

实时互动产业发展 研究报告

——互联网的下一个十年

(2022 年)

中国信息通信研究院泰尔终端实验室

2022 年 6 月

版权声明

本报告版权属于中国信息通信研究院，并受法律保护。
转载、摘编或利用其它方式使用本报告文字或者观点的，
应注明“来源：中国信息通信研究院”。违反上述声明者，
编者将追究其相关法律责任。

编制说明

本报告由中国信息通信研究院牵头撰写，限于撰写时间、知识局限等因素，内容恐有疏漏，烦请各位读者不吝指正。

本报告在撰写过程中得到了多家单位的大力支持，在此特别感谢感谢参编单位：北京百度网讯科技有限公司、北京火山引擎科技有限公司、北京金山云网络技术有限公司、北京易掌云峰科技有限公司（环信）、海马云（天津）信息技术有限公司、甲骨文（中国）软件系统有限公司、科大讯飞股份有限公司、摩尔线程智能科技（北京）有限责任公司、上海兆言网络科技有限公司（声网）、深圳华为云计算技术有限公司、深圳凡泰极客科技有限责任公司、腾讯云计算（北京）有限责任公司、英特尔（中国）有限公司、优刻得科技股份有限公司、元境生生（北京）科技有限公司、中国移动咪咕文化科技有限公司、中移（苏州）软件技术有限公司、中移物联网有限公司、珠海金山办公软件有限公司（按照拼音首字母排序，排名不分先后）的各位专家。

前 言

十四五规划中提出“加快发展现代产业体系，巩固壮大实体经济根基”、“加快数字化发展，建设数字中国”等部署以来，数字经济已成为产业转型升级的重要驱动力及改变全球竞争格局的关键力量。如今，数字化已从民众生活全面渗入至产业端各链条，实体经济的数字化进程被按下“快进键”，并且随着 5G、云计算、大数据、人工智能等新一代信息技术的日渐成熟，数字技术与实体经济集成融合的趋势更加凸显。实时互动技术作为新一代数字经济发展的底座支撑，将大幅提升行业全要素生产率，加速新型数字产业体系的形成与发展，开辟经济增长新空间。

在国际经济加速博弈、全球产业转型升级的当下，国内外均在积极探索下一代互联网的方向，争取领先地位。中国在下一代互联网关键技术及产业上的领先突破，将为中国经济发展和产业转型提供重要基础，为后续应对复杂竞争环境提供持续推动力。

实时互动与互联网行业、传统行业的融合不断深化，激发出更多的新场景与新应用。面向中国互联网的下一个十年，积极探索实时互动产业核心要素，明确行业发展难点痛点，提供行业发展参考路径成为关键所在。

本报告从实时互动产业发展的生态环境及历史必然性出发，首次提出了实时互动的定义，明确了其核心内涵与特征，并与 CPaaS、RTC、音视频云、元宇宙等相关概念进行了辨析。同时，根据产业现状，梳理了实时互动核心价值链与产业图谱，对市场规模进行了

预测分析，对未来趋势进行了深度研判。其次，从专有基础设施、专业组件与重要共生技术三个维度对实时互动所涉及的关键技术及内容框架进行了重点剖析。结合不同行业的特征与痛点，整理了实时互动与各行业场景的融合应用情况。最后，对整个产业面临的挑战进行了客观分析，提出了推进策略及发展建议。希望通过本报告，能够向大众普及实时互动的概念及作用，为社会各界提供借鉴和参考，有序推进产业的建设发展。

目 录

一、 实时互动发展背景.....	1
(一) 实时互动发展宏观生态环境	1
(二) 实时互动发展的历史必然性	2
二、 实时互动的概念和内涵.....	5
(一) 实时互动的概念	5
(二) 实时互动发展历程	6
(三) 实时互动的内涵	9
(四) 实时互动与相关概念辨析	11
三、 实时互动产业发展态势.....	16
(一) 实时互动产业图谱	16
(二) 实时互动产业发展现状与趋势	22
(三) 产业规模预测	27
四、 实时互动关键技术.....	30
(一) 专有基础设施	31
(二) 专业组件	36
(三) 重要共生技术	41
五、 实时互动赋能千行百业.....	42
(一) 实时互动与企业协作：优化流程，提升效率	44
(二) 实时互动与教育：促进公平，提升质量	45
(三) 实时互动与 IoT：高效连接，实时协同	46
(四) 实时互动与制造：远程作业，保障安全	47
(五) 实时互动与交通：远程处理，提高效率	48
(六) 实时互动与金融：远程服务，便捷高效	48
(七) 实时互动与房产：远程看房，促进交易	49
(八) 实时互动与医疗：普及资源，实时介入	49
(九) 实时互动与政务：减少繁琐，加速办理	50
(十) 实时互动与文旅：远程游览，沉浸体验	51
(十一) 实时互动与泛娱乐：创新玩法，改变潮流	52

(十二) 实时互动与游戏：丰富体验，随时可玩	53
(十三) 实时互动与社会关怀：关爱弱势，提升福祉	55
六、 产业发展挑战和建议	56
(一) 产业发展面临挑战	56
(二) 产业发展建议	58
附录 1：中国互联网发展简史	60
附录 2：缩略语	61
参考文献	63

图 目 录

图 1 实时互动技术集合..... 6

图 2 实时互动发展历程..... 6

图 3 实时互动相关产业图谱..... 16

图 4 实时互动渗透率..... 22

图 5 全球实时互动热门场景及新兴场景..... 23

图 6 实时互动产业微笑曲线..... 27

图 7 全球及中国实时互动产业规模统计预测..... 29

图 8 实时互动关键技术..... 31

表 目 录

表 1 互联网发展范式的三个阶段..... 3

表 2 实时互动相关技术概念对比..... 13

表 3 实时互动产业规模统计口径（To B 技术层） 28

表 4 实时互动应用新场景（示例）..... 44

一、实时互动发展背景

（一）实时互动发展宏观生态环境

外部竞争局势逐步加剧，科技引领趋势日益凸显。近年来，世界格局重塑、新冠疫情蔓延，贸易关系复杂，国际形势中的不确定性因素逐渐增多，全球经济面临着前所未有的严峻考验。科技创新是增强综合国力、提升国家地位的决定性因素，加大科技引领已成为众多国家的核心发展战略。

数字经济与实体经济深度融合，技术创新推动产业变革升级。数字经济已成为推动经济社会高质量发展的重要引擎和重要支撑，国家及地方相继推出有关产业数字化、数字产业化、建设数字中国等发展战略的利好政策。2022 年 1 月国务院印发的《“十四五”数字经济发展规划》，明确了社会发展的主要目标，强调建设高速泛在、天地一体、云网融合、智能敏捷、绿色低碳、安全可控的智能化综合性数字信息基础设施。同时，实体经济与数字经济的融合，也存在巨大的协同发展空间。根据中国信息通信研究院数据，2020 年中国数字经济规模达 39.2 万亿元，占 GDP 比重 38.6%，整体发展进入快车道。在具体行业上，《中国数字经济发展白皮书（2021）》数据显示，服务业、工业、农业数字经济占行业增加值比重分别为 40.7%、21%和 8.9%，传统产业数字化仍有很大的提升空间。

疫情常态化加速了线下活动的线上化，改变了人们的生产生活方式。在生活上，对于互联网的需求不再是单向的娱乐消费内容获取，原生于互联网的在线虚拟陪伴社交已开始替代线下生活的社交。

在工作上，协同办公工具的出现提升了沟通效率，改变了传统的团队协作方式，弥补了疫情下难以面对面交流的不足。同时，在线教育打破了传统线下教育的时空限制，使教育在技术、内容、形式及主体方面得到全方位的升级，尤其是在疫情形势十分严峻的当下，在线教育有效保障了学生学习的开展以及“停课不停学”政策的落实。受疫情影响，众多场景已实现利用数字化技术完成线下活动的线上化转移，线上线下融合发展已成为大势所趋。

云原生技术和分布式云不断成熟，成为业务创新发展的重要驱动力。云计算行业发展进入成熟期，企业上云的加速及云服务在传统行业的实践应用，推动了云原生技术体系与分布式云架构的不断完善。云原生底层核心技术已相对成熟，中间件、服务网格、无服务器等更上层的技术得到了快速发展，研发门槛大大降低，云端开发部署应用成为行业趋势。同时，边缘计算与边缘侧的算力需求，加速了云计算从中心向边缘延伸，分布式云逐渐成为云计算发展的新形态。

(二) 实时互动发展的历史必然性

通信技术、终端设备、开发框架等技术的升级，推动着互联网产业的发展，并以不同的应用形态为用户所感知。我国互联网从门户网站为主导的 PC 互联网，到智能手机激发的移动互联网，再至万物互联的产业互联网，深刻改变人们的生活和经济的运行。近年来在国内外已经掀起了一波对下一代互联网发展范式的讨论。根据文件传递类型、文件传递方式、网络组织形态、内容创建、商业模式

等进行阶段划分，可以将互联网范式分为 Web1.0、Web2.0 到 Web3.0 三个阶段，如表 1 所示。

表 1 互联网发展范式的三个阶段

发展阶段	Web 1.0	Web 2.0	Web 3.0
时间阶段	1989-2004	2004-2016	2016 ~ 至今
核心特征	可读	可写	可信
用户规模	百万级用户 以公司为主	十亿级用户 以社群为主	千亿级用户 以个体为主
展示形式	以文字、图片为主的 静态内容	以音视频流媒体为主的 动态内容	融合 AI、3D、实时音 视频的多维内容
内容拥有	占有内容	分享内容	共建内容
核心指标	用户浏览量	用户点击量	用户参与度
信息传递方式	单向传递	双向传递	多用户多向传递

来源：公开资料整理

从生产力角度而言，Web1.0 到 Web3.0 的迭代，使互联网用户的创作能力得到了更大化地释放。Web 1.0 以中心化的门户网站为主，为用户提供文字、图片等内容，信息为单向传递。Web 2.0 以中心化平台为主，用户在获取音视频等内容的同时，可以主动创作、分享内容，信息的传递方式变为双向。随着技术的演进，在互联网发展的新阶段，用户规模将持续扩大，融合 AI、3D 及实时音视频的多维内容可以得到充分展示，用户的参与度不断增强，多用户之间的多方向的事实信息传递成为可能。

关于 Web 3.0 的研究与探索，主要围绕开放、共建、隐私三个特征展开。首先，Web3.0 世界充分自由开放，用户行为不受中心化平台间的隔离与限制；其次，内容和应用将由用户创造和主导，可充分实现用户的共建、共治，实现平台价值的共享；最后，用户可

完全拥有属于自身的所有数据及隐私信息，相关数据信息能够通过加密算法和分布式存储等手段得到充分保护，从而不受平台的干涉。Web 3.0 充满想象力，但难以作为有指导性的概念代表下一代互联网发展范式的全貌。目前 Web3.0 缺少标准定义，对应的核心技术架构与应用落地形态尚不明晰，难以提供全面的技术路径指导。同时，Web 3.0 更强调线上内容的生产消费，难以与线下生活建立联系。

实时互动强调生产力的创新迭代，并且实现了从“在线”至“在场”的重要转变，推动互联网向以“临场感”为主要特征的下一代互联网演进方向升级变革。其所涉及的传感技术、边缘计算、感知交互、全息影像、仿真引擎、空间声场、AIGC 等技术会转化为新一代生产力，进一步改变生产和生活方式。目前实时互动概念已被国内外云厂商、互联网平台、实时音视频服务企业等逐步认可，实时互动与各行各业的深度融合也促使其需求呈指数级增长。未来，实时互动将逐步变成一种普适能力。

实时互动作为下一个十年引领互联网增长的技术，具有如下优势：

- (1) 可全面覆盖用户的生产生活需求，具有庞大的用户基础和不可替代性；
- (2) 可支撑从消费互联网到工业互联网的多种场景升级，是推进产业发展智能化、数字化的关键途径；
- (3) 可通过便于落地的场景化方案，从可操作、可测量的技术角度为互联网基础设施改造指明方向；

(4) 可融合多项关键技术，有目标性地激发多种国家关键技术的自主发展。

二、实时互动的概念和内涵

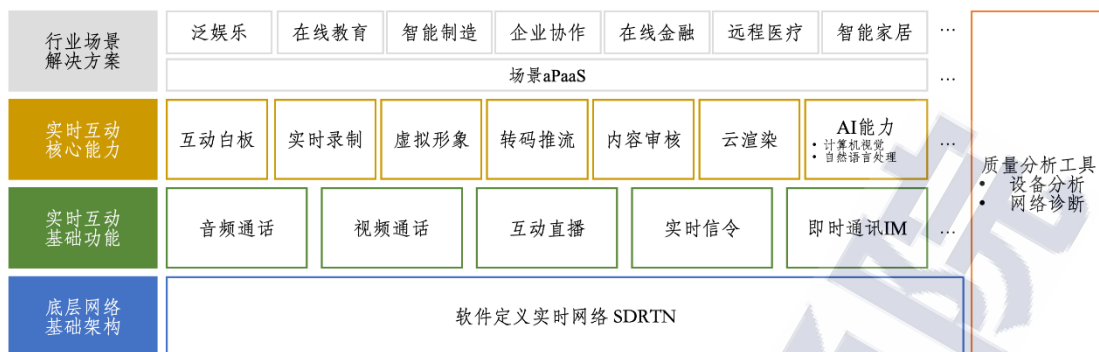
(一) 实时互动的概念

实时互动，指在远程条件下沟通、协作的多方能够随时随地接入、实时地传递虚实融合的多维信息，并体验身临其境的交互活动。

从广义角度而言，实时互动包含了特定场景下人、物、空间三者交互体验的所有内容，包括保障实时无延迟的通信网络、满足用户互动的技术组件及设备、提供可感知的实时互动场景应用等。

