景

中国“AI+大文娱”产业的生态架构，遵循清晰的“三层协同”结构逻辑：以算力基础设施为核心的基础算力层、以通用大模型与垂直应用模型并行为特征的模型能力层，以及以文娱垂直场景与广大创作者为核心的场景应用层，三层之间形成“算力支撑模型、模型赋能场景、场景反哺迭代”的闭环价值流动，共同构成中国“AI+大文娱”产业的生态图谱。

第一层是基础算力层，用于支撑“规模化训练”与“高实时推理”。基础算力层是整个“AI+大文娱”产业生态的物理基础，承担了为上层模型训练、微调与高并发推理部署的核心职能。

从算力供给结构看，当前国内“AI+大文娱”产业的算力底座正形成三位一体的“端云协同”框架：一是以国家新型数据中心与各地智算中心为代表的公共算力基础设施，为大规模模型训练提供集中式算力保障；二是云计算平台，以弹性算力租用模式降低中小文娱企业与独立创作者的算力获取门槛，推动 AI 能力的普惠化部署；三是以端侧芯片与轻量化推理设备为代表的边缘与终端算力，支持 AI 文娱应用在移动端本地的高效运行。

值得注意的是，大文娱领域的算力需求结构与工业或科研计算不同——视频生成、实时语音驱动、高保真数字人渲染等高频消费场景，正推动产业算力重心从“重训练”向“重推理”倾斜。

这类场景对推理算力的低延迟、高并发与调度弹性提出了高要求。未来，算力成本的下降与国产异构芯片的快速迭代，会在一定程度上影响“AI+大文娱”应用的普及程度。

第二层是模型能力层，构建“通用大模型+垂域模型”的立体智能中枢。模型能力层是承上启下的核心枢纽，也是当前“AI+大文娱”产业价值创造较为密集的高地。目前模型能力层已告别单点模型时期，呈现出鲜明的“通用大模型做宽底座+垂直大模型做深专业”的双轨并行格局。

在通用大模型侧，国内主流大模型厂商持续提升语言理解、多模态生成、逻辑推理等基础能力，并以开放 API 或插件生态向下游输出基础智能。其核心价值在于“广谱适配性”，能低成本完成创意发散、文本润色、资料检索等通用性辅助支持。但在面对复杂文娱专业需求时（如复杂情绪复刻、特定风格音乐生成等）往往面临“通用有余、垂直不足”的能力边界瓶颈。

在垂直应用模型侧，专注于特定内容介质的垂直大模型厂商正在加速崛起。其核心竞争逻辑在于以通用大模型为底座，注入大文娱细分场景的稀缺私有数据（如高质量版权音乐、多语种专业配音语料），通过精细化的工程打磨与 RLHF（基于人类反馈的强化学习），形成了通用大模型难以复制的专业深度优势。目前，在 AI 写作、AI 音乐、AI 语音、AI 视频生成及数字人等垂

类赛道，均已涌现出具备行业影响力的垂直大模型，构筑了通用厂商难以轻易跨越的专业壁垒。

第三层是场景应用层，通过多模态 workflows 影响五个领域的价值生态。场景应用层是“AI+大文娱”产业价值实现的最终落点，将底层技术转化为直接面向创作者与用户的产品与服务。围绕语音、视频、音乐、写作和数字人五大领域，目前已经形成了多样化的应用形态。

同时，场景应用层构成了整个生态的动力引擎。数以亿计的创作者与用户在真实场景中产生的高质量交互数据与反馈信号，形成了强大的“数据飞轮”（即“场景牵引数据、数据驱动模型、模型赋能应用、应用创造价值”）。这些数据持续向上输入，驱动模型能力迭代，进而拉动底座规模化的算力消耗。



图 1 国内“AI+大文娱”产业链上下游协同全景图

2. 竞争格局与市场主体分析

当前，大文娱 AI 应用市场的竞争逻辑已从单纯的“模型参数比拼”转向“场景聚焦度”与“价值化深度”（指产品是否形成闭环营收、用户付费意愿及复购率等指标）的综合较量。基于这两个核心维度，当前产业主体演化出了四个竞争阵地：

第一类是**生态引领者**，以字节跳动（如剪映、Seedance）、快手（“可灵”视频生成大模型）等企业的核心文娱业务线为代表。这类主体依托母公司的底层通用大模型能力，将其无缝接入自身庞大的泛娱乐产品矩阵，在自有的流量生态内，形成了“数据—模型—应用”的正向飞轮效应。

第二类是**垂直探索者**，以趣丸集团（趣丸千音、天谱乐、趣丸万相）、MiniMax（海螺 AI、星野）等为代表，构成了当前具备较强垂直场景价值生态的阵地。这类企业避开了在底层通用算力上的同质化竞争，将重心锚定在特定文娱场景的纵深发展上，探索在行业深水区实现价值化。

第三类是**基建赋能者**，以百度智能云、智谱 AI 和月之暗面等底层模型与算力提供方为代表。它们如同产业的“水电站”，致力于打造泛化能力较强的底层基座，通过 API 调用或 MaaS（模型即服务）模式，为大文娱赛道提供多模态生成的基础设施。

第四类是**单元创新者**，以众多切入细分领域（如单点 AI 配音、AI 换脸小工具）的初创企业为主。它们嗅觉敏锐、产品迭

代较快，是赛道活力的重要来源。但受限于单一的工具属性和较浅的用户留存，目前大多仍在探索价值化路径。

三、技术全景：驱动大文娱产业变革的核心 AI 技术

人工智能技术对大文娱产业的赋能，已从早期单点工具的浅层介入，演进为多技术协同驱动的模式。本节从四个技术维度，系统梳理驱动大文娱产业变革的核心技术图谱，并对各方向的技术成熟度、工程化壁垒及产业化进程进行分析。

（一）多模态大模型：打破内容载体壁垒的核心基座

1. 文本与图像生成：规模化成熟的基础设施

文生文与文生图技术是“AI+大文娱”最早实现规模化商用的领域之一，现已成为内容生成技术体系的基础底座，完成了从“实验室验证”到产业级“新基础设施”的角色转变，降低了图文内容的生产边际成本。

在文本生成方向，大语言模型已从“短文本浅层问答”向“长文本连贯控制”的阶段转变。依托超长上下文处理能力的突破，以及 RAG（检索增强生成）技术的成熟应用，模型已具备处理宏大逻辑推演与保持角色长期设定的能力。在规模化应用端，该技术已深度嵌入大文娱产业的核心创作流。网文平台、影视制作公司正在规模化引入大模型，进行剧本大纲生成、背景设定以及游戏 NPC 大规模动态交互对话库的批量构建。这不仅大幅缩短了内容孵化周期，更将核心创作者从繁复的基础案头工作中解

放，赋能其专注于高阶创意。

在图像生成方向，基于扩散模型的底层架构迭代，技术已摆脱早期依靠提示词“盲抽卡”的“基础语义生成”阶段，迈入高度可控的“图像精细化微调”时代。在规模化应用端，随着 ControlNet、LoRA、IP-Adapter 等条件控制与微调技术的普及，AI 在特定画风复刻、人体姿态精准控制、光影结构保留等方面已逐渐成熟。目前，AI 图像生成已成为 UI 资产量产、影视分镜可视化预演、互动小说插图及营销海报高频生成的“标配工具”，帮助企业在部分视觉资产生产环节显著提升效率、降低成本，其作为“工业级视觉底座”正在释放市场化价值。

整体而言，文本与图像生成技术的规模化成熟，不仅完成了单点效率的变革，更通过“图文协同”构筑“AI+大文娱”内容生成坚实的技术“底座”。

2. 跨模态理解与生成：前沿攻坚的技术突破口

跨模态理解与生成是当前“AI+大文娱”领域技术演进最具战略价值的前沿方向之一。其核心难点在于“语义对齐”，即如何将文本的离散符号、音频的连续波形与视频的时空张量，精准映射至统一的语义计算空间。当前，这种跨模态突破正从单一的图文领域，向视频、音频等多维度全面扩散，为用户带来了 AI 创作自由度与体验升级。

