

传媒数据标准化研究报告

（2025 版）

全国数据标准化技术委员会秘书处

二〇二五年八月

序 言

移动互联时代，随着社交媒体、短视频、直播等新兴媒体的强劲发展，传媒数据体量呈几何量级增长。以用户数据为内核，以个性化分发为引擎，推动传媒行业爆发式增长，成为数据要素驱动行业创新发展的成功典范。大模型时代到来后，文本生成、文生图、文生视频、图生视频等 AIGC 内容生成大模型不断涌现，图文、音频、视频等多模态传媒语料数据在大模型学习训练中扮演了更为至关重要的角色，特别是具有中国社会文化背景的各类传媒语料数据，成为当前我国自主大模型研发创新的关键。传媒数据展现出日益显著的巨大价值，成为推动社会文化高质量发展和人工智能技术创新的重要因素之一。

为促进传媒数据标准化工作，针对传媒数据技术迭代快、模式多样化、应用领域广的特点，为奠定传媒数据标准化良好工作基础，经认真研究，于 2024 年 5 月启动《传媒数据标准化研究报告》的编制。由中央广播电视总台组织成立传媒数据标准化研究报告编写组，中国电子技术标准化研究院、中国移动咪咕公司、阿里云计算有限公司、腾讯云计算（北京）有限责任公司、央视国际网络有限公司、北京启明星辰信息安全技术有限公司、中宣部电影数字节目管理中心、江苏省广播电视总台、贵州广播电视台、爱上电视传媒（北京）有限公司、东方有线网络有限公司、

东北大学、深圳广播电影电视集团、中国联通集团、北京百度网讯科技有限公司、北京爱奇艺科技有限公司、国家广播电视总局广播影视信息网络中心、央视频融媒体发展有限公司、中译语通科技股份有限公司、广东南方新媒体股份有限公司、未来电视有限公司、银河互联网电视有限公司、北京明略软件系统有限公司、央广云听文化传媒有限公司等 30 余家产学研单位先后参与相关工作。研究报告在总结传媒行业数据多年来研发、建设和应用经验的基础上，对传媒数据标准体系进行总体规划，为后续传媒数据标准的研究制定提供指南。

2024 年 7 月—9 月，编制组讨论研究传媒数据标准体系总体框架。

2024 年 10 月—11 月，编制组下设 6 个编制小组，分别负责各部分的编制工作。

2024 年 12 月，形成研究报告初稿。

2025 年 1 月，面向业内广泛征求意见，形成研究报告送审稿。

研究报告第一章是引言，主要介绍研究背景、目标及意义。由中国移动咪咕公司牵头，东北大学、中央广播电视总台共同参与编制。

第二章是传媒数据发展应用，主要介绍传媒行业数据技术发展趋势、面临挑战和热点方向。由中国移动咪咕公司牵头，东北大学、中央广播电视总台、央视国际网络有限公司、阿里云计算

有限公司、腾讯云计算（北京）有限责任公司共同参与编制。

第三章是传媒数据标准现状，主要介绍传媒数据相关的国际标准、国家标准和行业标准。由中央广播电视总台牵头，中国电子技术标准化研究院、国家广播电视总局广播影视信息网络中心、上海交通大学负责编制。

第四章是传媒数据标准体系，规划了传媒数据的总体架构、7类标准类别及主要内容。其中：

传媒数据标准体系框架由中央广播电视总台牵头，中国移动咪咕公司、阿里云计算有限公司、腾讯云计算（北京）有限责任公司、央视国际网络有限公司、北京启明星辰信息安全技术有限公司、央视频融媒体发展有限公司共同参与编制；

基础标准、数据资源标准、数据流通标准和技术标准由央视国际网络有限公司牵头，央视频融媒体发展有限公司、腾讯云计算（北京）有限责任公司、中译语通科技股份有限公司、未来电视有限公司、银河互联网电视有限公司、北京明略软件系统有限公司共同参与编制；

评估标准由腾讯云计算（北京）有限责任公司牵头，中国移动咪咕公司、爱上电视传媒（北京）有限公司、北京爱奇艺科技有限公司、贵州广播电视台、阿里云计算有限公司、广东南方新媒体股份有限公司、东方有线网络有限公司共同参与编制；