从技术角度而言，实时互动包含了能够实现实时通信和灵活互动能力的一系列技术的集合。实时互动最基础的功能为实时通信，即将用户在线下产生的音频、视频、文本、图片等媒体和非媒体数据进行实时传输。实时互动最核心的能力为灵活互动，即在信息传输的基础上，根据具体场景需求，借助多样化的插件组件和算法，灵活增加互动工具，提升效率与用户体验。

完整的实时互动技术集合，使行业应用方可基于底层实时网络，在实时通信的基础功能上，结合场景所需的核心互动能力，为用户提供场景解决方案。同时，相关应用方可借助质量分析工具，完成端到端的质量监控，提升用户实时互动体验。

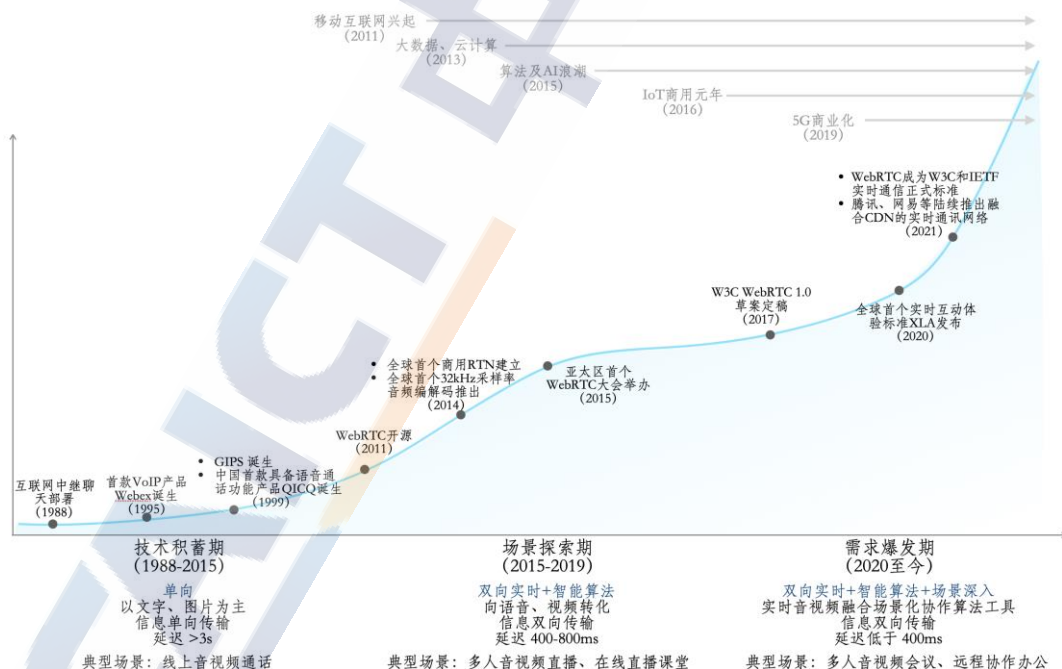


来源：中国信息通信研究院

图 1 实时互动技术集合

(二) 实时互动发展历程

实时互动是建立在长期技术发展基础上的可靠创新。随着 5G、大数据、云计算、AI、区块链、IoT 等技术的发展和成熟，实时互动的边界不断拓展。结合技术发展及商业化进程，可将实时互动产业的发展分为三个阶段。



来源：中国信息通信研究院

图 2 实时互动发展历程

第一阶段：技术积蓄期（1988-2015）

1988 年，互联网中继聊天（IRC）首次部署，拉开实时聊天和即时消息程序的先河。1995 年，首款基于 VoIP 的通信产品诞生，实现了基于 IP 传输的语音通话和多媒体会议。1999 年，GIPS 利用 PC 计算能力对声音算法进行创新，为互联网通信提供新引擎；同年，国内首个具备音视频功能的软件 OICQ 诞生。2003 年后，GIPS 陆续向 QQ、Skype、Webex 等产品提供音频引擎，为更多互联网通信产品提供服务。2011 年 GIPS 更名为 WebRTC 并正式开源，为实现视频流、音频流或者其他任意数据在浏览器之间的点对点连接传输提供了技术支撑。

2014 年，全球首个商用实时音视频传输网络 RTN 开始搭建，开创性地使用边缘网络节点和智能路由技术，实现快速高效连接，提供高质量、低延时、稳定可靠的音视频传输，为 WebRTC 的大规模应用提供了底层网络保障。

在技术积蓄期，实时互动的主要形式为即时通信和音视频通话，其中即时通信以文字和图片为主，音视频通信由于几秒甚至几十秒的延迟，且受限于网络条件，整体商业化水平较弱。

第二阶段：场景探索期（2015-2019）

2015 年，亚太区首届 WebRTC 大会的召开，拉开了实时互动场景大规模商业化探索的序幕。移动互联网浪潮下，智能手机的普及和 4G 商业化使得通信效率提升、资费下降，互联网用户规模不断扩大，内容消费形式从文字向语音、视频转变，对实时互动的技术与形式提出了新需求。

2016 年前后，具有互动直播能力的商业化服务与 SDK 陆续推出，低延迟、高清晰、大规模实时音视频传输成为可能，移动视频直播行业迎来爆发期。社交泛娱乐行业依托其庞大的用户基数与多样化的用户需求，积极利用互动直播能力探索新场景、新玩法，为用户创造无可替代的新体验。

实时互动的场景实践也推动了 RTC 技术的发展与成熟。在场景应用中，为保证用户体验，需要针对大规模、高并发、抗弱网、低能耗、小包体等各类具体问题优化与适配，因此，实时通信服务商的技术壁垒越来越高，网络效应愈发凸显。随着专业化的实时音视频服务提供商出现，垂直供应商能够通过其提供的 SDK 实现低代码接入，从而降低调用门槛，更专注于服务模式本身。长此以往，更多行业应用者加入实时互动的赛道，基于自身行业认知，进行场景创新。

第三阶段，需求爆发期（2020 年至今）

2020 年疫情期间，社交、办公、娱乐等大量线下活动被迫迁移至线上，视频会议、教育、电商直播、娱乐直播、教育信息化等实时音视频需求激增，机遇和挑战并存。

大规模爆发的用户规模与使用场景，对低延时的实时互动基础网络提出巨大挑战，从而推动了云网端的升级与优化。此外，AI、插件组件、云渲染等技术的融合，提供了多样化的交互形式，拓展了实时互动的边界。面对需求激增、类型丰富的创新场景，让实时互动从“可用”变为“好用”，各实时互动服务提供商从用户体验质

量标准、超低延时直播标准等多维度进行探索。经过多年商业模式的验证探索及技术标准的逐步完善，2021 年万维网联盟（W3C）和互联网工程任务组（IETF）宣布将 WebRTC 纳入官方标准，成为线上通信及协作服务的基石。自此，开发人员可在任意网页、应用程序内置实时影音通信系统，使用 WebRTC 构建实时互动应用，这代表着下一代互联网发展的重要方向。

（三）实时互动的内涵

1. 技术实现维度

从技术架构角度而言，实时互动是未来数字化生活的新一代基础设施。实时互动对原有互联网技术架构提出了更高的要求，其特性主要体现在“实时”和“互动”两个方面。

（1）实时

“实时”是对信息传递效率的变革，使用户随时随地即时获取无损信息。具体体现在：

接入的实时性：指用户接入的高登录成功率，保证用户不受连接设备、网络情况、地域环境的限制，随时随地接入网络、触达业务。

传输的实时性：指音视频数据传输的低延时和低卡顿，通过网络实时动态路由规划，以最优边缘节点接入，降低全球端到端的延时，保证传输流畅进行。

交互的实时性：指除基础音视频外，图像、声音、音效等多种交互方式反馈的实时性，需要 AI 算法、边缘计算等技术的综合处理。

渲染的实时性：指最终画面尤其是 3D 画面、多维场景展示的实时性，需要大量的场景计算能力，使得户获得逼真流畅感受。

（2）互动

“互动”是用户交互模式的变革，用户可以从多维度接触信息。具体体现在：

互动网络的升级：网络基础设施的升级、边缘计算能力的提升、SDN 的完善，以及 3GPP 标准、4A 技术规范的成熟，解决了在 1 对多、多对多场景下常遇的大规模、高并发网络问题。

互动设备的突破：低功耗微型处理器的迭代、物联网传输协议的成熟标准化、AI 技术的融合与加速、传感器技术的迭代，推动了 IoT 设备的快速发展，丰富了人与物、物与物的交互模式。智能手表、智能门锁等 IoT 设备提升了人们的生活质量，AR/VR 等设备为用户提供了全新沉浸式体验。

互动组件的丰富：低代码平台的兴起、以 API、SDK 形式接入的插件组件的大量出现，降低了研发的门槛，提高了场景构建的灵活度与丰富度。如诞生于教育场景的互动白板，同时可广泛应用于会议、泛娱乐场景，为用户提供更丰富的信息交换维度。

互动场景的拓展：AI、IoT 等技术的演进、线下业务的线上化，拓展了实时互动的应用新场景。如疫情下兴起的云会展、云旅游，将用户的线下游览体验迁移至线上，融合中加入的虚拟人，又为用户创造了全新的交互体验。

2.用户体验维度

从用户体验角度而言，实时互动提供了越来越接近无损的信息传输，以及提供真假难辨的数字仿真能力。实时互动真正实现了人体感官的复制和延伸，让用户在足不出户的情况下，以远程的虚拟的化身亲临现场，与置身同一数字空间的其他人进行沟通和协作。与传统音视频通信相比，实时互动的核心特性为：

共享情境：共享人与人的交流内容，还原交流时的感官、道具及环境，完整展示人、物、环境之间的状态与关系；

任意规模和形式：支持不同人数规模的实时交流并可实现场景快速拓展，可还原在同一时刻多向信息交叉互动的复杂情形；

无处不在：实时互动可跨平台、跨终端进行，并且不受网络环境限制完成快速切换，用户可随时随地获得流畅体验。

(四) 实时互动与相关概念辨析

1. 实时互动和 CPaaS、RTC、音视频云的关系和区别

实时互动最基础功能是对信息的实时传输，与此相关的概念包括 CPaaS、RTC 和音视频云，但在具体的定义上，三者各不相同。

CPaaS 指通信平台即服务，是以用户通信为核心的云服务，强调通过 API 和 SDK，将语音、SMS、数字消息、视频等通信功能嵌入到面向客户和合作伙伴的应用中。常见应用场景为：即时消息、音视频通信、多因素身份认证等。

RTC 指实时通信，广义包括各种数据的实时传输，狭义更强调实时音视频的传输。RTC 与 CPaaS 的本质都是对信息进行转发，但 RTC 对网络延迟的要求更高，用户在正常通信过程中基本感受不到

延迟存在。常见应用场景为：网络电话 VoIP、视频电话会议、远程在线/网真。

音视频云是指为满足视频与音频制作、存储、处理、分发以及分析、审核、检索、推荐、理解等特定需求而定制的云解决方案。具体而言，音视频云包括视频 CDN，以及在视频云基础设施上部署的云平台、应用解决方案。常见应用场景包括长短视频点播、在线视频直播、智能媒体处理、云渲染等。在音视频云中，只有注重互动、对延迟低的内容需要通过实时音视频进行传输，大部分通过传统的 CDN 即可实现分发。

CPaaS、RTC、音视频云已相对成熟，但无法完整描述用户对场景真实性、信息多样化和交互多元化方面日益增长的需求：

用户对场景真实性的需求不断提升，需要将道具、物品、介质真实状态进行同步展示。比如教育和协同办公场景下的白板、金融场景下的屏幕双录、社交场景下的 K 歌房等，不仅需要传递音视频信息，更需要增加人与物、人与空间的多种交互关系，实时动态地传递同一场景下的多维信息，因此大幅提高了场景的复杂度。

用户对信息多样化的需求不断增加，需要包含大量非语义信息交换来丰富信息交换的维度。信息交换是通信的基础，随着线下场景的线上化，更多“陪伴”、“现场体验”类场景不断涌现，如在线自习室、虚拟演唱会等，该类场景的核心价值在于共享虚拟空间下的陪伴，而非传统的信息交换。此时，用户体验的“真实性和还原性”比“通畅”更为重要。

用户对交互多元化的需求不断加，互动形式需要有拓展性。首先，需要满足用户互动规模的多样化，实现从一对一到多对多规模的快速拓展。其次，需要涵盖人和人、人和物的多样化的互动形式；最后需要考虑场景内，由空间位置差异引起的互动强弱差异。

具体而言，实时互动与 CPaaS、RTC、音视频云在信息形式、沟通方向、网络延时及核心指标上具有明显区别，决定了典型应用场景的差异性。

表 2 实时互动相关技术概念对比

概念	信息形式	沟通方向	速度延时	核心指标	典型应用场景
CPaaS	静态简单信息（文字、图片）	双向	异步信息，无低延时要求	送达率	即时消息、音视频通信、多因素身份认证等
RTC	实时动态复杂信息（实时音频、实时视频）	双向	同步信息，延时百毫秒级	网络延时和接通率	网络电话 VoIP、视频会议、远程在线/网真等
音视频云	动态复杂信息（动态图片、视频流）	双向	异步信息，延时秒级	接通率与送达率	长短视频点播、在线视频直播、智能媒体处理、云渲染等
实时互动	实时音频复杂信息+互动（实时音频、实时视频、实时信令）	多向	同步信息，延迟几十毫秒级	接通率、网络延时、卡顿率、送达率	互动小班课、实时合唱、实时语聊、企业协作、虚拟会展等