在音乐领域，跨模态理解与生成打破了传统音频创作的专业

门槛。目前以趣丸“天谱乐”等为代表的音乐大模型已实现了从“自然语言语义”到“旋律特征、编曲风格、情感基调”的跨模态联合理解与自动化生成。这种多模态底座支撑，不仅大幅提升了泛人群的 AI 音乐创作体验，更在影视配乐自动化、短视频音轨智能匹配等高频场景中展现出一定程度的市场潜力。

在视频生成领域，基于“时空注意力机制+扩散过程”或深度融合 **Transformer** 的混合架构成为主流范式。部分模型已在短片级视频生成、广告素材预演、微短剧分镜等场景进入商业试用或规模化使用阶段，率先在微短剧开发、广告素材效果预演中落地。但向院线级高阶场景渗透仍面临两重挑战：一是物理世界规律的高保真模拟（如光影折射、流体动力学的失真现象）；二是长视频时序的绝对连贯性（随生成时长延伸，帧间漂移与角色“变脸”概率呈指数上升）。

（二）智能语音与 AI 音频技术：促进听觉交互与内容生产的底层引擎

1. 高拟真语音合成与跨语种复刻技术

智能语音是“AI+大文娱”产业化落地较为深入的方向之一，已覆盖有声读物、短视频配音、直播互动、影视译制等核心场景。当前技术的演进主线是持续突破情感表现力、多语种零样本生成与 AI 语音翻译。

智能语音与 AI 音频技术的演进，正沿着“高拟真表现力”

与“跨语种泛化性”两条主线，全面跨越从实验室算法向项目工程落地的鸿沟，并在四个核心维度构筑了较高的技术壁垒：首先，在情感表现与控制颗粒度上，语音大模型已具备对哽咽、笑腔、叹气、颤音等细颗粒度情感音色的可控生成能力，推动机器语音从传统的“固定音调”向“语境感知的动态情绪响应”转变，这已成为高端语音产品的核心差异化壁垒。其次，在音色复刻的工程化方面，高保真音色复刻所需的参考音频时长已压缩至 10 秒以内，直接赋能版权声音资产的数字化保存与有声书平台音色的规模化复用。再次，在多语种少样本生成与跨语种翻译上，以 MaskGCT 等新一代架构为代表，模型不仅支持基于少样本的小语种及方言生成，其跨语种语音复刻技术更能在保留原始音色特征与情感质感的前提下，直接以源语言音色生成目标语言，将影视剧出海的本地化译制周期从数周大幅压缩至数小时。最后，在实时交互的时效性上，以 MiniMax Speech 为代表的语音大模型端到端语音合成的时延已成功突破数百毫秒级响应区间，为数字人及智能体在复杂互娱场景中的“无缝实时互动”提供了技术参考。

2. AI 音乐生成模型的算法演进与可控性评估

在算法演进维度，AI 音乐生成经历了两条早期路线的融合。符号路线（以 MIDI 为代表的乐谱类符号表示）结构可控、可编辑，但需要额外合成环节，音色相对机械化；音频路线（直接建

模频谱/波形）音色真实、表现力强，但早期音质粗糙，也缺乏长程音乐结构。当前主流模型走向端到端生成——并非抛弃中间表征，而是用模型自学的音频表征（离散 **token** 或连续 **latent**）取代人工定义的符号表征，以此为桥直接建立从“文本/音乐描述”到“高保真完整音频”的映射，兼顾结构可控、音乐性与音质。

在工程能力验证方面，音乐大模型已跨过“能不能生成”的阶段。音频域生成早期有三个典型问题：一是音质粗糙，人声与乐器不够真实；二是长曲结构松散，段落之间连贯性差；三是歌词多字漏字，唱词与输入不一致。归结起来，是音乐性（旋律、节奏、段落结构、歌词一致性）与音质（音色真实、细节清晰）之间难以平衡——早期的模型迭代，基本都在解决“生成质量”这一件事。

目前主流模型的生成质量已达到可用水准，行业竞争进入新阶段。与语言模型的演进路径相似，音乐生成模型也开始具备类似的思考、规划与编辑调整能力：先规划歌曲的整体结构，再生成具体内容；听得懂“减少鼓点”、“突出某种乐器”这类具体的修改意见，支持旋律与音色的精准控制；并能在保留原曲动机与辨识度的前提下，完成翻唱、改编、局部编辑等修改。竞争焦点由此从“第一次生成是否惊艳”，转向“能否把一首歌持续改到满意”。与此同时，创作入口也在从文本向图片、视频等多模态扩展，视频自动配乐等能力开始进入短视频、影视等内容生产

场景。AI 音乐正由“一键出结果”的工具，走向“可反复打磨”的创作 workflow。

然而，AI 音乐生成技术向影视级、唱片级等专业编曲 workflow 渗透时，仍面临“多轨独立可控性”这一技术壁垒。现阶段的端到端模型主要输出混合单轨音频，缺乏原生分轨生成与编辑能力。如何在生成阶段实现人声、弦乐、鼓点等特定乐器声部的解耦，并在维持全曲听感平衡的前提下，支持特定声部的局部重绘与音量独立控制，是加大该技术在专业级文娱内容生产应用时亟待突破的工程难点。

（三）数字人与空间计算技术：拓展虚实融合内容边界

1. 高保真虚拟形象的实时渲染与纯视觉驱动

数字人技术的产业价值，高度依赖于驱动方案的轻量化演进。技术路径已从专业级多模态传感器动捕方案向单目 RGB 摄像头纯视觉驱动延伸，推动动捕能力向消费级场景逐步下沉。依托深度学习驱动的面部关键点检测、三维姿态估计与骨骼自动绑定技术，当前主流数字人系统已能仅凭普通摄像头输入实时驱动数字形象，显著降低硬件部署门槛，为直播带货数字人、在线教育虚拟讲师等高并发场景的规模化落地提供关键支撑。

在大模型的赋能下，数字人的智能化演进体现在三个层面：一是口型音频毫秒级同步，大模型可基于语音特征实时预测高精度口型；二是微表情的自动化生成，使数字人情感表达从“规则

驱动的固定表情序列”向“语境感知的动态情绪响应”转变；三是**端侧轻量化部署**，通过模型量化、知识蒸馏与神经架构搜索等技术的综合应用，将实时渲染能力下沉至消费级移动终端，为虚拟形象全民创作提供技术基础。

2. 3D 资产的自动化生成与轻量化部署

三维数字资产构建是大文娱产业数字化转型的重资产环节。基于 NeRF（神经辐射场）与 3DGS（三维高斯泼溅）的最新算法演进，结合扩散模型的生成能力，业界已打通了“图生 3D/文生 3D”的技术闭环。

产业赋能层面，上述技术已在多个典型场景得到验证：在沉浸式演艺领域，大幅降低虚拟演唱会、XR 互动演艺的舞台资产制作成本；在数字典藏与实体衍生领域，“扫描建模—AI 优化—三维打印”全链路开辟了从数字生成到实体消费的新型商业闭环；在移动端部署层面，依托 LOD 自适应与 WebGL/WebGPU 渲染加速，轻量化三维内容正向消费级终端普及推进。

（四）智能体应用：打造沉浸式文娱互动生态

1. 单体智能体的技术创新

智能体技术是大文娱内容消费从“单向接受”向“双向情感互动”转变的核心引擎。通过长文本记忆机制与 RAG（检索增强生成）技术的融合，大模型赋予了 NPC 或虚拟伴侣持久的“长期记忆”。结合深度强化学习，情感陪伴类智能体实现了从

“单轮指令问答”向“多轮主动拟真交互”的技术变化，能够敏锐感知用户情绪并主动提供情感反馈。

2. 多智能体协同与沉浸式互动生态构建

多智能体协同技术的意义在于，构建具备自组织叙事能力的沉浸式大文娱互动生态。技术框架层面，多智能体系统依托任务分解、角色专属知识库与智能体间通信协议实现自主协作。以沉浸式音频互动剧为例，叙事导演智能体、角色扮演智能体与环境音效智能体可实现实时协同，根据用户行为动态生成差异化剧情走向与沉浸式听觉环境，构建传统静态内容无法企及的个性化叙事体验。

智能体技术在大文娱产业中的落地仍面临三项瓶颈：一是**协同一致性维护**，即并发交互中如何保持叙事等内容逻辑的一致性；二是**实时性保障**，即高并发场景下多智能体调度延迟控制对推理算力提出高要求；三是**内容安全和数字伦理**，即开放式互动叙事生成中需有效防范不良内容，且平台治理措施必须符合我国有关政策法规监管规定，加强动态更新、确保安全合规，这是产业在转型发展中的必答题。

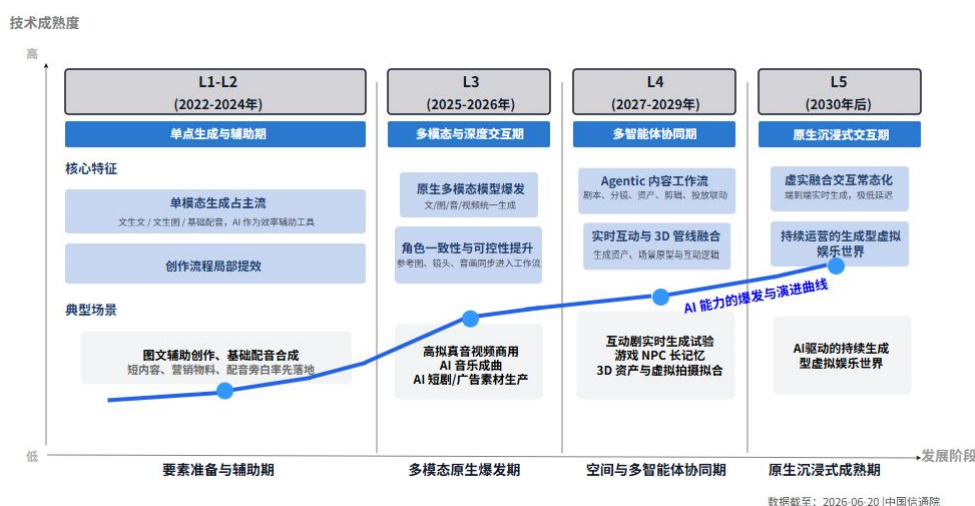


图 2 国内“AI+大文娱”技术演进路径图

四、应用场景：AI 赋能大文娱产业的典型场景与模式创新

本章按照语音、视频、音乐、写作和数字人五个领域依次展开，结合各细分领域的市场需求与产业痛点，系统梳理 AI 技术在不同应用场景下的赋能路径与商业实践。案例遴选以商业模式差异为核心维度，在每一领域分别选取代表不同市场参与逻辑的典型产品服务，涵盖垂直场景深耕、底层能力开放输出、平台生态融合等差异化路径，同时兼顾 B 端专业市场与 C 端大众市场的双线视角，力求在 AI 技术方向下呈现多元化的产业落地图景，为行业参与者提供具有横向对比价值的参考样本。

（一）“AI+语音”：打造沉浸式听觉新体验

1. “情绪价值”驱动与高保真拟真交互需求增长

近年来，泛娱乐音频市场规模持续增长，传统 TTS（将文本自动转换为语音的合成技术）正在面临体验天花板。用户对音频内容的消费诉求，正经历从“信息获取”向“情感陪伴”的转变。然而，传统基于拼接合成或早期参数合成的 TTS 技术，普遍存

在机械感强、语调平淡、情感表达僵化等“恐怖谷效应”，难以满足受众日益增长的“情感共鸣”需求。同时，传统配音工业流面临着“质量、效率、成本”的不可能三角，面临选角门槛高、排戏对戏效率低、录制修改周期长等痛点，制约了文娱产业的产能释放。

新一代语音大模型正推动数字语音生产方式发生变化，探索实现从“机械复述”到“情感演绎”，精准还原“喜怒哀乐”等情感，推动提升“高保真拟真交互”的体验，同时把声音变成可复用、可授权、可跨语种迁移的内容资产。

2. 典型实践案例

在“AI+语音”的商业化落地进程中，行业呈现出“通用基座效率优势”与“垂直模型深耕场景”并行的双线发展态势。本节选取三个产品服务：“星火”语音大模型、“趣丸千音”AI语音平台和“豆包”语音大模型，展现AI语音在不同维度下如何打破技术壁垒，实现价值增量。

● 典型实践一：“星火”语音大模型：深耕严苛专业标准，赋能融媒高拟真播报

相较于C端（消费端）泛娱乐应用对“互动性与网感”的侧重，政企机关与广电媒体市场在引入AI语音技术时，对信息传递的准确性、严肃性与技术规范有着更为严格的要求。作为国内智能语音领域的代表性企业，科大讯飞依托“讯飞星火语音大模

型”，聚焦专业文娱与融媒体生产场景，构建了具备高标准与高可靠性的行业解决方案。

（1）适配广电级标准的智能配音与自动化音频播发应用

在电视台及政企宣传部门，传统的播报工作高度依赖专业采编播人员，制作流转周期较长，且难以满足突发新闻或资讯全天候的播报需求。依托大模型底座的自然语言理解与语音合成技术，科大讯飞推出了广电级 AI 智能配音与音频生成系统。该系统能够精准分析复杂的新闻文本，自动生成发音规范、停连准确、基调庄重的新闻语音。

目前，该项纯音频生成技术已广泛应用于图文资讯自动转音频、政务短视频一键配音及全天候在线的 AI 广播频道。例如，2025 年全国两会期间，讯飞智作投放 200 余款虚拟主播，服务全国超 300 家地方媒体，对应相关视频总播放量超 2 亿次。有关新型声音技术的应用有效补齐了基层媒体在专业声音产能上的短板，在保障内容严肃性与播出安全性的前提下，提升了政务宣传与新闻资讯的数字化生产效能。

（2）多语种及方言保护，赋能影视及专业译制

在专业性极强的高端纪录片配音、海外影视正剧译制及地方戏曲或方言保护等方面，星火语音大模型展现了其在“多语种与方言微调”上的积淀。科大讯飞语音识别技术已于 2024 年实现了全国地级市方言全覆盖，涵盖了全国 288 个地市的 202 种方言。

同时，在跨国译制与出海场景中，最新升级的星火大模型支持 100 余种语言的高精度语音识别以及 55 种语言的语音合成。其合成音频在信噪比、发音自然度及专业术语的停顿处理上，比较适合政务、新闻、广电等场景，形成了面向 G 端和 B 端市场的差异化优势。

● 典型实践二：“趣丸千音” AI 语音平台：以情绪拟人为核心突破，构建高情感浓度语音交互

趣丸科技依托 AI 技术底蕴，推出“趣丸千音（All Voice Lab）” AI 配音平台，以“高情感浓度与高保真拟人”为核心技术突破，将传统高度依赖人工经验的“感性配音艺术”转化为标准、可控的“数字化工业流”，为文娱行业推动“AI+语音”转型发展提供新实践。

（1）搭载开源语音模型 MaskGCT，构建语音高保真能力

趣丸千音的情感表现力，源于其底层的大模型基座。平台基于趣丸科技与香港中文大学（深圳）联合研发的 MaskGCT 语音大模型构建。该模型在若干 TTS 基准上表现较优，在包含 10 万小时多语言数据的 Emilia 数据集上进行训练。2024 年 10 月，在三个国际 TTS 基准数据集上均达到了行业较高水平，并成功入选深度学习领域的知名国际会议。

该模型突破了传统拼接合成的技术瓶颈，创作者可在较短样本基础上用较短时间实现音色复刻和风格迁移。同时，模型能够

精细化解析语音和文本中的情绪起伏，在生成语音时精准匹配“喜怒哀乐”等情感及自然的呼吸气口，其情绪拟真度与表现力在部分标准化配音、短视频解说和译制场景中，已接近专业配音的可用水准。

（2）重构配音 workflow，打造大文娱内容生产新范式

针对大文娱 B 端生产场景，趣丸千音打造了全自动化的高效配音 workflow。系统可实现多音轨分离、台词文本智能提取与角色识别，并推出“AI 音色一键配音”功能，内置数百款专业高情感音色供自动匹配。

通过“配音节奏调控”与“口型精准匹配”技术，趣丸千音使得 AI 配音的视听融合度达到工业级水准。根据平台运行数据验证，通过一站式 AI workflow，企业的内容制作成本较传统人工配音大幅降低，整体制作速度明显提升，实现了从“手工作坊”向“AI 规模化量产”的提升。