来源：中国信息通信研究院

CPaaS、RTC 围绕着以人为核心的通信展开，其核心价值来自于所传达的信息本身。而实时互动除了信息本身，还有很重要的一部分价值，来自与传输过程中的交互方式、交互场景，由此创造的信息维度更丰富、信息量更大。实时互动通过对 CPaaS、RTC、音视频云技术的组合、对实时性体验的极致突破、对场景化的加强，

形成了一个更广泛、更全面的互联网基础设施。

与 CPaaS 相比，实时互动更强调情境信息、多种沟通关系的传递，以及对实时性的极致突破。在体验上，CPaaS 强调跟传统电信能力的结合，并非互联网原生，通过短信、电话、电子邮件等方式与用户建立连接后，跳转至 App 内部进行深度沟通，无法创造连续的、沉浸式的体验。在核心指标上，CPaaS 注重到达率，以异步信息处理为主，而实时互动强调互动连续性和实时性，需要完成多维信息的实时同步处理。在行业参与者上，CPaaS 偏重于 PaaS 的参与者，实时互动更加全面，包含所有层面的致力于在实时互动领域进行研发和创新的参与者。

与 RTC 相比，实时互动是基于其技术的、涵纳多场景的全方位升级。RTC 强调“通信”（Communication），要求对语义信息进行高质量与高效率的传递。而实时互动更强调“互动”（Interaction）和“情感交流”（Engagement）。实时互动不以语义信息的交换为唯一目标，旨在以数字化方式，创造用户所需的所有共享时空，满足其精神需求。

与音视频云相比，实时互动包更强调实时的传输速度与多类型的信息内容。音视频云提供的音视频工作流，通常为单向传输，对时延要求不高。实时互动不仅需要满足音视频的双向传输与实时传递，还需添加大量其他信息如实时信令。多向的互动方式和多类型的信息内容，决定了实时互动所面对的网络优化和延迟问题更加复杂，对于网络时延、卡顿率、到达率有着更高要求。

2.实时互动与元宇宙的关系

实时互动技术是元宇宙的底层技术支撑，元宇宙是融合各项实时互动技术和具体场景需求的数字化应用场景。

2021 年 3 月，UGC 游戏开发商 Roblox 的上市引发了元宇宙热潮，同年 10 月 Facebook 改名为 Meta，彻底引爆元宇宙。元宇宙是发展中的概念，是整合多种新技术而产生的新型虚实相融的互联网应用和社会形态。它基于扩展现实技术提供沉浸式体验，基于数字孪生技术生成现实世界的镜像，通过区块链技术搭建经济体系，将虚拟世界与现实世界在经济系统、社交系统、身份系统上密切融合，并且允许每个用户进行内容生产和编辑。

就技术要素而言，元宇宙包含了底层计算与 AI 能力、硬件计算与展示设备、操作系统及应用软件、以及区块链技术。元宇宙与当前互联网所使用的技术类似，但更多的是对于实时互动、区块链等底层技术的组合和综合利用。

在始终在线、实时渲染、沉浸式的元宇宙里，所有的信息可以实现多向的实时传输与交互，用户在现实生活中的基础五感甚至是空间感、平衡感的感知系统都需要通过新技术得到调动，因此需更低延时的传输网络、更强大的边缘算力、更丰富的交互工具与设备来提供支持。总而言之，实时互动技术的升级极大地推动了元宇宙的发展，同时，元宇宙场景中极致体验的探索激发了实时互动技术的迭代。

(一) 实时互动产业图谱

在实时互动行业爆发阶段的早期，行业的发展重点主要聚焦在底层基础设施的建设及中间工具的研发上。随着稳定高质量的基础设施与简单低门槛的接入方式的完善，更多行业应用者深挖需求，融合实时互动，进行场景创新与业务突破。根据行业生态结构可将实时互动行业分为基础设施提供方、专业服务提供方、行业场景应用方和场景生态企业四大类。基础设施提供方通过底层技术的迭代、专业服务提供方通过核心技术的突破、行业场景应用方通过应用场景的创新、场景生态企业通过全链路的生态运营，共同推动与拓展实时互动行业边界。



来源：中国信息通信研究院

图 3 实时互动相关产业图谱

1.基础设施提供方

基础设施提供方包括了最底层的硬件厂商、运营商和云服务提供商。在实时互动赛道中，基础设施提供方的投入成本大、研发周期长、技术门槛高、规模效应显著，具有明显的头部效应。基础设施提供方往往属于重资产模式，产品差异不大，寡头竞争格局稳定。

（1）硬件厂商

硬件厂商包括芯片厂商、服务器、交换机等基础设施的提供方，其核心壁垒在于硬件集成性和一体化能力。硬件厂商具有成熟的、规模化的硬件设计、制造和分发的能力，并依托自身能力禀赋切入实时互动的细分赛道。如作为 GPU 行业领导者英伟达推出了适用于 3D 设计协作和模拟的 Omniverse 平台，国内 GPU 芯片设计创新者摩尔线程推出了 GPU 物理引擎 AlphaCore 及数字人解决方案 DIGITALME，拥有全场景通信技术的紫光展锐则以面向 5G 领域的芯片为万物互联赋能。

在实时互动的场景创新上，硬件厂商常以自身计算资源或硬件为基础，撬动生态资源，以获取更高层面的技术类、应用类创业者的合作，例如英伟达推出的初创加速计划“INCEPTION”。

（2）运营商及云服务提供商

运营商及云服务提供商集合硬件厂商的能力，通过在全球部署服务器、交换机等，提供稳定的通信服务。其核心壁垒在于强大的网络效应，在追求卓越的产品稳定质量和标准化程度时，通过规模效应降低成本，提升可靠性。

在实时互动的场景创新上，运营商及云服务提供商利用云计算资源和丰富完善的生态，为行业创新者提供成本支持与场景孵化。例如移动云的万象计划，亚马逊云科技推出的云创计划，甲骨文的创新扶持计划等。

2. 专业服务提供方

专业服务提供方可分为核心 PaaS 服务商、专业工具与算法提供商，以及行业 SaaS 服务商，通过技术自研，为下游的场景应用提供专业服务。在实时互动产业中，专业服务提供商业务发展时间较长，技术相对成熟，并且通过长期的实践经验，形成了独特的竞争优势。但这种竞争态势并不稳定，极易受上下游厂商业务边界拓展的影响。

（1）核心 PaaS 服务商

核心 PaaS 服务商打通了实时互动技术的全链路，提供从端到端的完整解决方案，为实时互动行业应用提供针对性的基础能力支持。具体而言，核心 PaaS 服务商提供的能力涵盖音视频及其他互动信息的采集、处理、编解码及传递。

核心 PaaS 服务商通过强大的自研能力、快速的工程能力与深入的行业整合能力三方面建立核心壁垒。首先，强大的自研能力构成了 PaaS 厂商的核心竞争力，通过长期的研发投入，构建具有稳定性与先进性的核心网络，为场景业务提供强有力的支撑；其次，快速的工程优势，使 PaaS 企业能够快速融入第三方能力，共同打造面向行业特性的 aPaaS 解决方案，提高产品的易用性；最后，深入的行业整合能力，让 PaaS 厂商能快速完成行业洞察，抽象出行业痛点与

解决方案，并横向输出至其他应用场景。

核心 PaaS 厂商的典型代表为各类实时互动 PaaS 厂商，例如提供 RTE-PaaS 的声网，提供 CPaaS 的 Twilio、网易云信和环信等，提供云游戏服务的阿里元境、海马云和蔚领时代等。同时各大云厂商也推出了相应的实时互动服务，例如腾讯云 TRTC、火山引擎 RTC、华为 SparkRTC、移动云 MRTC 和优刻得 UEC 等；在 OS 方面，还有中移物联网的 OneOS 等。

对于核心 PaaS 厂商而言，建立生态、连接底层基础设施服务商和上层应用服务商、提供一体化服务能力是参与场景创新的主要模式。通过技术合作伙伴完成能力互补，提升基础设施的易用性与实用性；通过场景合作伙伴完成创新场景的探索与验证，丰富技术产品矩阵；通过服务价值链的构建与完善，实现更高效的增长闭环。典型代表为声网的声选计划、腾讯云音视频的 V+计划、元境的云游戏研运一体化服务、Twilio 的 build 项目等。

（2）专业工具与算法提供商

专业工具与算法提供商专注于某个功能插件、技术能力的提升，通过长期研发与产业实践，实现产品的高效能与高可用，为其他企业提供技术支持。专业工具与算法企业通过将核心能力标准化，并以 API、SDK 的调用形式对外提供服务，赋能更多行业场景应用。典型代表为各提供 AI 算法、实时引擎、具体插件组件等的厂商，例如智能语音领域的科大讯飞、计算机视觉领域的商汤、游戏引擎领域的 unity、cocos 等。

专业工具和算法提供商核心壁垒是建立算法和工具的性能优势与易用优势。通过建立高精尖人才队伍，保持技术的持续投入、痛点需求的持续洞察，完成算法与工具的优化迭代，提升核心性能；同时降低调用门槛，提升易用性、可复制性与可扩展性。

开源降低了技术的使用成本，但对专业工具和算法提供商提出了商业层面的挑战。对于此类企业而言，建立严格的授权认证体系，深入真实业务场景、完成应用绑定，是进行场景创新的重要途径。

（3）行业 SaaS 服务商

行业 SaaS 服务商兼具行业洞察与技术研发两大优势，为用户具体的行业场景提供一站式的解决方案，帮助客户完成快速部署。其核心优势在于用高效的产品与技术解决用户痛点问题，帮助用户进行商业落地。SaaS 服务商的场景创新来自于与上下游的持续沟通与交流，通过与上游客户的交流，明确行业需求与痛点，打造具有针对性的产品或解决方案；通过与底层技术厂商的交流，了解技术演进动态，共同以行业需求为导向探索技术创新。

3. 行业场景应用方

行业场景应用方具有明显的行业属性，深耕于用户需求，对特定场景需求具有深刻理解，同时可以针对行业痛点，借助上述技术厂商的能力，提出用户可感知的解决方案。行业场景应用方的核心优势来自于对用户的理解，通过明确的场景洞察、闭环的用户运营、优质的产品体验建立核心壁垒。

行业场景应用方是当前探索新场景应用的先行者，主要专注于

垂直行业的研究，直面最终用户，极大丰富了实时互动的场景版图。目前实时互动场景创新主要集中在泛社交、游戏、企业协作、教育等领域。庞大的用户基础与直观的用户反馈，使应用方可快速完成产品的最小可行性验证，实现创新。对于数字化水平相对薄弱的传统行业，更需借助业务流程的数字化重塑，抽丝剥茧发掘新场景，同时通过引入创新技术与智能设备，完成实践升级。

4. 场景生态企业

场景生态企业往往具备雄厚的技术积淀，通过长期耕耘，积累了海量客户企业资源，具有丰富的行业实践经验，了解客户核心痛点及关键能力。这类公司核心优势来自于强大的资源整合能力，通过调用上下游资源，结合自身能力迭代，提供面向多场景、涵盖多技术的一站式灵活解决方案，是实时互动最重要的领导者与支持者。在面对新场景时，场景生态企业会快速更新知识结构，与关键合作伙伴共同打造具有核心意义的技术应用方案，如泛娱乐中的实时合唱、元宇宙中的语聊房。在这类企业的生态中，往往可以往看到排列组合丰富的云市场、插件市场等。

场景生态玩家的典型代表既包括腾讯、阿里、火山引擎、咪咕、甲骨文等具备综合业务的公有云厂商，也包括声网、移动云、中移物联网等可提供一站式服务的 PaaS 服务提供商。

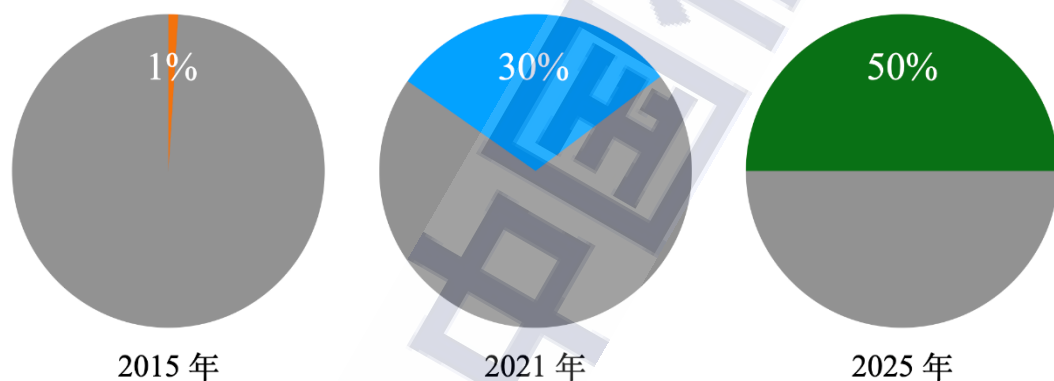
在参与实时互动场景创新时，场景生态企业会综合自己在实时互动领域的技术能力与合作伙伴的技术能力，协同其他非技术类资源，例如分发流量、IP 版权等，共同推动行业创新。例如腾讯的千