（3）AI 解码传统文化与泛文娱，构建多元场景商业闭环

在国内多元化文娱场景中，趣丸千音正加大探索价值化进程。在动态漫与短剧领域，趣丸千音通过其一站式视频翻译与 AI 配音平台，为内容平台与创作者提供了从“字幕擦除—翻译—配音—后期—成片交付—海外自动分发”的全链路自动化解决方案。

此外，在“AI+文娱”的跨界融合上，趣丸千音打造了具有社会价值的行业标杆。例如，由广东省委宣传部与趣丸千音等联

合出品的“寻迹文化岭南主题宣传·广东 100 个馆藏故事”项目中，趣丸千音深度参与了历史纪录片（如《中英街历史博物馆》）的 AI 配音制作。该系列视频全网累计播放量突破 100 万+，趣丸千音也凭借该项目获评“2025 年广东‘人工智能+文旅’应用场景典型案例”，体现了 AI 语音在文化传播中的应用潜力。

● 典型实践三：“豆包”语音大模型：夯实模型支撑，推动短视频迸发创作者经济活力

在大文娱赛道，火山引擎依托“豆包”大模型，在用户生成内容（UGC）市场占有一席之地。其语音大模型作为底层技术支撑，深度整合于各类 C 端数字应用之中，通过规模化赋能提升了创作者经济的发展活力。

（1）赋能 C 端创作工具，推动语音合成技术的规模化应用

作为底层技术支持，“豆包”语音大模型通过 API（应用程序接口）深度接入音视频创作平台。通过提供“文本输入至语音生成”的模块化功能，大幅降低了用户的技术操作门槛，实现了自然对话式语音的快速合成。基于对海量应用场景数据的深度学习，该模型构建了包含地方方言、特色人声等在内的多元化音色矩阵，有效适配了不同类型短视频的情感表达需求。这种将 AI 能力组件化、工具化的方式，促进了视听内容的规模化生产与高频互动。

（2）降低专业制作壁垒，创新 UGC 内容生产与分发生态

在微短剧、视频日志（Vlog）、影视解说等 UGC 生态中，个体创作者长期面临音频制作成本高、专业配音资源匮乏等痛点。“豆包”语音大模型凭借高并发处理能力与普惠性的技术调用机制，普通用户能够低成本获取接近专业级的音频生成能力。这一赋能机制丰富了短视频平台的高质量内容供给。火山引擎借此将智能语音技术转化为底层的通用数字基础设施，形成了从“底层模型供给”到“前端工具赋能”，再到“创作生态”的产业链条。

（二）“AI+视频”：影视工业与微短剧的“视觉升级”

1. 影视工业降本增效与原生 AI 视频活力涌现

近年来，随着移动互联网深度普及与用户视觉审美不断提升，微短剧、网络剧集、商业视频等内容形态的市场需求持续增加。国家广播电视总局数据显示，2025 年国内上线微短剧数量已超过 3.3 万部，市场规模突破千亿元人民币，用户规模接近 7 亿。然而，传统视频生产长期受制于实地取景、特效渲染、演员档期协调等重物理资源约束，制约了产能的规模化释放。

早期的 AI 视频生成技术往往只能输出 2 至 4 秒的低分辨率片段，且存在“时空不一致”与“物理规律扭曲”（如人物瞬间变脸、物体诡异形变等）问题。这导致 AI 视频大多停留在“盲盒式抽卡”的玩具阶段，难以真正切入严苛的影视工业化生产线。

新一代视频生成大模型的技术正打破上述瓶颈。部分领先视

频模型已能生成更长时长、更高分辨率的视频片段。这一技术正在推动“脚本辅助生成分镜、分镜生成视频素材、人工再剪辑与审美把控”的新型 workflow。

AI 视频工具的引入，使得传统的选角、场地租赁、绿幕抠像等重物理环节中的部分工作正在被“数字化替代”。据测算，在部分微短剧或广告素材生产场景中，AI 视频 workflow 可显著压缩拍摄、置景和后期成本。从导演调度到后期渲染，AI 正重塑影视工业链条，不仅大幅降低了专业内容制作门槛，更催生了“AI 导演”“AI 调度师”等新兴职业，推动视频内容生产效率持续提升。

2. 典型实践案例

在“AI+视频”的产业化落地进程中，核心痛点已从最初的能否生成转变为如今的能否实现工业级高保真，能否提升 IP 一致性和可控性。本节围绕上述核心痛点，选取两个产品服务：

“可灵”视频生成大模型和“Vidu”视频大模型，以呈现 AI 视频技术从底层支撑到工业应用的产业链条。

● 典型实践一：“可灵”视频生成大模型：以物理模拟与长视频能力为核心，创新微短剧生态

在视频内容生产对“视觉真实感与叙事连贯性”要求不断攀升的背景下，视频生成大模型“可灵”以“物理高保真模拟与长视频连贯生成”为核心突破，探索将传统依赖实地拍摄与特效制

作的视频工业流程转化为标准、可控的“AI 数字化生产线”。

（1）自研时空扩散架构，探索实现物理模拟与长视频生成的双重突破

可灵采用 3D 时空联合建模架构，核心包括 3D Causal VAE（3D 因果变分自编码器，一种用于对视频的时间和空间维度进行联合压缩与特征提取的模型组件）等时空压缩组件，并结合扩散/Transformer 类生成框架提升视频时序一致性，将时间与空间维度纳入统一建模框架，改变了传统模型将时序与空间割裂处理的架构局限。物理模拟层面，内置仿真引擎可精准还原液体流动等复杂动态场景，运动连贯性与材质真实感已在核心评测指标上接近传统实拍效果。在长视频生成层面，可灵支持最高 2 分钟、1080p 分辨率及 30fps 超高帧率的端到端生成。通过“首尾帧锁定”与“多镜头连贯生成”技术，确保跨镜头叙事的视觉统一性，改善了长视频生成中普遍存在的帧间抖动与角色漂移问题。创作者还可通过“镜头语言指令”对推、拉、摇、移等专业运镜方式进行精细化干预，提升了成片的电影级质感。

（2）以“业态深度绑定”切入微短剧赛道，推动实现 AI 短剧原生商业闭环

可灵的核心商业逻辑在于将能力直接嵌入微短剧生产流，显著降低传统拍摄流程门槛。通过将选角、实地取景、绿幕特效等部分重资源环节以 AI 进行替代，有效节约人力成本。截至 2026

年6月，可灵AI全球用户规模已突破1亿，覆盖224个国家和地区，服务过近5万家的企业客户和开发者。

● 典型实践二：“Vidu”视频大模型：凭借多主体一致性与极速生成，开辟动画工业量产新路径

“角色变脸、场景不连贯、多人物混淆”是长期困扰动画工业AI转型的行业痛点。脱胎于清华大学的生数科技，依托“Vidu”大模型，以底层架构创新为支点，聚焦影视动画IP一致性与风格化量产场景，构建了高保真、高可控的动画工业量产新路径。

（1）U-ViT 架构驱动多主体一致性，攻克IP量产技术瓶颈

Vidu 基于 U-ViT（U 型视觉 Transformer，一种通过跨层特征复用提升生成过程中特征一致性的视觉生成模型架构）底层架构创新，通过跨层特征复用机制在生成过程中持续锚定角色外观特征，从架构层面一定程度上解决了长序列生成中普遍存在的时序特征漂移问题。在此基础上，Vidu “多参考图联合输入”技术支持同时输入主角、配角、场景等多组参考图像，自动提取并锁定各主体的五官、服装、色调等关键视觉特征。即使在复杂的交互动作或剧烈运镜下，也能实现多主体在不同场景与动作状态下的高保真外观复现，适合动漫 IP、广告电商、品牌角色和风格化短片等对主体一致性要求较高的场景。U-ViT 架构带来的单步极速生成能力，配合内置多类主流动画风格模板与批量并行处理机

制，为动画 IP 的高频次、大体量连载生产提供了技术工具底座。

（2）以“角色或 IP 驱动”切入动漫工业，构建“二次元”IP 视频化量产闭环

针对传统动画制作中原画绘制、中间帧补全、逐帧上色等高度依赖人工的核心环节，Vidu 将其整合为“分镜脚本直生动画视频”的端到端单步流程，使动画 IP 视频化生产从“手工精绘”向“AI 高频量产”实现转变。在 IP 衍生开发场景中，多主体一致性能力使同一角色 IP 可快速扩展至系列短片、衍生其他作品等多类视频化产品。目前 Vidu 已与多家动画公司建立合作，基于 Vidu 量产的动画系列在主流平台持续获得高播放量，提供了国内 AI 驱动动画 IP 视频化量产并实现价值化的参考实践。

（三）“AI+音乐”：以技术演进推动音乐创作的大众化探索

1. 多模态输入拓展创作方式，AI 音乐进入应用探索阶段

音乐创作涉及乐理知识、编曲技能与版权成本等多重门槛，长期以来主要由专业人士承担。近年来，随着生成式人工智能技术发展，AI 音乐生成能力持续演进。早期形态以“AI 翻唱”或单一风格套用为主，目前已逐步支持文本、图像、音频、视频等多模态输入；生成内容也从简单的音频片段，向词曲结构相对完整、人声演唱与编曲混音一体的成品歌曲延伸，在部分场景中已具备一定的应用价值。这一变化的技术基础，主要在于底层模型在长序列音乐语义建模、音频空间建模与多模态语义对齐等方向

的持续改进，使模型逐步具备从多种形态输入到较完整音频输出的映射能力。

从市场应用看，AI 音乐在面向个人用户（C 端）与面向企业用户（B 端）两个方向同步探索。面向 C 端，以自然语言、图片或视频为输入的“大众写歌”，正在成为一种新的社交娱乐与个人表达方式；面向 B 端，AI 音乐在交付周期、定制成本等方面具有一定优势，可在短剧配乐、游戏音效、广告背景音乐等场景中对传统模式形成补充。总体来看，随着生成质量提升与平台授权机制逐步完善，行业正由技术验证向规模化应用探索过渡，相关版权规则和商业模式尚在持续探索之中。

2. 典型实践案例

当前，AI 音乐的应用落地正从早期试点逐步走向多路径、多层次的探索格局。本节案例以“创作能力来源×应用路径”为切分维度，选取三个产品服务，背后的三种模式在价值链中分别侧重创作生态、底层能力供给与平台价值化，反映出 AI 音乐从技术探索走向更广泛应用的多元路径。

● 典型实践一：“Tempolor V4.7”天谱乐多模态音乐模型：

“模型+应用+硬件”协同探索

趣丸科技从早期的演唱类应用出发，逐步搭建起覆盖底层模型、C 端创作应用与智能硬件终端的产品体系，在 AI 音乐领域前瞻探索“模型—应用—硬件”协同布局。

（1）自研多模态音乐模型“天谱乐”，夯实技术基础

趣丸科技 AI 音乐布局的底层支撑，是其自研的多模态音乐生成模型“天谱乐”。“天谱乐”于 2024 年 7 月实现支持文生音乐、音频生音乐，还首创图片、视频生成音乐的能力，实现多模态音乐生成产品化应用。借助兼具音乐创作与人工智能研发背景的人才，“天谱乐”在模型训练与产品设计中融入对音乐结构等的专业理解。在技术方面，构建了以音乐语义为基础的级联式生成架构，可支持生成最长 5 分钟的成品歌曲；在中文人声唱词生成能力上也有“真假难辨”的效果。相关项目在地方组织举办的算法大赛、AI 游戏及应用创新大赛中表现突出。

（2）推出对话式创作智能体与生成式 AI 吉他，探索软硬件协同

在底层模型基础上，趣丸科技进一步向应用层与硬件层延伸，于 2025 年 9 月推出对话式音乐创作 Agent “Tunee”与生成式 AI 吉他“天谱乐 TemPolor”，在使用效率与体验两个方向同步探索。

Tunee 基于多代理（Multi-Agent）架构构建，支持音乐生成、MV 制作等功能。用户无需掌握复杂的提示词技巧，以自然语言描述创作意图后，Tunee 可追问细节并输出音乐方案。付费订阅用户可获得所生成音乐的商业使用授权。当前，Tunee 已在专业创作者与大众用户中形成一定应用。例如，有创作者借其制作原创 AI 音乐作品，并获得较高的平台播放量。

天谱乐 TemPolor AI 吉他将模型能力嵌入实体乐器终端，是探索将生成式 AI 技术融入吉他演奏体验的智能乐器。用户上传喜爱的歌曲或哼唱片段后，模型可在数秒内生成对应的指弹旋律，并通过琴身内置屏幕以动态形式进行指引。在设计层面，还获得多项国际设计奖项。

（3）坚持全链路布局，多年积累探索大众化路径

2019 年，趣丸科技推出唱鸭 App，作为“新一代 AI+大众应用型音乐创作平台”，普及“无弦弹唱”玩法，让不具备乐器基础的用户也能实现自弹自唱，曾入选文化和旅游部“文化和旅游数字化创新实践十佳案例”；随后，趣丸科技开始布局 AI 音乐生成技术研发，并于 2024 年发布“天谱乐”；2025 年模型持续升级，同年 9 月全球公测 Tunee：集成视频生成能力，进一步打通从音乐创作到 MV 制作的内容生产流程。

从突破演唱门槛，到对音乐创作的支撑，再到以对话方式降低操作门槛、以智能硬件拓展演奏场景，趣丸科技在 AI 音乐领域形成了从技术研发到终端应用的全链条覆盖。这一“模型支撑应用、应用积累数据、硬件拓展场景”的协同布局，为深耕垂直行业提供了可参考的实践路径。

● 典型实践二：“MiniMax Music”音乐生成模型：以底层通用算法赋能的多模态应用探索

MiniMax 以通用多模态大模型研发为主要方向，相较于聚焦单一音乐赛道的垂直模式，MiniMax 更侧重以底层算法能力为支撑，向多类内容生成场景提供能力输出。

（1）底层模型持续迭代，兼顾可控性、声学品质与多模态协同

据企业公开信息，MiniMax 于 2026 年 4 月正式发布新一代音乐生成模型 Music 2.6，对底层生成架构进行了重构。在生成效率方面，首包延迟压缩至 20 秒以内，有效改善了此前版本生成周期较长的问题。在可控性方面，Music 2.6 实现了三项关键突破：一是 BPM（每分钟节拍数）与调性精准锁定，创作者设定的节拍速度与调式将被有效还原；二是段落结构精细控制，模型对编曲结构的理解力显著增强；三是情绪表达进一步优化，人声与乐器音色的情感还原度进一步提升。在声学品质方面，Music 2.6 优化了音色自然度与中低频声学表现，扩充了乐器种类的多样性，并改善了整体时序上的演奏逻辑。

（2）以综合应用切入市场，探索多模态内容生成路径

在应用路径上，MiniMax 旗下“海螺 AI”多模态智能创作平台整合 Music 2.6 音乐生成能力，并与文本、图像、视频生成等能力形成协同，用户可在同一平台内完成从文字描述到音乐、

视频的内容制作流程，有效降低跨平台使用成本。在功能扩展方面，Music 2.6 推出 Cover 跨风格翻唱功能，支持用户上传参考音频后进行旋律特征提取与跨风格迁移。这种“自有应用验证能力、开放 API 拓展生态”的方式，是通用大模型企业在垂直内容方向探索应用的一种典型路径。

● 典型实践三：“Mureka”音乐大模型：以“模型+平台”协同探索 AI 音乐创作与分发生态

昆仑万维将 Mureka 音乐大模型作为其 AIGC 内容矩阵的核心组成部分之一，以 AI 音乐平台 Mureka 为载体，探索从音乐创作、分发到价值化的全链路应用路径。

（1）依托专业架构，强化音乐生成的精准可控性

技术层面，Mureka 以持续演进的 MusiCoT（音乐思维链）体系为核心，进一步协同音乐结构、歌词语义与声音呈现，在段落内文本控制精准度、人声情感表达与混音质感等方面，实现一定程度的升级。Mureka 音乐大模型在生成效率、混音质量等维度持续进化，逐步推动 AI 音乐从“能够生成一首歌”向“按创作意图完成一首歌”迈进。

（2）以平台化路径探索创作、分发与价值化协同

在应用路径上，昆仑万维以 Mureka 平台为载体，构建涵盖创作工具、内容分发与用户运营的 AI 音乐生态。据公开信息，Mureka2025 年化流水约 1200 万美元，一定程度上验证了 AI 音