帆计划，字节跳动的火种计划等。

（二）实时互动产业发展现状与趋势

1. 应用场景持续创新，行业渗透率不断提升

C 端场景爆发，传统数字化转型加速，实时互动渗透率不断提升。2015 年，能够包含实时互动功能的 App 仅有 1%，只有少数像 Skype 的专用型应用才会使用包含实时音视频技术模块，而实时互动其他技术应用就更少。2021 年，实时互动的渗透率¹已经超过了 30%。据公开资料显示，2025 年实时互动的渗透率将超过 50%。



来源：公开资料整理

图 4 实时互动渗透率

C 端新场景的爆发，引领实时互动创新浪潮。根据 2021 年 10 月对全球实时互动热门场景的统计，C 端应用成为了实时互动爆发的主力增长点，各种新场景新玩法迅速涌出。在国内，为用户带来全新交互体验的在线 K 歌房、智能台灯，成为新一代生活娱乐方式的风向标。在海外，元宇宙以及与体育相关的远程加油、一起看比赛、体育脱口秀蓬勃发展。

¹实时互动渗透率：指包含音视频功能的 App 在整体 App 的渗透率，样本来源于国内 360、华为、小米安卓应用市场三大应用商店在教育、泛娱乐、购物、金融、医疗、企业通信等行业的近万个应用。

区域	中国	其他地区			
		美洲	中东/非洲	欧洲	亚太地区（中国除外）
热门场景	语聊房 游戏语音 小班课	互动播客 视频相亲 1对1课程/小班课	1对1课程/小班课 语聊房 远程问诊	1对1课程/小班课 语聊房 视频相亲	互动播客 秀场直播 教育大班课
新兴场景	在线K歌房 智能台灯 云包厢 元宇宙	虚拟活动 元宇宙	云音乐服务	虚拟活动 主播带货直播	韩国：元宇宙 日本：体育远程加油 印度：一起看比赛 东南亚：体育脱口秀 大洋洲：云冥想

来源：公开资料整理

图 5 全球实时互动热门场景及新兴场景

社交泛娱乐是与实时互动结合最早的行业，在秀场直播、电商直播、直播课、演唱会直播等场景中，实时互动仅为工具，承担信息传输职责。但随着用户习惯的变迁、实时互动的技术进步，以及疫情的催化，更多数字原生、互动性、陪伴性更强的场景诞生。例如，2020 年初互动播客中大量科技名人的参与引发巨大的社会反响和关注。互动播客在降低仰视感的基础上进一步通过更自然、直接、真实的实时语音交流拉近名人、意见领袖和所有社会参与者间的距离感，从而创造了爆发式吸引力，验证了实时音视频互动能力对于引爆新媒体形态、改造社会交流形式的巨大影响力。

数字化转型的加速，为实时互动创新提供持续动力。实时互动对于线下应用主要的发展推动体现在对于数字化转型类业务互动和连接方式的推动。数字化转型的趋势和潮流经常被提及，需要重点强调的是，真正的数字化只是基础，而数字化之后的信息如何彼此连接、与线下互动形成联动则是核心，只有数字化普及并与互动结合，才能创造丰富的场景，这就是目前实时互动无论是应用使用量、渗透率提高，还是场景爆发的根本原因。

无论在办公、公众服务还是企业业务经营方面，每一个线下的

场景都持续有线上的创新替代场景出现，这些场景提升了企业效率，提供了疫情下的支持，也代表着未来巨大的机会。

2. 用户规模快速增长，算法兼容性需求加强

用户规模的增长加剧了接入设备多样化的问题，算法向下兼容成为必然。以设备多元化最严重的东南亚为例，共有 2 万种不同的手机和智能设备在活跃地使用实时互动技术；在网络环境上，尽管 5G 商用已经推出，但 Wi-Fi 及 4G 依然为主流，仍有少量用户通过 3G、2G 甚至 LAN 接入服务。此情景推动了实时互动技术向着兼容多价位机型、适配多类系统、满足多种网络环境的技术方向发展。

实时互动服务商基于实时网络，利用 AI 与边缘计算，对传输的音视频质量进行极致优化。未来随着 AI 算法的不断加持，实时互动在网页端也能够更好地实现本地客户端感受，千元机型也可以满足各类 AI 算法应用条件。

3. 终端设备不断进化，交互界面持续更新

交互终端的进化，为全新、多元的实时互动场景提供新的载体。更多的 IoT 设备、移动智能终端设备融入实时互动功能，持续解锁场景。

人和世界互动的界面在进化。大量穿衣镜、冰箱、抽油烟机通过接入语音模组和视觉模组，提供实时交互体验。在 Peloton 之后，健身行业已被实时互动技术形态和能力深深地冲击和改造，进而引发中国的“千镜大战”，中国消费者也开始形成利用具有实时互动功能的智能设备进行健身。

此外，移动智能终端也在不断进化，随着 VR/AR 技术的突破和产业链的完善，2021 年 VR 头显销量突破千万，AR 眼镜光波导量产工艺等技术瓶颈取得突破，有望走进消费级市场。很多科技巨头都认为 VR/AR 终端将取代智能手机成为下一代计算平台。终端的升级，必然带来体验和交互的升级，比如交互界面从智能手机的平面交互拓展到空间 6DoF 交互，交互信息从传统的声音、图像升级到运动信息、空间定位跟踪信息等，这些对实时交互会提出新的要求，也带来新的机会。

在未来，更多的消费生活方式都将被实时互动行业改变，这些场景释放了人机交互界面的普及度和数字化无处不在的特性，提供了随时随地以海量数据为支撑的沉浸体验。

4. 技术提升用户体验，业务价值优势凸显

技术质量的进步提升了用户使用体验，应用场景中的内容价值得以充分放大。实时互动的超低延时带来了体验优化，高清和互动带来了沉浸感和参与感的提升。

在强调内容知识价值的教育场景，在优化延时、卡顿率的基础上，将大班课转变为互动性更强的 RTC 超级小班课后，转化率提升超过 25%。通过互动的参与，学生的注意力更为集中，学习效果提升明显，进而带来学生对课程价值的认可、家长对课程效果满意度的提升。

在强调娱乐消费价值的秀场直播场景中，当实时互动卡顿率下降 22% 时，观众的日平均使用时长可增加 30%。此外，在某直播应

用中，当视频分辨率从 SD 升级到 HD 时，月平均收入增加了 10%。画质的显著提升带来沉浸感的明显加强，自然带来了业务价值的增长。连麦直播、多人语聊更加增强了用户的参与感，建立了品牌情感，将某直播 App 的日活用户数提高 48%。

5. 功能组件加速融合，推动行业场景创新

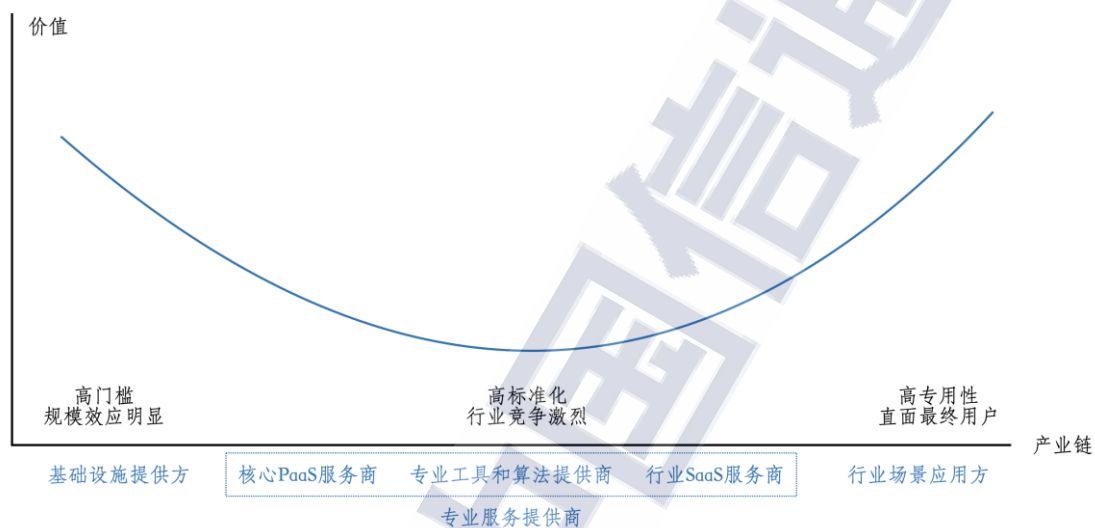
专业化、低代码的功能组件不断成熟，为场景创新提供了专用性与互动性的要素支持。实时互动对于行业的升级是内生的，互动性与沉浸感吸引着更多行业参与者对新场景的深挖。以互动白板为例，完备的白板包含了画笔、橡皮、选择、图形、Float Bar 等基础工具，同时提供轨迹同步、白板分页、回放功能等高级工具，不仅可以赋能教育场景，使老师与学生能针对课件进行动态展示与详细讲解，亦可赋能会议场景，使参会人能进行同屏标注与共同修订，提高协同效率。

以社交泛娱乐行业为例，除了低时延技术的突破，以音乐版权、3D 场景模型、虚拟形象、空间音频为代表的功能组件共同融入，使线上 KTV 场景搭建成为可能，线下合唱的线上化推动了社交泛娱乐的爆发式增长。在直播场景中，互动小游戏组件的加入，提升了主播与观众间的互动，丰富了直播的趣味性。

6. 产业合作更为紧密，生态化运营成为必然

实时互动产业上下游的联系加强，行业的生态化运营成为行业发展的重要趋势。实时互动赛道的生态变化趋势是跟着新技术、新场景、新业务逻辑而进行的。技术创新与商业模式创新融合演进，

加速实时互动产业微笑价值曲线的形成。基础设施层通过高门槛的技术壁垒，放大规模效应，获取高价值收益。直面最终用户的场景应用层因解决问题的高专用性，获取丰富价值收益。中层核心组件与算法由于产出可复制、可标准化，竞争态势更明显，溢价能力稍弱。



来源：中国信息通信研究院

图 6 实时互动产业微笑曲线

微笑价值曲线的存在使行业参与者积极向上下游拓展，生态化运营成为必然。实时互动赛道的专业服务提供方从单一的技术能力输出，转向生态聚合能力建设，以整合能力共同推动行业进步。创业者可以利用头部企业提供的顶尖技术与先进算法，快速进行场景探索与商业验证。成熟企业可借助上下游力量，进行业务实践与落地，推动技术的快速升级。

（三）产业规模预测

实时互动产业包含了从底层的基础设施，到上层的行业场景应

用，是一个复合的、创新的行业，其行业规模可从 To B 技术层口径与 To C 应用层口径两个维度进行预估。

从 To B 技术层口径而言，实时互动的市场规模可从基础设施提供方、专业服务提供方、行业场景应用方的相关业务收入维度进行统计。

表 3 实时互动产业规模统计口径（To B 技术层）

角色	相关厂商	相关业务
基础设施提供方	运营商	实时通信类业务，例如 VoIP 业务，5G 消息，5G 新通话等
	云厂商	<ul style="list-style-type: none"> ● 实时通信产品：实时音视频、实时信令等 ● 云厂商提供的云加速、云渲染产品服务 ● 互动解决方案中配套实时互动的音视频录制、存储、转码等方面的产品收入
专业服务提供方	核心实时互动服务商	核心实时互动服务提供商业务收入
	云加速服务商	云加速服务商业务收入
行业场景应用方	实时互动直接应用方	教育、泛娱乐、IoT 等大型应用企业针对实时音视频的自研投入
	整体解决方案提供方	教育、工业、交通、政务等实时互动整体解决方案厂商的配套软硬件基础设施投入

来源：中国信息通信研究院

根据公开数据和专业机构的调研数据，2021 年全球实时互动服务规模约为 35.6 亿美金，三方实时音视频服务为 7.32 亿美金。中国的实时互动服务规模为 77.13 亿人民币，三方实时音视频服务规模为 15.86 亿人民币。类比 CPaaS、RTC、音视频云的复合增长率，预计