乐付费订阅模式的商业可行性。在内容生态层面，Mureka 平台整合创作与分发功能，用户可在平台内完成从生成到发布的流程，探索“创作—分发—价值化”闭环路径。

（四）“AI+写作”：文本创作从人工主导向人机协同演进

1. 文本生成能力持续提升，内容创作向人机协同方向发展

近年来，大语言模型在文本创作领域的能力持续提升，从早期的碎片化问答与短文本生成，逐步向长篇叙事的逻辑连贯性与复杂结构延伸，为文本生成技术向专业创作场景渗透提供了基础。

从市场应用看，AI 文本生成工具同样在 C 端与 B 端两个方向发展。面向 C 端，大众化“智能写作”正逐步成为个体表达的辅助工具，用户借助 AI 完成日常表达、创意记录与轻量创作；在 B 端，围绕网络文学、营销文案、专业文书等场景，AI 的定位从早期执行具体写作指令的辅助工具，逐步向参与选题策划、结构设计与内容迭代的协作方向演进。随着这一变化持续，行业关注点也从单纯的文本生成质量，转向高质量语料的积累运用与人机协同的实际价值探索。

2. 典型实践案例

本节以“文本体裁特征×核心应用场景”为切分维度，选取两个产品服务，分别对应创意叙事和商业应用两类文本需求，共同反映 AI 写作在不同场景下的差异化应用路径。

● 典型实践一：“阅文妙笔”模型：赋能“作家助手”，探索网文人机协同创作

网络文学的人物设定、剧情线索等信息量较大，增加了长篇创作中内容设定一致性的难度。阅文集团自研网文垂类模型“阅文妙笔”并落地配套工具“作家助手妙笔版”，探索模型赋能长篇叙事创作的路径。

（1）自研网文垂类模型，助力解决内容设定一致性问题

相较于通用模型，“阅文妙笔”依托平台多年积累的网文语料完成垂类训练，在叙事解析、逻辑推演、文风适配等方面具备较强的场景适配能力。基于该模型的“作家助手妙笔版”围绕人物、场景、情节等模块搭建功能，可自动生成设定素材并校验逻辑冲突，有助于缓解长篇连载中的设定割裂问题。自接入 DeepSeek 以来，“作家助手”日活用户增长超过 30%，其中 AI 功能的周使用率超过 50%。

（2）以 AI 辅助创作贯通内容链条，支撑 IP 开发

从产业链看，“作家助手”不仅用于单篇作品创作，也是阅文“以 AIGC 辅助原创、贯通 IP 链条”布局的载体之一，平台已将相关技术应用于网文漫改，并向有声、动画等业态拓展。2025 年阅文平台内容生态持续扩张，全年吸引 40 万名新作家入驻，诞生超 80 万本小说，网络文学内容供给规模进一步扩大。

● 典型实践二：阿里妈妈“悉语”智能文案：支持商家提升内容生产与投放效率

在电商营销场景中，文案是连接商品与消费者的重要触点。阿里妈妈依托智能文案生成能力，为商家提供内容生产工具。

（1）快速生成多风格文案，满足规模化生产需求

阿里妈妈将大语言模型嵌入商家经营工具，商家录入商品链接等信息之后，系统即可产出多种风格文案，适配不同投放渠道与客群。2025年，在电商促销节点，不少商家依托AI智能创意批量产出内容营销素材，节省了制作成本。

（2）结合大数据预测，提升营销推广效率

与仅完成文本输出的通用工具相比，阿里妈妈“悉语”智能文案的差异化之处在于与平台大数据预测模型结合，进而提升营销效果。在落地形态上，智能文案生成能力已嵌入阿里妈妈商家经营流程，覆盖商品详情文案、推广广告语、直播间脚本等场景，并与选品、投放、数据分析等模块形成协同，成为商家日常经营中的一项辅助能力，有助于中小商家降低内容生产成本、提升营销效率。

（五）“AI+数字人”：加快多模态交互应用拓展

1. 多模态技术持续演进，数字人应用场景加快拓展

数字人技术的发展，经历了从早期动作捕捉驱动到如今大语言模型与多模态技术深度融合的阶段性演进。据中国互联网协会

《中国数字人发展报告（2024）》，数字人（Digital Human）是通过建模等多种数字智能技术创建，具备人类外观形象、声音语言、肢体动作或思维功能，并能在大模型支撑下，形成学习、生成、互动等功能的数字智能体（AI Agent）。

在早期阶段，无论是用于影视特效的虚拟角色，还是用于展览展示的虚拟主播，其交互方式大多依赖预设脚本或动作捕捉设备驱动，交互逻辑相对固定，制作成本较高，在实时互动等场景中的适配能力相对有限。

随着大语言模型推理能力的提升与语音、图像等多模态感知技术的发展，数字人的交互能力发生变化。在 B 端市场，数字人正逐步被应用于直播、客服接待、企业培训等内容服务场景，在一定程度上辅助降低部分标准化内容生产和运营投入；在 C 端市场，虚拟形象产品也从早期以外观定制为主的“捏脸换装”模式，逐步向具备记忆、“人设”设定与互动反馈能力的产品方向演进。

与此同时，行业关注的重点也在发生变化。过去，数字人的竞争焦点主要集中在美术渲染精度与外观写实程度；当前，多模态实时交互的响应延迟控制、情感计算的自然度以及规模化生成能力，正成为观察数字人产品应用效果的重要参考维度。从虚拟空间中的数字形象，到可打印落地的实体艺术品，AIGC 技术正在推动数字人从单一的线上展示场景向更多元的应用形态延伸。

2. 典型实践案例

本节在案例选取上，从“线上内容服务与线下定制服务”的维度进行切分，选取两个产品服务，通过“一虚（线上直播）一实（线下文创）”“2D、3D”的双重视角，观察数字人产业在不同业务环节中的应用路径。

● 典型实践一：“硅语”数字人平台：以 AIGC 重构制作流程，探索数字人落地路径

在数字人应用落地的早期阶段，制作成本、交互能力和内容生产效率是影响数字人场景应用的重要因素。硅基智能以 AIGC 技术为核心驱动，围绕降低数字人制作门槛开展探索，推动数字人在场景化服务中的应用。

（1）以 AIGC 技术降低数字人制作门槛，提升内容生产效率

硅基智能的核心技术路径在于，将传统依赖人工建模、动作捕捉与后期渲染的数字人制作流程，转化为以 AIGC 为支撑的标准化生产流程。通过自研的 2D 数字人生成与驱动技术，平台支持基于真人照片或视频素材快速生成可交互的 2D 数字人形象。在模型能力层面，硅基智能于 2024 年推出多模态大模型 DUIX ONE，集成了多模态融合处理能力，可同时处理文本、语音、视觉等多种输入信息，并搭载情感捕捉技术，支持实时识别与响应多种复杂情绪，有助于提升数字人在表情与语态呈现上的自然度。

依托于此，平台可赋予数字人实时问答与对话能力，辅助完成标准化问答、内容播报和互动响应等任务。

（2）聚焦直播带货与本地生活等商业场景

在应用层面，硅基智能的数字人产品主要聚焦于低人力成本需求和内容产出频次高的直播带货、本地生活播报与企业知识培训等场景。

在直播带货场景中，可辅助开展长时段直播，并完成商品介绍等基础交互任务，一定程度上降低重复性工作投入；在本地生活与资讯播报场景中，数字人可快速将文字内容转化为视频播报形式，适用于信息发布、商家活动推广等对内容更新频次要求较高的应用场景；在企业培训场景中，数字人可作为标准化“培训讲师”，实现培训内容的统一化交付。

● 典型实践二：“趣丸万相”AI 3D 生成交付一体化平台：探索 AI 三维生成与全彩 3D 打印结合的个性化定制服务

趣丸万相是趣丸科技围绕 AI 三维生成与全彩打印场景开展的产品探索，平台主要面向人物、宠物、潮玩等个性化定制需求，提供全过程服务链路。

（1）围绕高精度写实人像等场景，完善“图像、扫描素材输入—三维模型生成—模型修调—全彩 3D 打印交付”全链服务

趣丸万相的技术优势主要体现在面向三维生成与全彩 3D 打印场景的模型和流程适配。区别于面向虚拟展示的泛化 3D 生成

模型，趣丸万相的算法围绕可打印、可交付的应用场景进行适配，重点探索攻克模型物理结构完整性与打印可行性难题。平台支持基于单张图片生成较高精细度的 3D 模型，用户可上传符合规范的照片素材，在平台完成 AI 建模操作。平台提供半身人像、全身模型、手办定制、萌宠定制、浮雕定制、扫描建模及图生 3D 等多类建模功能，覆盖个人写实形象、宠物形象与潮玩手办等不同应用需求。