实时互动市场未来 5 年平均每年将迎来 35% 的增长。预计到 2025 年，全球实时互动服务规模将达 120 亿美金，其中，中国市场规模达 256 亿人民币。

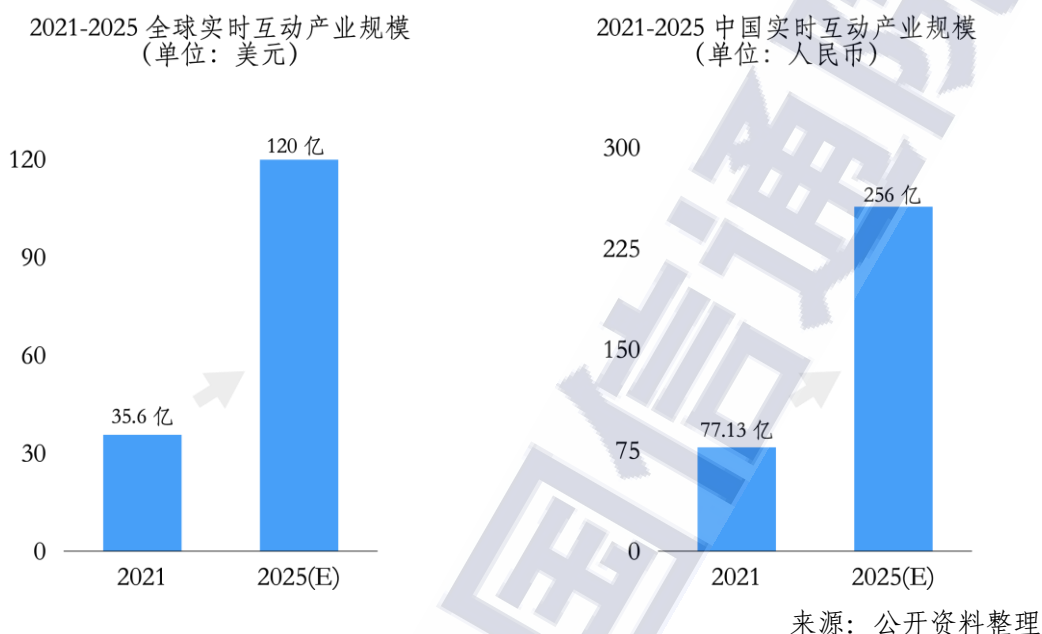


图 7 全球及中国实时互动产业规模统计预测

从 To C 应用层口径而言，实时互动为直播、电商、社交泛娱乐、企业协作、健康医疗等行业提供技术支持，已经成为互联网产业的底层新基建之一。

在直播电商领域，截至 2020 年底，我国直播电商用户规模已达 3.88 亿人，占整体网民的近四成，直播下单用户占观看直播用户的 66.2%，基数潜力庞大。据数据统计，2020 年底中国直播电商市场规模达 1.2 万亿，并预计在未来三年年均复合增长 58.3%。

在泛娱乐行业，根据工信部的估计，2020 年中国网络视频活跃用户规模达到 10.01 亿，市场规模达到 2412 亿元，同比增长 44%，增长趋势明显。中国演出行业协会发布的《2020 年中国网络表演（直播）行业发展报告》显示，2020 年我国网络表演(直播)行业用户规

模已达 6.17 亿，市场规模达 1930.3 亿元，增长空间巨大。

在社交行业，借助实时互动的音视频社交增速明显，发展空间广阔。据弗若斯特沙利文预测，2021-2024 年全球社交媒体市场规模年复合增长率有望达到 15.2%，其中音频社交市场规模 CAGR 为 27.6%，音频社交市场规模为 25.8%。同时预计 2024 年全球社交媒体市场整体规模达 3000 亿美元，其中视频社交为 1287 亿美元，音频社交为 526 亿美元。

在企业协作市场中，实时互动可作为构建视频会议、远程协作解决方案的基础。根据 IDC 预测，企业协作应用市场预计将从 2018 年的 97 亿美元增长到 2023 年的 177 亿美元，2018-2023 年的 CAGR 为 12.8%。

健康医疗方面，根据弗若斯特沙利文报告显示，2024 年中国在线问诊市场占总问诊量的百分比预计将从 2019 年的 6% 扩大至 42%，在此背景下，在线问诊市场规模将就会由 2019 年的 90 亿元扩大至 1550 亿元。

实时互动与每一个场景的结合都将激发出更多潜力空间，随着实时互动的渗透率加大，实时互动服务能够撬动的应用市场将突破万亿人民币。

四、实时互动关键技术

实时互动关键技术包括专有基础设施、专业组件和重要共生技术。其中，专有基础设施是面向实时互动场景的基础设施，核心技术包括实时网络、传输协议、边缘计算和终端适配等；专业组件是

在专用基础设施上，用以处理实时音视频等数据的音视频引擎、云加速、实时信令、数据处理、人工智能等技术；重要共生技术是指实时互动应用过程中，为解决重要衍生问题诞生的技术。如下图所示：



来源：中国信息通信研究院

图 8 实时互动关键技术

（一）专有基础设施

1.实时网络：以大规模分布式的软件定义实时网络为基础，提供实时性保障

实时网络利用分布广泛的边缘计算节点，运用智能路由、智能调度等技术，开创性地实现了实时互动领域上层应用与底层网络传输横向分层的解耦，为实时互动场景应用开发者解决了传输延时、网络波动以及研发成本等问题。

（1）软件定义实时网络 SDRTN

SDRTN 是专为实时互动设计的虚拟通信网，是基于底层互联网，通过软件定义算法构建的、有 QoS 保障的虚拟实时网络。

SDRTN 服务商依照 SLA 协议，向网络运营商采购底层的公共

互联网带宽，提供有 QoS 保证的叠加层实时互联网，来支持最终客户在应用层设计与构建实时业务场景。SDRTN 技术的发展，进一步推动了智能路由与智能调度的算法迭代和技术革新。

（2）智能路由

智能路由指的是以提供 QoS 质量保障为目标的、建立在 SDRTN 上的路由算法与工程技术，又称 QoS 路由（QoS R）。实时互动行业与 SDRTN 技术的发展，进一步推动以 QoS 为目标的、在 SDRTN 多种传输模式下的智能路由与智能调度的算法和技术的革新。目前 SDRTN 通过分布式与集中控制相结合、理论计算与负反馈系统相结合的方式，保证了在 Internet 复杂网络模型下的算法正确性。

（3）智能调度

SDRTN 的智能调度包含两方面：终端设备接入 SDRTN 边缘节点的选择和 SDRTN 系统内部带宽的调度。智能调度的本质是在实时互动数据传输质量与传输成本等诸多约束条件下，综合运用工程与算法手段，得到实时近似最优解。质量方面需要考虑用户地理位置、网络运营商、本地网络特征、设备性能、实时互动的场景类型、场景对质量要求的 SLA、参与互动的人数等条件；成本方面需要考虑资源预留、资源利用率、带宽削峰填谷等问题。目前行业领先的实践是在接入负载均衡、骨干网路由规划、实时质量检测的基础上，结合机器学习预测方法，将多类场景需求抽象转化为数学优化问题，并通过采用最优化算法，寻求最优决策方案。

2.传输协议：新一代 UDP 协议，推动抗弱网能力、媒体协议升级

在实时互动多数应用系统中，大量客户/服务器模式的网络应用都需要使用 UDP 协议。UDP 是一种无连接的传输层协议，提供面向事务的简单不可靠信息传送服务，其与 TCP 协议一样用于处理数据包，处于 IP 协议的上一层。UDP 不提供数据包分组、组装、排序，不产生任何额外的数据，因此数据传输时间短，实时性高。在强调传输性能而不是传输的完整性的场景中，UDP 可被视为最优选择。由于 UDP 不强调传输的完整性，实时互动在其传输上需要新的弱网²对抗能力和新的媒体文件协议来支持。

（1）弱网对抗

弱网对抗技术主要包括**抗丢包**、**拥塞退避**、**带宽估计**、**抗抖动**四个方面，在面对各种复杂网络条件时，保障实时互动过程流畅清晰。

（2）IETF 的媒体协议 RTP 以及媒体质量反馈协议 RTCP

RTP/RTCP 协议是实时互动通信的基石。RTP 协议（实时传输协议）定义流媒体数据在互联网上传输的数据包格式，RTCP 协议（实时传输控制协议）负责可靠传输、流量控制和拥塞控制等服务质量保证。RTP/RTCP 模块作为传输模块的一部分，负责对发送端采集到的媒体数据进行封包，然后交给上层网络模块发送；接收端在

²网络的链接形态丰富多样，除了有线连接，还有数据网络，WIFI 等多种移动网络连接。人们在使用实时互动随时随地沟通过程中，难免会遇到网络拥塞，信号干扰，信号弱，防火墙拦截等各种原因导致网络传输出现丢包，时延，抖动。通常将有高丢包，高时延，高抖动的网络称为弱网。

RTP/RTCP 模块收到上层模块的数据包后，对媒体数据进行解包操作，最后把负载发送到解码模块。

3.边缘计算：边缘下沉部署、就近接入计算推动提升实时交互性体验

随着直播、云游戏、边缘 AI 等场景的爆发，实时网络需要在高并发、广覆盖、低延迟的基础上进行计算能力的扩展。边缘计算通过云-边-端无缝协同，将传输数据在边缘节点就近处理，减少边缘节点到骨干网之间的数据传输量，降低延迟与卡顿。

实时互动场景下的边缘计算存在巨大的技术挑战：（1）实时互动场景往往面临全球化、广域覆盖问题，这要求边缘计算必须建立在可靠通信与传输的基础之上，以保证边缘与边缘间的交互质量；

（2）实时互动引入了大量有状态、长生命周期的分布式服务，对系统架构和可维护性提出了很高要求；（3）实时互动场景更多是计算密集型场景，对算力、带宽等资源要求更为苛刻。总之，在不可靠的边缘基础设施上，构建复杂的业务场景，同时兼顾成本与体验，这对边缘计算中的边缘部署和分布式计算都提出了极高要求。

4.终端适配：信号处理结合人工智能持续推进终端设备覆盖千家万户

实时互动场景的实现需要多种终端设备的支持。智能终端设备作为实时互动的接入设备，提供了感知、连接、计算与交互的功能。

（1）感知能力

智能感知设备是具有计算能力的终端设备，可对线下行为进行

捕捉并转换为线上数据。目前智能终端设备中应用较多的传感器包括视觉传感器、声音传感器、运动传感器、环境传感器以及生物传感器等五大类。

（2）网络连接能力

智能终端设备需要具备网络连接的能力。终端设备有多种网络接入方式，比如：固定网络、WiFi、4G/5G 移动网络、蓝牙、ZigBee、NB-IOT 等。IoT 设备低功耗、低成本的设计，导致其网络通信能力偏弱。若想在已有的 IoT 设备中增加实时互动能力，需要对现有网络进行改造，例如增强媒体引擎、优化网络对抗策略等。

（3）计算能力

随着人工智能技术的成熟与普及，基于深度学习的人工智能算法能够带来更加智能化的体验，因此端侧 AI 的算力也成为了智能硬件长期发展的制高点。智能终端的计算能力主要包括**通用计算**（由 CPU 进行）、**图形渲染**（由 GPU 进行）、**图像处理**（用于图像采集）、**视频编解码**（Codec，用于图像/视频的压缩与还原）和**神经网络/AI 处理**（用于端侧 AI 硬件加速）等。尤其是在以元宇宙为代表的实时互动新场景中，需要同时拥有强大的渲染能力，音视频编解码能力和 AI 加速能力。多功能 GPU 不仅能够进行实时渲染，还支持丰富的音视频格式、8K 超高清、一体化快编和多路极速超高清转码，同时通过 AI 加速视频画面的分类、检测和识别过程，实现窄带高清，有效提升画面质量和流畅度并节约带宽，满足 RTE 的各种应用场景对算力的需求。

（4）交互技术和设备

随着交互方式从传统的单类型、单方向、异步交互向多类型、多方向、实时交互演变，交互技术、终端设备类型都发生了巨大的变化。

交互技术：终端的交互技术包括触觉交互、语音交互、视觉交互等。触觉交互从传统鼠标、键盘的机械反馈向新型设备的震动反馈发展。语音交互经历了单声道、双声道、立体声的发展过程，正向空间音频演进。视觉交互从低像素的 VGA 向 4K/8K 超高清、从平面视觉向立体视觉发展；并且随着 AR/VR 等新型设备的出现，视频交互逐渐朝着沉浸式体验的方向演进。

终端设备：既包括手机、PC、AR/VR 眼镜、智能可穿戴硬件等常见设备，也包括虚拟现实一体机、全息投影仪等新型设备。这些硬件集成了多种交互技术，保障了交互的实时性、同步性。

（二）专业组件

1. 音视频引擎：场景化、智能化的编解码与前后处理技术为实时互动赋能

音视频引擎是实时互动应用系统中处理音视频的核心技术组合，也是开发实时互动程序或系统的核心组件。利用音视频引擎，开发者可迅速建立、铺设程序所需的音视频功能。音视频引擎相关技术包括音视频编解码技术、AI 编解码技术、音频 3A 技术、网络体验技术、空间音频技术等。

音视频编解码是将数字音视频数据进行压缩和解压缩的算法与

工程技术，是实时音视频互动场景中的基础技术之一。音视频压缩可以有效的减少音视频传输所需要的带宽，进而提高音视频传输在实时互动场景中的效率与鲁棒性。音视频编解码技术解决了数据传输中的低采样率、低编解码还原度、低编解码算法、高丢包率的问题，使用户听到更真实的音频原声，看到更高清的视频内容，减少内容损伤、质量损失带来的不好体验，保证实时音视频的稳定传输。

AI 编解码是指基于人工智能的音视频编解码算法与工程技术。AI 音视频编解码能在保证音视频质量的前提下，用更少的比特进行数据压缩，从而节省更多的带宽，提高实时音视频互动的并发数和低带宽情况下音视频的质量。

音频 3A 技术即语音增强前处理，主要负责拾取近端环境中的目标声音，同时抑制包括回声和噪声在内的干扰信号，提高通话、录音、识别等场景中的上行音频信号质量。随着近几年 AI 领域技术的不断进步和成熟，在 3A 任务中有越来越多的 AI 技术实现了商业应用，并显著提升了 3A 的性能和功能。

网络体验技术是实时互动场景中为了保证音视频数据在网络中的传输质量而采取的多种网络技术和算法相结合的工程技术的统称，是实时音视频互动技术的重要环节，其中包括 ARQ 自动重传请求、FEC 前向纠错、NetEQ 音频网络抖动缓冲器、Jitter Buffer 视频网络抖动缓冲器、GCC 拥塞控制算法等。通过合理运用各种网络技术和算法，来对抗网络传输过程中出现的丢包、高时延、窄带宽、网络抖动等问题，从而达到低延时、低卡顿的效果，显著提升实时互动

场景的用户体验。

空间音频技术是指为使用户在音频听感上具有空间感，通过模拟头部球面区域立体声场而构建的纯算法方案。通过空间音频技术让听音者感知声源方向、距离以及所处的声场环境，从而提升在线互动的沉浸式体验。根据实时互动应用场景，可以分为真实环境的空间音频重放和虚拟环境的空间音频渲染。

2.云加速：云边一体、就近处理，减少接入与链路时延

云加速指通过广泛部署的边缘服务节点，提供靠近用户的网站服务、文件下载、音视频推拉流等加速服务，实现显著的用户体验提升。

互联网的泛在性和开放性，为全球各地的用户获取信息提供了极大的便利。但在提供优越连通性的同时，难以保障数据传输的质量，只能提供尽可能送达的服务。在全网用量高峰期或移动终端信号不稳定时（特别是在跨国、跨区域、跨运营商互联时），会出现端到端延迟增加、音视频卡顿的现象，导致用户体验感下降。实时互动场景下，用户对交互延迟、卡顿等问题极为敏感，稳定可靠的数据传输至关重要。云加速技术通过靠近用户的边缘服务节点，提供稳定可靠的接入服务；通过对大量信息的就近处理和边缘节点缓存技术，减少链路时延。此外，云加速技术，通过云边端一体的全链路音视频服务能力，完成实时互动的数据传输优化，提升时延、卡顿、网络兼容等多个维度的指标质量，从而保障端到端的数据传输质量，满足实时互动的极致体验要求。

3.实时信令：可靠、高并发的实时分布式信令控制技术丰富实时互动场景的沉浸感

实时信令技术铺设在实时网络基础设施上，支持超低延迟、支持数以亿级的高并发，在高丢包环境下依然具有高可靠率。分布式的实时信令技术不仅可以传输实时控制指令和消息，同时也可以实时同步用户自定义状态等，能够满足应用内的文字消息、弹幕等轻聊天场景。在面向未来的应用中，云信令不仅要传输实时控制指令和消息，同时也需要实时同步用户自定义状态等。会议、教育场景中的实时聊天，游戏中的玩家操作动态，直播中的礼物特效，均可通过实时信令进行传输。

4.数据处理：分布式数据体系从监控面与调试面为实时互动保驾护航

实时互动对实时可感知的服务提出了更高要求。在常见的实时互动应用场景下，如大规模的互动直播、视频会议、在线教育，客户需要实时了解该场景下从整体到个别用户的体验情况，并快速响应由于个别用户的设备、网络环境差异造成的体验问题。分布式数据库和实时数据监控分析工具的使用，让每个用户的体验状况都可实现实时反馈，为技术开发人员针对性的解决问题和总结经验提供了很大的便利。

5.人工智能：AI 算法提升交互质量与效率，丰富用户体验

人工智能推动实时互动在各个行业的各个场景下的应用更加智

能和完善，对于音视频产业的发展也愈加重要。人工智能技术的应用主要体现在以下几个方面：

（1）人工智能增强实时互动体验，助力体验质量提升

人工智能增强了交互上的体验，最具代表性的场景是超分。视频在端到端传输时，通过 AI 算法识别特征信息进行视频超分，最终在设备端来显示更高清的视频效果。针对实时互动的不同场景，人工智能能够根据场景适配不同的音频算法以达到最佳质量，同时进行智能降噪处理，提升实时互动的用户体验。此外，AI 推理引擎降低了对 AI 算力的需求，针对需要大规模算力的音视频传输场景，更好地发挥芯片性能与算法效用，实现了提效降费。

（2）人工智能丰富交互体验形式

实时互动把线下的场景搬到线上，在人工智能的助力下，甚至出现了超越线下体验的交互形式。例如在实时互动中，通过人脸识别、可完成面具变脸与情绪识别，通过音效变换可增强用户的娱乐体验，通过实时翻译与实时转写可实现跨国会议的顺畅沟通。这些人工智能技术包括但不限于：**语音合成(TTS)、自动语音识别(ASR)、声纹识别技术、实时变声技术、背景分割、美颜、手势/表情/姿态识别、三维重建、数字孪生、虚拟替身等。**

（3）人工智能提升业务效率

实时互动场景下，在同一时间会产生海量内容，为人工内容监管和质量评价带来了很大挑战。训练有素的 AI 可以很好的通过语音、图像识别和图片、音视频审核对实时互动的内容进行质量评估和内

容监管，极大减少人工审核的量级，实现业务增值。

（三）重要共生技术

1.低代码平台：组件化与编辑引擎降低开发门槛、提升场景制作效率

实时互动场景下，需要大量的 UGC 内容和更多场景化组件来支持沉浸感和代入感，因此很多实时互动的服务提供商都会建立自己的低代码平台。低代码平台也能够帮助普通开发者构建出优质体验的专业实时互动应用。低代码平台主要通过提供基于低代码平台标准化交付的组件、模板、数据连接器，和支持拖放功能与可视化指导实现编排与部署能力的可视化编辑工具，来简化实时互动应用程序开发。

2.安全合规：维护实时场景秩序，保障用户安全

实时互动提供了全新的场景和体验，也带来了更复杂的交互关系与更多、更快的数据交换，这对于基础设施安全、网络安全、数据安全、运营安全等都提出了更高的要求。

（1）基础设施安全

在实时互动场景下，需要基于不同承载功能和安全级别，将网络划分出核心、边缘、IT 等几大安全区域。在不同的安全域之间，需要根据不同的业务访问需求和安全级别，制定不同的路由策略以及严格的安全访问策略。

（2）网络安全

为了确保实时互动服务的合规及安全性，在架构设计上，需要充分考虑实时互动在互联网环境面临的安全风险，通过**资源隔离、频道隔离、加密传输、身份认证、SDK 安全**等控制措施为开发者及用户提供安全稳定的服务。

（3）数据安全

数据作为实时互动信息活动的载体，经过合法合规且安全的处理尤为重要。数据安全技术是贯穿在实时互动的全生命周期中的，包括：**最小化的数据采集原则、数据脱敏、数据保护和加密传输、数据使用和存储。**

（4）运营安全

在实时互动服务的设计和架构阶段，为了能够更早的发现风险，会通过威胁建模来识别潜在的安全问题并实施响应措施。在安全测试层面，实时互动服务的运营更注重 DevSecOps 崇尚的内置安全防护，在 CI/CD 层面进行黑白盒工具的集成，包含开源代码扫描工具 SonarQube，组件及合规扫描工具，App/SDK 扫描工具等，从而完善在集成发布过程中的风险监测。在落实纵深防御以应对威胁的基础之上，实时互动服务在最小权限范围内采集用于安全分析的日志，对确认的风险，根据应急响应机制进行处置和追踪，以保障业务系统的安全性和可用性。

五、实时互动赋能千行百业

实时互动加速产业数字化渗透，助力数字经济规模化增长。实时互动将成为数字经济新的生产力来源，通过提供创新的生产工具

卷入更多的生产资料，并与其他的生产力形成协同，加速经济的发展。

一方面，以实时互动技术为基础的在线办公、在线教育、远程医疗、网络视频、数字文旅等，作为数字化新业态新模式蓬勃涌现；另一方面，实时互动将促进数字经济对于第一产业、第二产业的渗透，加速生产体系的完备升级，实现供需精准对接、高效生产和统筹调配的科技化运作模式。产业数字化，尤其是实体经济数字化转型，对于人工智能、大数据、区块链、IoT 等技术上的需求层级进一步提升。产业数字化依托于实时互动技术建立的万物互联、低延时、高可用的网络，辅之以满足具体业务场景需求的强互动工具，实现产业的自动化与智能化升级。实时互动技术作为新一代数字经济发展的底座支撑，将大幅提升行业全要素生产率，助力传统产业的转型升级，加速新型数字产业体系的形成与发展，开辟经济增长的新空间。

实时互动通过技术创新创造新体验，推动新增长，主要从以下几方面推动新场景的产生：（1）超分、插帧、空间音效等技术**为线下场景线上化带来了业务流程重塑**；（2）互动白板、远程控制、自动转写等为**互动工具创新带来了效率提升**；（3）超低延时、高并发、抗弱网的网络传输技术突破带来了**信息传输的高质量变革**；（4）美颜、变声、虚拟人等 AI 技术发展**实现了超越自然的新体验**。这些技术创新给不同行业的用户带来可感知的新体验，在与具体场景交叉融合中，激发、赋能千行百业的新场景。

表 4 实时互动应用新场景（示例）

应用行业	线下场景 线上化	互动工具 创新	网络传输 技术突破	AI 技术发展
企业协作	在线多人会议	互动白板协作	在线会展	虚拟形象会议
教育	三个课堂	互动白板	实时大班课 小班课	音乐、美术陪 练
IoT	VR 高清会议	互动健身镜 实时力反馈手 套	儿童智能手表 智能台灯	
制造	数字孪生制造 OS	远程实时 AR 专家指导	远程无人作业	
交通	远程车险处理		自动泊车	
金融	虚拟营业厅	金融双录		虚拟柜员
房产	VR 远程看房	远程协作渲染		
医疗	视频问诊		远程超声会诊 视频急救	远程医疗机器人
政务	视频办医保			
文旅	线上虚拟展馆	AR 寻宝 AR 导览 AR 文物复原		沉浸式视觉展
泛娱乐	虚拟会展 线上演唱会	语聊房 云包厢	直播连麦 在线合唱	虚拟偶像
游戏	剧本杀类游戏	游戏内语音	云游戏	虚拟人 NPC
社会关怀	免费视障服务 平台		老人监控系统	定向智能耳蜗

来源：中国信息通信研究院

（一）实时互动与企业协作：优化流程，提升效率

疫情加速远程办公的应用，企业服务迎来发展新机遇。同时，在线协作场景逐渐丰富，视频会议、协作文档、企业培训、远程面试等线上线下协同的工作方式都已成为日常所需。以典型的视频会议为例，用户可通过手机、电脑、平板、小程序、会议系统等多终

端接入，利用屏幕共享、互动批注、协作白板、在线文档等功能插件，完成多人办公协作的需求。疫情期间，腾讯会议、WPS+会议、钉钉等向社会免费开放多人会议，帮助企业维持日常运营与发展。基于实时互动音视频云的在线会议以大规模、低延时、抗弱网、高接通率的特性，得到企业和用户的高度认可，据 IDC 估计，未来几年中国云视频会议市场规模的 CAGR 将保持在 25%左右。

在线协作正成为实时互动在企业办公的应用新场景，以腾讯文档、飞书、石墨、金山文档、蓝湖为代表的协作工具，通过将生产资料统一放在云端，不同角色参与者在线接入服务，完成多人实时协作，降低沟通成本、提升工作效率。

企业的在线协作同时也是对“碳中和”的最佳践行，通过减少线下通勤、降低会务用品损耗、增加线上场景的复用，有效减少碳排放量。

（二）实时互动与教育：促进公平，提升质量

优质教育资源具有稀缺性，在线教育打破了时空限制，为用户提供了低门槛、高灵活的触达方式，对促进教育公平、提升教学质量具有重要意义。2020 年 3 月，教育部发布《教育部关于加强“三个课堂”应用的指导意见》，提出促进信息技术与教育教学融合应用、探索信息化背景下育人方式和教研模式，积极推进“互联网+教育”发展。

在互联网+教育的实践过程中，抗弱网、高并发、高可用、低延迟等技术的攻坚，完成了在线课堂从录播课、直播课向实时互动课

堂的突破迭代，信息传输从单向、异步变为双向、同步，教学质量与效率得到显著提升。针对基础教育、职业教育、高等教育、继续教育等教育信息化场景拓展出多样化教学模式。

新东方、好未来等行业领先者在大量教学实践中，平衡教师资源与学生需求，拓展出小班课、大班课、超级小班课等多类型教学模式；火花思维、小叶子陪练、美术宝等素质教育平台结合课程对图像、声音等内容特征，融入计算机视觉、音频编解码等技术，提供低延迟的强互动直播课程；疫情期间，以晓黑板、伯索云、超星为代表的教育信息化提供商，在各地教育局的指导下快速完成空中课堂搭建，为“停课不停学”提供了重要的技术保障。

（三）实时互动与 IoT：高效连接，实时协同

2020 年，全球物联网设备连接数正式超过非物联网设备连接数，物联网产业发展进入新阶段。2021 年，工信部、网信办等 8 部门联合发布《物联网新型基础设施建设三年行动计划（2021-2023）》，明确提出到 2023 年底，在国内主要城市初步建成物联网新型基础设施。据 GSMA Intelligence 预测，2025 年全球物联网设备连接量将由 2019 年的 120 亿增长到 246 亿，智能家居和智能建筑将分别引领消费级物联网和企业级物联网连接量的增长，人与物、物与物之间的各种信令、即时消息以及音视频形式的实时互动通信都将创造新的互联网通信需求。

IoT 对于低能耗、抗弱网、小包体、跨平台、强安全的要求性更高，意味着研发过程中需要从底层芯片到云端服务进行全链路适配

与优化。以中移物联网 OneOS 为代表的物联网轻量级操作系统，向上承接业务应用，向下屏蔽硬件差异、提供丰富的组件能力，更好的牵引行业创新，降低技术门槛。

IoT 技术厂商通过与实时互动网络服务提供商的合作，面向具体场景进行持续优化，让实时互动功能随着可穿戴设备无处不在。HTC Vive、Pico 等 XR 设备则强调超低延时、超高清清晰度及实时渲染，对网络传输质量提出更高要求。以小天才手表为代表的智能手表，强调超长的续航能力与稳定的高清通话功能，这要求对低能耗、抗弱网方案进行持续优化。