在生产流程上，平台覆盖了从素材上传、AI 建模、模型修调到全彩打印与质检交付的服务链路。建模完成后，用户可在平台内对模型进行 360 度查看，并通过灯光切换、纹理展示、白模查看等选项核验模型质量；确认无误后可直接在平台内置打印中心提交打印。此外，平台同时支持 API 接口接入，支持从个体定制向较大规模生成能力的衔接。

（2）面向多元定制场景，探索 AI 三维生成服务应用

在 B 端合作生态方面，趣丸万相通过 AI 三维生成服务能力，向电商及定制商家、全彩打印机厂商、潮玩与文创厂商、扫描仪设备商等提供技术支持。例如：自动化建模能力帮助电商商家提升定制产品建模和生产环节效率；面向打印机厂商提供模型生态内容支持，丰富设备应用场景；支持潮玩与文创厂商实现 IP 三维化的快速打样，缩短产品上市周期；面向扫描仪设备商则通过建模-修模-全彩 3D 打印的流程，为 3D 真人手办产品提供持续产

能。目前，趣丸万相已与珠海赛纳三维科技有限公司建立战略合作关系，结合趣丸万相的三维生成能力及赛纳三维全彩 3D 打印设备，推出“从图到物”的一站式解决方案，业务覆盖日韩、美洲、澳洲、俄罗斯等海外市场。

综合来看，趣丸万相所探索的“AI 建模×打印交付×个性化定制服务”路径，将 AI 生成能力与实体制造能力相结合，在定制礼品、个人典藏、IP 衍生周边等细分消费场景中形成了差异化的应用实践。

五、 全球化传播：AI 驱动中国文化出海的新路径

（一）发展机遇：政策环境与技术支撑持续推动文化出海

当前，国家层面对文化出海和文化数字化的鼓励政策持续完善。一方面，“十五五”规划中提出“鼓励文化企业国际化经营，推动优质网络文学、网络游戏、影视动漫、精品展览等出海，加强国际传播重点基地、国家文化出口基地建设”。同时，在推进数智技术赋能方面，“全面实施‘人工智能+’行动，加强人工智能同文化建设等领域相结合，抢占人工智能产业应用制高点，全方面赋能千行百业。”

另一方面，AI 技术为文娱内容的生产、译制、本地化和传播分发提供了新的工具支撑。AI 不再只是提升单一环节效率的生产工具，而是正在成为提升内容本地化传播效率和传播适配能力的重要工具。AI 多模态技术在这一转变中扮演了重要角色。

多语种语音合成、视觉内容自动生成等能力的发展，有助于推动中国文化出海。AI 大模型不仅改变了“内容如何生产”，也为内容触达海外用户提供了新的工具选择。在这一背景下，一些将 AI 技术能力与文化出海需求深度融合的企业，已在部分内容生产和本地化服务场景中开展应用探索。

（二）关键路径：多模态技术提升内容本地化生产效率

1. 跨语种转换辅助：AI 多语种配音与表达适配

在文化出海的全链路中，语言本地化始终是重要环节。传统 TTS 技术受限于早期算法框架，能力较为局限；新一代基于 AI 大模型的相关技术实现了升级，主要体现在两个维度：其一是音色复刻能力，模型能够基于一定长度的授权原声样本提取说话人的声纹特征，并在目标语种中保留这一音色特征，使译制后的配音在听感上保留一定声音特征；其二是情感迁移能力，模型不仅翻译内容本身，还能识别原声中的情绪状态，并将对应的情感韵律映射到目标语种的语音输出中，改善跨语种表达中的情感呈现。

2. 视觉与听觉的自适应：动态对口型与音画同步演进

如果说语音克隆解决了“听觉层”的本地化问题，那么唇同步技术（Lip Sync）则是改善视频内容出海音画适配的重要技术方向。长期以来，影视与短剧内容的跨语种出海面临一个常见适配环节：不同语言在发音时，口腔、唇形与面部肌肉的运动规律存在显著差异。例如，中文的发音系统与英语、日语、阿拉伯语

的唇形运动模式差异悬殊，当译制配音与原始视频画面叠加时，演员嘴部动作与实际听到的语音内容往往明显不匹配，可能影响部分观看体验和-content表达效果。

AI 驱动的动态唇同步技术，可从技术流程上改善这一适配环节。其核心流程可分为三个层次：首先，通过面部关键特征点追踪技术，提取视频中演员面部的嘴唇轮廓、牙齿位置与面部肌肉运动数据，建立针对特定演员的面部运动模型；其次，基于目标语言的音素序列与对应唇形运动规律，生成与新配音语音在时序上匹配的目标唇形序列；最后，通过生成式模型将改善后的唇部动画自然融合回原始视频画面，提升面部整体动态的连贯性。经过这一流程处理后，海外观众看到的画面，在目标语言配音下具有较好的音画一致性，从而改善本地化观看体验。

这一技术与 AI 多语种配音能力，构成了当前 AI 辅助内容本地化生产的一种代表性技术组合：语音层改善多语种表达，视觉层改善唇形适配，两者协同作用，有助于提升译制效率，并改善多语种配音的自然度和音画一致性。

（三）典型实践案例

从语音情感复刻到唇同步技术完善，相关技术能力正在为中国文娱内容的多语种生产和海外传播提供支撑。以下选取两家企业，基于产品矩阵布局等维度，呈现 AI 技术赋能文化出海的实践情况。

1. 典型实践一：趣丸科技：AI 语音与 AI 音乐工具助力内容本地化传播

（1）AI 语音：以 B 端服务为基础，切入短剧出海场景

趣丸千音 AI 语音平台的海外拓展，主要依托其在短剧配音与视频译制场景中积累的 B 端服务能力。传统人工译制流程周期较长、成本较高，制约了短剧内容的规模化出海。针对这一需求，趣丸千音通过视频翻译功能切入短剧译制场景。在影视动漫与短剧出海领域，趣丸千音视频翻译功能可实现 24 小时译制 30000 分钟剧集，效率较人工提升 10 倍以上，成本下降 20 倍以上。趣丸千音还支持中文、英语、日语等 48 个语种生成，具备一定的多语种服务基础。此外，在新闻传媒领域，助力多家媒体机构将国际版视频一键生成多语种版本，同步分发至海外平台。

（2）AI 音乐：以社群运营与用户共创为核心，向全球拓展

Tunee 在产品设计阶段就将全球化作为基本定位，并面向专业创作者群体展开海外运营。依托内容生态与社区运营，构建起一套低成本、高转化的全球市场发展路径。Tunee 在不同地区采取差异化的运营策略，以适配当地用户的内容消费习惯。例如，在北美市场，Tunee 侧重真人题材内容的生成体验；在欧洲市场，则增加艺术电影感、增强实验性的音乐风格支持。在用户运营机制上，主要通过社区维护海外核心用户，同时对优质创作者进行邮件触达与深度访谈。Tunee 产品迭代中约 70% 的需求来源于用

户反馈。此外，设有“创作者计划”，支持核心市场用户通过发布创作内容获取平台积分，形成一定的内容传播机制。

2. 典型实践二：昆仑万维：以大模型与文娱应用矩阵协同布局，探索出海实践

据公开披露信息，2025 年，昆仑万维实现海外业务收入达 77.23 亿元，同比增长 49.91%，为“AI+大文娱”出海探索提供了应用实践。

昆仑万维围绕 AI 大模型和文娱应用开展全球化布局，将相关技术能力应用于短剧、音乐创作、AI 社交等产品方向。在内容消费赛道，昆仑万维以 AI 短剧平台 DramaWave 为核心出海支点，依托自研 SkyReels 视频生成大模型支撑内容的 AI 化生产，并辅以 FreeReels 免费平台扩大产品覆盖。截至 2025 年底，双平台合计 MAU（月活跃用户人数）在海外短剧市场排名靠前，单月流水接近 3600 万美元。在创作者经济赛道，AI 音乐创作平台 Mureka 面向全球用户提供从词曲生成到分轨编辑的全流程服务，其底层模型的部分维度具备行业影响力。