(四) 实时互动与制造：远程作业，保障安全

2021 年工信部等多部门印发《“十四五”智能制造发展规划》，强调以数据为基础，构建虚实融合、知识驱动、动态优化、安全高效的智能制造系统。实时互动技术从实时通信、远程控制等方面为实体经济数字化、智能化升级转型保驾护航。传统制造行业对于抗弱网及低延时有着极高要求，毫秒之差都可能引发巨大的生产事故。

传统行业的生产作业环境常往往缺少完善的网络基础设施支持，实时通信异常困难。通过解决弱网传输核心问题，大量石油、矿山通过公有/专用 IP 网络完成内部实时对讲系统的部署，通过专用智能通信设备整合对讲机和视频录制，在保障实时通话的同时完成远程视频的实时回传，为开采现场的安全提供保障。

矿山、港口等行业作业环境辛苦、作业时间长，常年面临招工难的问题。三一智矿的无人矿车通过远程操作系统，去除本地安全

员，实现车上无人化。同时引入远程介入作为应急场景的处理方式，完成了全场景闭环的作业流程，不仅保障了作业时间，也降低作业危险、提高作业效率。

(五) 实时互动与交通：远程处理，提高效率

2021 年 9 月，交通部印发《交通运输领域新型基础设施建设行动方案（2021-2025）》，强调推动交通运输领域新型基础设施建设。随着人们生活水平的提高，全国汽车保有量进一步提高，根据乘联会数据，截至 2021 年底，全国汽车保有量超过 3 亿辆。随之而来的是汽车周边配套服务的飞速发展，其中车险视频见证定损、可视化紧急救援、远程自动泊车等业务是实时互动技术在交通汽车领域落地的典型应用场景。

在车险视频见证定损中，实时互动技术提供了低延时、高质量的双向视频通话能力和实时回传的双录功能，让车险公司的勘查人员在车辆事故发生的第一时间，通过远程视频获得必要证据并进行记录，免去前往现场的时间成本，提升定损的处理效率与经济性。在远程自动泊车场景中，通过采集传输的高清实时视频，多信号实时反馈车身和周边情况，超低延时进行信令传输，减少车辆泊车过程中的剐蹭和事故，保障财产安全。

(六) 实时互动与金融：远程服务，便捷高效

银行业核心业务的办理方式主要是线下操作，疫情期间，小微企业难以办理金融服务恢复生产，居民线下办理业务繁琐受阻，银行职员缺少远程办公条件影响事物处理，数字化转型迫在眉睫。

搭载实时音视频及 AI 技术的虚拟营业厅、视频银行成为最佳实践形式。在符合金融监管情况下，通过实时双录、人脸识别、身份验证、电子存证等技术，完成业务的线上化、视频化，高效准确可追溯地完成金融服务的可达性、及时性、便捷性、互动性，让客户足不出户就可以享受“面对面、有温度”的金融服务。

(七) 实时互动与房产：远程看房，促进交易

实时互动技术的发展改变了传统地产售房商的销售模式，通过将部分看房流程转移至线上，使买家、租客能够完成沉浸式的远程看房，提高沟通及交易效率。

远程在线看房的出现，大大节约了用户在线下看房中因路途奔波造成的时间成本。借助 VR 技术，贝壳找房实现了虚拟场景内的交互应用，为消费者提供更多的决策信息及沉浸式看房体验，推动产业升级。云渲染技术的应用，将查看房屋渲染效果时的算力需求放在了云端，用户随时随地通过轻量化的小程序进行接入，降低参与门槛。同时，超高清、低延时的实时通话，保障了用户与房主、中介之间的顺畅交流，更好促进买卖、租赁交易的达成。

(八) 实时互动与医疗：普及资源，实时介入

我国人口占世界人口的 22%，但医疗卫生资源仅占世界的 2%，医疗服务供给不足。我国的医疗资源近 80%集中在城镇，农村医疗卫生资源严重不足，设备少、服务能力低，部分偏远地区缺医少药的现状仍存在。面对我国医疗资源短缺，区域发展不平衡的问题，大力发展以远程医疗为代表的智慧医疗服务将是解决目前医疗资源

分布不平衡的有效方法。

远程医疗，包括远程超声、远程会诊、远程手术、远程教学等一系列场景，通过将远端病人侧的视频画面传输到医生端，由医生做出诊断决策和治疗，能够突破原有的地域限制，助力优质医疗资源的均衡配置。相对于普通视频问诊，远程超声检查对远程会诊系统、实时影像传输、实时音视频互动提出了更高的要求，不仅需要满足超声实时影像、视频流、音频流的三路无损高清传输，还需提供动态+静态影像采集与实时标注，提供录制回放，完成与电子病历的深度融合。

急救场景中的实时互动实现了急救与院内诊疗的无缝连接。以重庆急视救为例，在用户拨打 120 后，急救调度中心通过实时音视频，为呼救者提供远程自救指导，进行紧急处理。急救医护人员同时可以通过视频快速判断现场病情、人数及性别，缩短了 80% 的沟通时间，提升医患沟通效率。此外，急视救也还增加了对聋哑人的急救，通过设置手语解说和字幕功能，首次解决了自电话 120 启用的 30 多年来聋人无法报警的问题。

(九) 实时互动与政务：减少繁琐，加速办理

2019 年 4 月 30 日，《国务院关于在线政府服务的若干规定》正式施行，提出政务服务事项全国统一，让“让数据多跑路，让群众少跑腿”，业务办理“最多跑一次”。实时互动的加入帮助民生政务不断优化其业务流程，通过将大量线下业务、活动的线上实时化、可互动化，省去繁琐不必要的流程细节，提升沟通交流效率，

甚至做到“一次都不跑”，助力社会资源得到更充分的调度及利用。

以黑龙江省医保局为例，2020 年上线首个系统视频远程服务平台，用户只需通过 App 或者小程序向工作人员发起视频通话，通过人脸识别认证，就能快速办理医保。以往线下办理医保业务时的材料、证件遗漏问题也有效缓解，提高业务办理效率的同时，减轻老百姓负担。

（十）实时互动与文旅：远程游览，沉浸体验

2021 年 1 月，文化和旅游部提出要加快推进以数字化、网络化、智能化为特征的智慧旅游。数字文旅主要是基于数字化手段、游戏化思维和云基础设施来重塑文博文旅产业。

实时互动在数字文旅中常见的应用场景有云景区、虚拟人、数字藏品等。云景区将线下纪念品、当地文化特色产品等商品进行 1 比 1 建模复刻，并陈列至虚拟商店。游客游览景区时可进入商店，多角度观摩商品细节，并直接在线下单购买。整个云景区游览体验将具备极高的真实感、趣味性和沉浸感。例如实体博物馆，海马云通过使用数据采集设备及无人取景方式，在不对景区和博物馆造成损害的基础上，实现 0.02 毫米精度的采集效果。虚拟场景搭建在云端，依托成熟和领先的云服务技术，用户可通过电脑、电视、手机、平板等设备进入云端场景，实现景区自由移动和游览，游客可像玩游戏一般在景区自由行动。

数字文旅更加升维的方式是从双向弱互动升级成全媒体实时互动，从线下游览升级成线上虚拟空间与线下现实增强相结合的全新

体验。在云景区中设置超写实虚拟人作为导览员，游客还能够在游览时与同游的游客进行互动。数字藏品技术能够让景区增加更多的运营增长空间。例如，基于 3D 实时云渲染、智能虚拟人、云原生联机、数字藏品等技术和手段，阿里元境的数字互动行业应用方案帮助景区实现跨端跨平台、线上线下联动、人与人/人与物之间的实时互动的沉浸式体验，并基于景区 IP 拓展新的运营空间，吸引更多年轻游客，激发文旅新商业价值。

此外传统的博物馆亦可引入互动游戏，提升游客体验。咪咕和博物馆合作在文博场景的创新中，除了推出云博物馆、提供 AR 讲解，更引入了 AR 历史合拍、AR 探究镜、AR 寻宝等多种虚实结合的实时互动游戏，丰富用户的游览体验。

（十一）实时互动与泛娱乐：创新玩法，改变潮流

多元化、互动化已成为泛娱乐平台迭代演进新趋势，相较于传统单一的电台播客、视频直播与社交平台，新型泛娱乐平台通常融合“短视频+互动游戏+连麦 PK+在线语聊”等多类互动场景，以丰富的内容元素创造多样玩法，提升用户参感，增强用户粘性。

泛娱乐行业的用户范围广、用户体验要求高，大量的场景创新推动新一轮的技术升级。声网与陌陌在打磨连麦 PK 场景的过程中首创了直播旁路推流的服务，将 UDP 协议转换成 RTMP 协议，在直播中需要强互动连麦的部分采纳 UDP，对延时要求不高、需要大规模分发的部分采用 CDN，为整个直播行业提供了参考借鉴。

泛娱乐场景同时将更多传统行业带入实时互动，借助各自资源

禀赋共同进行场景创新。以在线 K 歌为例，玩法从本地录唱到排麦单唱、多人轮唱，到在线合唱，对于实时性和技术难度的要求越来越高。全民 K 歌与腾讯云合作，同时引入正版音乐版权，通过全球部署的 RT-ONE™的音视频通信基础网络和 RT-CUBE™音视频终端引擎，快速构建直播、电台、歌房等多样的业务场景，为用户带来流畅稳定的新型体验。

在云观赛场景中，咪咕视频“云包厢”实现云聚会看比赛，在享受精彩赛事的同时，基于中国移动的 MRTC 实时互动网络，身处各地的朋友可以如面对面般进行赛事评论和聊天，极大提升观赛的娱乐性和沉浸感。在常规实时互动基础上，融合引入 VR/AR、数智人、虚拟演播等技术，创新互动场景，实现全场景沉浸式极致体验。在 2022 年北京冬奥会期间，中国移动 5G 冰雪推广大使谷爱凌的数智分身 Meet Gu “走进”咪咕演播室，运用 AR、虚拟演播、实时互动等技术，实现场内与场外、虚拟与现实的同台互动，带给观众耳目一新的体验。

除了 K 歌、音乐、云包厢外，新型泛娱乐平台通常融合“短视频+互动游戏+连麦 PK+在线语聊”等多类互动场景，例如，火山引擎与抖音共同打磨各种生动有趣的内容社交场景。

(十二) 实时互动与游戏：丰富体验，随时可玩

实时互动的加入让游戏表现更加优异。从可感知的用户侧而言，游戏与实时音视频的融合体现在两方面：一方面，在游戏内增加实时音视频，增加了新的可玩元素内容，提升了游戏的社交属性；另

一方面，在直播内增加游戏内容，增加了用户参与感、提升了直播间活跃度，同时提高了用户粘性及变现效率。例如太空狼人杀等。

此外，基于实时互动技术的云游戏，突破了硬件设备的限制，将用户、开发者与终端算力成功解绑，用户只需配备具备基本编解码和显示性能的互动性配屏终端即可接入云游戏服务，使得在手机上玩 PC 游戏、在 Mac 端运行 Windows 游戏、在智能大屏上玩 PC 游戏等跨端跨平台体验成为可能。云游戏降低了大型游戏的入局门槛，提高了游戏的可玩性与创新性，扩大了游戏的用户规模。

实时互动在云游戏领域的关键要素之一是“流畅的体验”。端到端的延迟是否足够低，可用带宽下画质是否清楚，不同网络条件下是否足够流畅、稳定，共同决定了玩家的游戏体验。而从当前的整体趋势来看，5G、Wi-Fi、固网、编解码等技术都呈现向好态势，但在网络、编码、传输、端侧优化等方面还需要核心技术攻坚的地方。以阿里元境为例，针对 4G、5G 移动网络和 Wi-Fi 下的网络条件进行了带宽侦测、卡顿识别、码率自适应、帧率自适应等优化，并通过分区部署、城市边缘网关部署及就近接入等手段实现全网、全国区域的覆盖，使云游戏的体验始终处于特定网络条件下的较高水位。元境作为《云·原神》的云游戏服务商之一，定位于为游戏厂商提供云游戏的研发、运行、运维、运营的一体化服务，实现了全平台、全终端的游戏云化，并通过打通中心云和边缘云，建立了云游戏规模、性能、弹性领先的基础设施，为云游戏的规模化增长奠定了基础。

咪咕快游云游戏基于中国移动“5G+云+边”算力和网络能力，将游戏与视频结合的方式为玩家带来创新的游戏场景。例如实时互动的游戏直播，玩家可以围观游戏主播，并实现连麦、游戏控制权转移等新的实时互动方式。针对跨端或跨设备场景，解决用户因设备、地点等因素带来的游戏体验限制，在手机、电脑、平板之间实现跨平台跨端游戏内容随意无缝切换。

（十三）实时互动与社会关怀：关爱弱势，提升福祉

实时互动是多类型技术进步与场景创新的产物，润物无声改造日常生活的同时，也在提升社会福祉，为更多视障、听障人群提供便利。

“小艾帮帮”作为纯公益的免费视障服务平台，通过实时的音视频通话，让志愿者或者专业服务人员通过远程线上的实时互动方式连接来帮助视障人士，为视障人士提供生活、出行、工作上必要的指导和帮助，让视障人士更好的融入社会。