六、 挑战与治理：“AI+大文娱”的风险识别与规范发展

（一）风险透视：“AI+大文娱”发展面临的四重挑战

AI 技术在大文娱领域的应用推动生产效率提升的同时，也衍生出一些新风险。

1. 内容安全：价值观偏离与“深度伪造”

大模型训练引入质量参差不齐的语料，可能导致生成内容存在偏激观点与有害信息。以 AI 换脸等为代表的“伪造”技术已出现侵犯名誉权及用于电信欺诈的实际案例。当前已有《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》明确合规底线，但在执行层面，技术迭代与监管响应之间可能存在一定时滞，内容安全的动态防控能力仍有待加强。

2. 版权保护：训练语料侵权争议与生成物确权难题

AIGC 对版权的影响集中于两个层面：在训练端，大模型预训练或微调需大规模抓取部分受版权保护的文娱内容。训练语料使用的授权来源、权利边界及合理使用原则等需进一步明确。在生成端，司法实践虽已探索回应 AI 生成内容在著作权等方面的权属问题，但有关裁判依据仍有待细化完善。一些大文娱市场主体对 AI 生成内容商业化的合规路径持观望态度。

3. 个人信息保护与数据合规：采集边界和授权机制

个性化体验依赖对用户数据的持续采集与分析，一些平台在采集范围与使用告知方面的透明度需进一步提升。《中华人民共和国个人信息保护法》《个人信息保护合规审计管理办法》等对有关方面已有明确规制，但在“AI+大文娱”场景中，高风险数据的采集边界与授权同意机制仍需进一步细化。此外，用户在情感陪伴类交互中主动输入的私密信息，其合规处理亦需纳入重点

关注范畴。

4. 算法治理：算法歧视与人机协同边界

训练数据的结构性偏差可能导致模型输出存在表达偏差。当训练语料集聚于某一地域、语言时，算法歧视便在所难免。此外，大文娱创作的人机协同边界需要审慎把握，在提升效率的同时，需保留创作者在创意判断与质量把关中的主体作用。这也让“负责任 AI”成为大文娱行业创新不可忽视的新要求。

（二）多方共治：产业规范发展建议

在风险识别的基础上，推动“AI+大文娱”形成创新治理路径，需要多方协同、多措并举。

1. 加强监管，包容审慎推动行业创新

目前《人工智能生成合成内容标识办法》已确立“显式标识+隐式标识”双重体系。日后，还需进一步完善 AI 生成合成内容的全链路追溯机制。同时，鼓励“AI+大文娱”探索“沙盒监管”机制试点，在可控范围内给予一定的试验空间，以包容审慎的态度，推动行业技术迭代和监管规则完善。

2. “以技治技”，夯实双重防线

逐步完善大模型内容安全评测基准，明确合规性测试的核心指标与评估方法，助力守好内容安全“底线”。鼓励“以技治技”，加大深度伪造检测与识别技术的研发投入，提升大文娱内容“去伪存真”的能力，夯实技术安全“底座”。

3. 多元共治，以“人工智能+”创新产业生态

持续加强多元共治，推动政府、高校、文娱行业企业加大协同合作。企业端需将“负责任 AI”理念前置，在合规边界下推动内容和技术创新；行业力量需在版权保护、数据合规等共性问题上引领行业自律；产业端和高校科研机构还可对标“AI+大文娱”发展需求，联合加快培养复合型人才，促进人机协同在行业深化落地，把握“人工智能+”浪潮，建设欣欣向荣的大文娱产业生态。

七、趋势观察：“AI+大文娱”未来发展新风向

人工智能技术演进对大文娱产业的影响已从工具层面向多层次延伸。基于公开探索实践，本章节从五个维度梳理“AI+大文娱”行业发展的新趋势。

（一）生产范式演进：高水平人机协同赋能高效率内容生产

当前，人工智能技术在大文娱产业中已广泛应用。未来，将涌现出更多 AI 深度参与行业创新实践的探索。从经济成本看，AI 原生内容初始模型与创意框架一旦建立，后续个性化生成所需的增量投入相对有限，边际生产成本较传统标准化内容生产较低，这将在一定程度上对内容生产方式与成本核算逻辑产生影响。放眼未来，内容个性化程度与交互深度将继续提升，用户逐步参与内容实时生成，消费体验的沉浸性与专属感将更受关注。然而，创意方向设定与价值导向审核，仍需要专业人员的深度参与，这

离不开更高水平的人机协同。

（二）产业空间拓展：“AI+大文娱”出海促进数字 IP 跨文化传播

未来，以“新三样”为代表的中国文化出海热度仍将不减。从技术层面看，大模型在多语种翻译、本地化配音生成与文化表达转化方面的能力持续提升，有助于压缩海外本地化传播的时间与资金成本，语言差异对出海效率的制约有所减弱。政策层面也为大文娱行业开拓海外市场提供了有力支撑。技术能力提升与政策支持形成合力，共同推动优质文娱内容出海向多语种、多市场拓展。在 IP 运营层面，企业可基于同一 IP 核心设定，通过 AI 生成工具针对不同市场输出差异化版本，在维持 IP 一致性的同时适应本地化表达偏好。这种“一个 IP、多向生长”的运营模式，可在一定程度上降低内容适配成本。但出海内容的文化适应性终究无法完全依赖技术手段，对传播目标人文逻辑的深度理解，仍是跨文化传播取得较好反响的基础。

（三）媒介形态演进：“AIGC+空间计算”推动沉浸式文娱场景拓展

现阶段，AI 文娱应用的主要载体仍集中于二维屏幕。随着技术的发展，大模型驱动的 3D 资产实时生成能力，正在为沉浸式文娱内容生产提供新的工具选择。从“AIGC+空间计算”技术成熟度与硬件普及进程来看，相关场景的大规模商业落地仍需一

定时间，但相关探索已对文娱体验形态形成新启发。然而，媒介形态演进同样带来新的合规要求，沉浸式体验的高感知强度，对内容安全、未成年人保护等方面的平台责任、动态监管等提出了更高要求。

（四）价值闭环探索：向“个性体验服务”与“数据资产运营”延伸

传统文娱的商业逻辑大多指向对内容产品本身的价值化。随着 AI 技术的发展与应用，逐步推动这一逻辑向两个方向延伸：一是面向 C 端的“个性化体验”服务模式。当内容可以根据用户偏好与实时行为动态生成时，用户购买的已是一段专属化的交互体验（如互动短剧、虚拟陪伴形象等），其对应的服务形式，将从一次性购买或流量广告，逐步向基于持续服务的订阅制收费演进。二是面向 B 端的“能力与资产”输出模式。对于具备 AI 能力积累的文娱企业而言，其核心资产将延伸至高质量的专有语料库、经过专业调优的垂类模型以及虚拟 IP 的授权体系。以“模型即服务”为形态的 B 端业务已在部分企业服务场景中开展探索，为行业提供基于 AI 能力输出和数据资产化的服务路径。

（五）人才迭代升级：新职业逐步形成

AI 工具的普及正在改变大文娱产业的人力组织逻辑，并对从业者能力提出新要求。个体创作能力的上限得以抬升。借助 AI 工具，具备清晰创意框架与审美判断力的个人或小型团队，

能在视效、音乐、叙事等多环节调用 AI 辅助，完成过去需要较大规模团队才能承接的创作任务，这种“一人公司”（One Person Company，简称 OPC）创业模式将为“AI+大文娱”带来更多新活力。**新职业需求正在形成。**随着 AI 深度介入文娱生产流程，生成式人工智能系统应用员、人工智能训练师、虚拟现实产品设计师和提示词工程师等新兴岗位正逐步形成。从人才培养机制看，现有的教育课程设置与发展需求之间仍存在一定差距。AI 工具的演进为从业者打开新发展空间的同时，也对其持续学习与适应变化的能力提出了更高要求。



人民数据管理（北京）有限公司

中国信息通信研究院云计算与数字化研究所

广州趣丸网络科技有限公司