《中国听力健康报告（2021）》显示，过度的噪音暴露，正让全球 11 亿年轻人面临听力受损的风险；而随着社会老龄化进程的加速，老年性耳聋问题愈加凸显。2020 年 9 月，腾讯天籁实验室联合腾讯公益慈善基金会、深圳市信息无障碍研究会等机构在国际聋人日发起“天籁行动”，从“技术开放”、“听力筛查”和“康复行动”三方面着手，向公益开发者、设备厂商、NGO 及行业免费开放腾讯天籁音频 AI 技术，为听障人群开创更多元化的支持。

六、产业发展挑战和建议

（一）产业发展面临挑战

实时互动关键技术尚未完全成熟，大量技术思路待验证。实时互动处于产业培育期，很多技术思路还处于试验阶段，没有经历过大规模应用的验证，木桶效应明显，需要底层硬件、基础应用和核心算法等多方面技术共同推进方能实现设想中的诸多场景探索。

交互层面，目前实时互动主流交互设备中，XR 设备在显示分辨率、续航、重量等方面仍不理想，与实时互动所要求的无感佩戴相距甚远。脑机接口技术当前还处于较早期的研究阶段，当前可实现的性能距离目标理想状态仍有较大差距。一方面，人类对大脑工作机制的了解相对有限，手术植入的安全风险有待评估；另一方面，脑机接口技术的适用范围与运作模式尚不明确，存在大量伦理风险

网络层面，完全沉浸感需要高分辨率、低时延的画面传输，这对于网络传输的带宽、延迟提出了新的需求，同时亿级的交互用户规模将对网络传输的可靠性提出更高要求。未来实时互动还将占据大量带宽，或对网络稳定性提出新的要求。

算力层面，基础工艺能力不足，工业母机、高端芯片、基础软硬件、开发平台、基本算法、基础元器件、基础材料等方面的瓶颈依然存在。算力网络面临技术不完备，产业标准和理解不统一等问题，在未来亟待解决。

在互联网以及传统数字化转型的应用中，实时互动场景待充分挖掘。相比于数据库等其他互联网基础设施，实时互动目前的渗透

率仍存在不足，具备巨大发展空间。同时，在已包含实时互动的相关应用中，人均使用时长相对较低，实时互动在互联网的发展仍处于早期阶段，存在巨大发展空间。处于数字化转型中的实时互动应用在未来存在大幅扩张空间，尤其是实体经济的众多线下场景还未能充分的搬到线上，疫情对工业制造、安全生产、医疗、金融等行业的实时互动的大量需求还没有专门的服务企业进行深耕，大量的新场景等待开发。

实时互动产业整体缺乏统一标准，互联互通有待提升。实时互动的底层是通信，而通信技术发展到现在，已经应用到各行各业，对于已经存在的通信系统如何兼容、互通，是现在实时互动的一大挑战。上一个阶段通信基础设施的发展得益于移动互联网的爆发，催生了一大批优秀的互联网原生企业，较早形成了这部分的技术融合。处于数字化转型阶段的工业制造、安全生产、医疗、金融等传统行业，在增加实时互动能力时将面临与已有系统和标准的互通兼容问题，实时互动服务提供商需要进行充分考虑。

产业发展处于早期，法律法规和行业监管尚待完善。实时互动是当前的新兴赛道，用户层面、技术层面、市场层面的相关法律法规和行业标准仍需完善。由于互动方式的创新、技术的突破，会带来一些新的隐私保护、数据安全、合规等方面的问题，需要加强监管及行业自治；行业对于服务到达质量、网络延迟与卡顿等尚未形成统一的系统化标准，因此行业协作、跨企业共建效率有待进一步提升。

（二）产业发展建议

加强专有基础技术的深入研究，提供核心技术支持。终端侧光学成像、储能、空间定位、传感器，网络侧基础通信技术，算力侧芯片、算法、计算架构等众多基础技术方面，需要聚集优势资源，集中攻坚，才能实现实时互动产业的宏伟远景目标。针对有示范性、影响力、权威性的高势能客户需求，提供定制化的产品试用策略，有效引导、释放、扩大对关键产品的消费需求。

开展规模化应用试点，探索新场景和新商业模式。尝试推动扩大实时互动在大众生活和娱乐领域的应用范围，进一步探索和融合在实体经济领域中的应用。在数字文旅、教育、企业协同、智能制造、科技金融、数字政府等具备业务变现能力、产品能力相对成熟、政府示范效应显著的细分行业应用领域，落地一批应用示范。通过行业优秀案例征集、创新应用大赛多维度宣传推广，鼓励一批应用先跑起来。

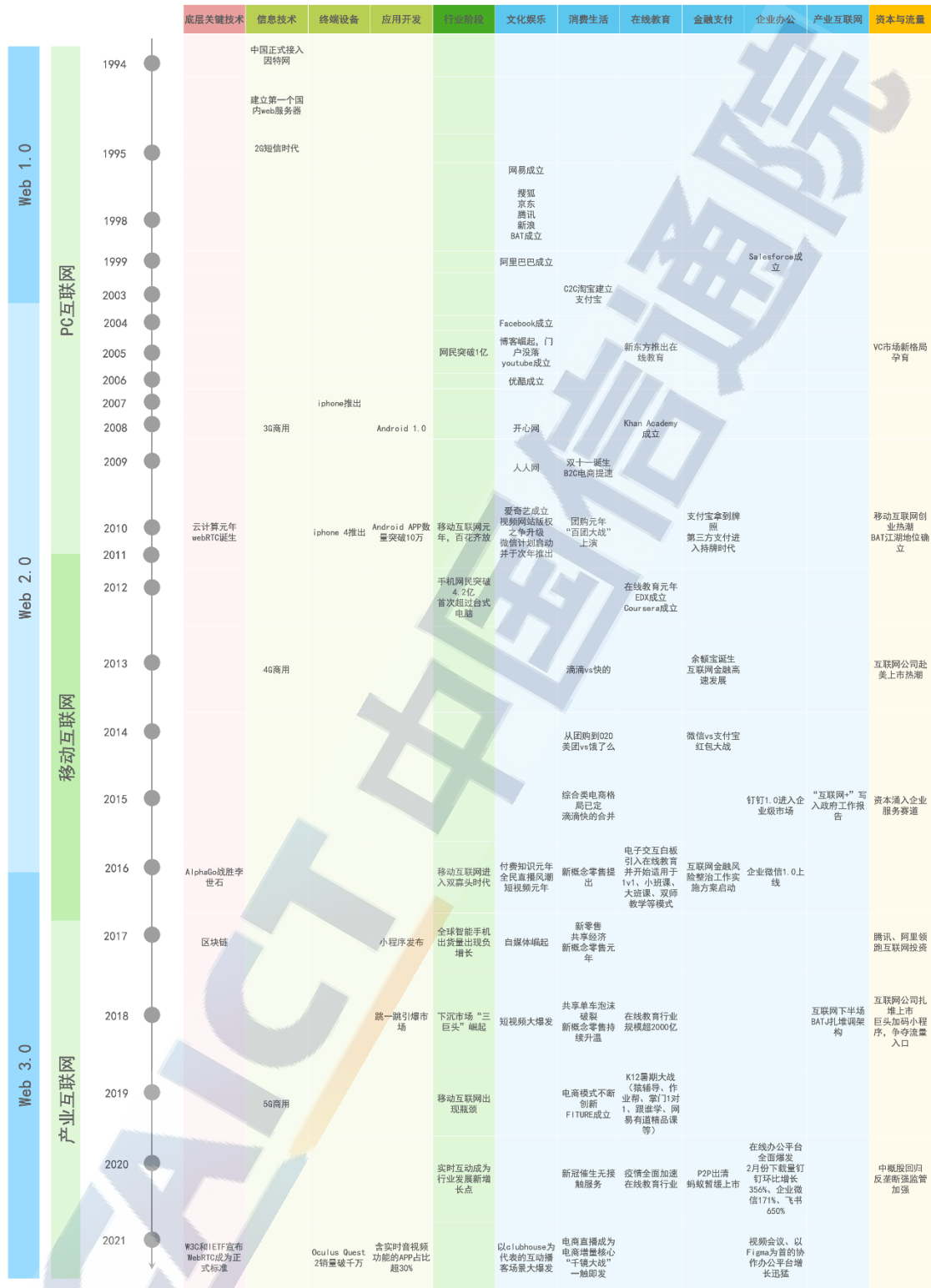
推动产业聚集融合，优化扶持政策。支持具有技术优势的龙头企业联合高校、科研院所组建实时互动应用创新中心和联合实验室，串联产业链不同领域骨干企业组建联盟生态。建立实时互动产业基地和产业聚集发展示范园区，围绕产业链上下游生态的芯片研发、网络传输、算力供应、开发工具、内容制作、应用分发等领域，发展一批面向新兴业态与跨界创新的市场主体，培育和招引一批实时互动细分领域优势企业，推动产业聚集融合。优化实时互动产业扶持政策，鼓励平台生态和产业协作，提高财税政策利用效率，实现

由政策输血向政策造血的转变。

构建公共服务平台，提升产业服务能力。构建实时互动产业链上下游生态体系和基础设施软硬件工程体系，提供面向全行业资源供应、试验验证、投融资对接等公共服务能力。提供面向用户体验、软硬件兼容、功能健全完备、性能和成熟度指标、平台数据安全的产品测评和检测认证服务。充分发挥资本和地方投资对新兴产业和技术的激励作用，鼓励和引导地方加大资源投入力度，通过设立专项基金支持实时互动产业发展与应用推广。建立交叉学科来进行专门的人才培养。鼓励高校和企业支持实时互动领域的专业人才的培养项目、科研项目，积极进行深度创新场景的技术研发，并鼓励产学研结合，争取做到“成熟一批，产业化一批”。

完善法律法规行业标准，推动行业稳健有序发展。出台相关法律法规，保护个人隐私和数据安全。除了支持立法保护数据安全合规、规范竞争行为之外，行业监管部门和行业自治联盟的职责还有如下：（1）制定行业标准，推进行业互助合作；（2）构建信任机制，引导健康有序的商业竞争：成立专门的监管部门，例如行业治理委员会，合理规划引导产业发展，推动开放平台生态建立，提高资源配置效率；（3）保护知识产权：政府有关监管部门应推动知识产权等相关法律法规的完善，保护实时互动领域中信息流动的自主性和产权归属，促进实时互动创业者加速入局。

附录 1：中国互联网发展简史



来源：公开资料整理

附录 2: 缩略语

3GPP	3rd Generation Partnership Project	第三代伙伴计划协议
AIGC	AI-Generated Content	人工智能生产内容
API	Application Programming Interface	应用程序接口
AR	Augmented Reality	增强现实
ARQ	Automatic Repeat-reQuest	自动重传请求
ASR	Automatic Speech Recognition	自动语音识别技术
CDN	Content Delivery Network	视频内容分发网络
CI/CD	Continuous Integration/Continuous Delivery	持续集成/持续交付
Codec	COder-DECoder	编译码器
CPaaS	Communication Platform as a Service	通信平台即服务
FEC	Forward Error Correction	前向纠错
GCC	Google Congestion Control	一种拥塞控制算法
GIPS	Global IP Solutions	一种语音引擎
IoT	Internet of Things	物联网
IRC	Internet Relay Chat	因特网中继聊天
NB-IOT	Narrow Band Internet of Things	窄带物联网
NetEQ	Network Equalizer	音频网络抖动缓冲器
QoS	Quality of Service	服务质量
RTC	Real-time Communication	实时通信
RTMP	Real-time Messaging Protocol	实时传输控制协议
RTN	Real-time Network	实时网络
RTP	Real-time Transport Protocol	实时传输协议

SDK	Software Development Kit	软件开发工具包
SDN	Software Defined Network	软件定义网络
SDRTN	Software Defined Real-time Network	软件定义实时网络
SLA	Service Level Agreement	服务级别协议
SMS	Short Message Service	短信息服务
TTS	Text To Speech	从文本到语音
UDP	User Datagram Protocol	用户数据报协议
UGC	User Generated Content	用户生产内容
VGA	Video Graphics Array	视频图形阵列
VoIP	Voice over Internet Protocol	基于 IP 的语音传输
VR	Virtual Reality	虚拟现实
WebRTC	Web Real-time Communication	网页即时通信

参考文献

- [1]中国数字经济发展白皮书（2021）[R]. 北京：中国信息通信研究院，2022.
- [2]Web3.0 时代：开放、隐私、共建 [R]. 国盛证券研究所，2022.
- [3]探索元宇宙框架，生产力的第三次革命[R]. 天风证券研究所，2021.
- [4]2020 年-2021 年元宇宙发展研究报告[R]. 清华大学新闻与传播学院新媒体研究中心，2021.
- [5]中国视频云市场跟踪（2021 上半年）[R]. 国际数据公司（IDC），2022.
- [6]2021 年全球实时互动云（RTE-PaaS）行业研究报告[R]. CIC 灼识咨询，2022.
- [7]2021 年中国直播电商行业研究报告[R]. 艾瑞咨询，2022.
- [8]中国互联网发展报告(2021)[R]. 北京：中国互联网协会，2021.
- [9]实时互动场景创新生态报告[R]. 声网，2021.
- [10]Zhang S , Lei W , Zhang W , et al. An online learning based path selection for multipath real-time video transmission in overlay network[J]. Transactions on Emerging Telecommunications Technologies, 2020.

中国信息通信研究院 泰尔终端实验室

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62304633-2923

传真：010-62304633

网址：www.caict.ac.cn