安全标准由北京启明星辰信息安全技术有限公司牵头，贵州广播电视台、东北大学共同参与编制；

应用标准由阿里云计算有限公司牵头，东方有线网络有限公司、江苏省广播电视总台、深圳广播电影电视集团共同参与编制。

第五章是标准化工作建议，对传媒数据后续的标准化工作提出建议。由中央广播电视总台牵头，爱上电视传媒（北京）有限公司共同参与编制。

《传媒数据标准化研究报告（2025版）》

编制组

梅剑平	李杰	芦丽丽	王为中	李冰	李琳	赵磊
方强	曾亮	杨天识	赵明	潘红梅	张弘	张青
尚峰	王明敏	刘斌	齐勇	赵为纲	胡苏	黄卓伟
单华琦	李忠炤	刘翼光	高天寒	孙鹤立	李向荣	王峰
肖云	褚雷	李凯东	傅力军	闫国卿	李义彪	章国良
汪睿棋	崔菲菲	杨质祺	桂祖宏	胡强	张建军	谢晴平
吴慧慧	刘香云	杨岚	陈珊珊	米庆巍	张晋	吴量
朱承松	黄小勇	张小云	韦梁春	王兵	陈志业	黄昭婷
朱红梅	张超超	官世良	姜昊延	马金韬	张丽	胡春磊
马千里	王亮	许辰铭	邵勇	臧亮	孟峰	王学奎
张宁	张桐瑞	刘增志	潘登	夏蓉	陈森	李凯
王显辉	商常捷	曲明				

编写单位

中央广播电视总台	中国电子技术标准化研究院
中国移动咪咕公司	阿里云计算有限公司
腾讯云计算（北京）有限责任公司	央视国际网络有限公司
北京启明星辰信息安全技术有限公司	国家广播电视总局广播影视信息网络中心
中宣部电影数字节目管理中心	央视频融媒体发展有限公司
爱上电视传媒（北京）有限公司	东北大学
贵州广播电视台	上海交通大学
广东南方新媒体股份有限公司	西安交通大学
银河互联网电视有限公司	未来电视有限公司
中译语通科技股份有限公司	北京爱奇艺科技有限公司
北京明略软件系统有限公司	东方有线网络有限公司
江苏广播电视台	上海广播电视台
央广云听文化传媒有限公司	深圳广播电影电视集团
北京广播电视台	北京百度网讯科技有限公司
中国联合网络通信集团有限公司	北京歌华有线电视网络股份有限公司
北京宏天信业信息技术股份有限公司	东软集团股份有限公司

目 录

一、引言	1
1.1 研究背景	1
1.2 研究目标及意义	3
二、传媒数据发展与应用	4
2.1 传媒数据技术发展	4
2.2 传媒数据行业应用	13
2.3 传媒数据热点方向	23
三、传媒数据标准现状	28
3.1 国际标准化现状	28
3.2 国家标准化现状	31
3.3 行业标准化现状	33
四、传媒数据标准体系	34
4.1 基本原则	34
4.2 指导思想	36
4.3 标准体系框架	37
4.4 标准规划建议	47
五、标准化工作建议	49

一、引言

1.1 研究背景

数字经济时代，数据是重要的生产要素，是国家基础性战略资源。数据作为以数据采集、存储、计算、加工、分析、服务为主的战略性新兴产业，是激活数据要素潜能的关键支撑，和加快经济社会发展质量变革、效率变革、动力变革的重要引擎。面对新一轮科技革命和产业变革深入发展的机遇期，世界各国纷纷出台数据战略，开启数据产业创新发展新赛道，聚力数据要素多重价值挖掘，抢占数据产业发展制高点。党中央、国务院高度重视数据产业发展，推动实施国家数据战略，在《国家“十四五”规划纲要》围绕“打造数字经济新优势”，做出了培育壮大大数据等新兴数字产业的明确部署。2022年中共中央、国务院印发《关于构建数据基础制度 更好发挥数据要素作用的意见》，从数据产权、流通交易、收益分配、安全治理等四方面提出加快构建数据基础制度体系，让数据可确权、可流通、可交易，为促进数据要素赋能经济高质量发展点明实现路径、提供重要指引。2024年国家发展改革委等部门印发《国家数据标准体系建设指南》，遵循数字经济发展规律，以促进数据“供得出、流得动、用得好、保安全”为主线，遵循顶层设计、协同推进，问题导向、务实有效，应用牵引、鼓励创新，立足国内、开放合作的基本原则，建立国家数据标准体系，为推动数据要素高水平应用提供有力支撑。与此同时，传媒行业作为信息传播的重要载体，在数字经济时代

也面临着深刻的变革需求。党的二十届三中全会通过的《中共中央关于进一步全面深化改革、推进中国式现代化的决定》中，特别强调了传媒工作的重要性，提出“构建适应全媒体生产传播工作机制和评价体系，推进主流媒体系统性变革”“构建更有效力的国际传播体系。推进国际传播格局重构，深化主流媒体国际传播机制改革创新，加快构建多渠道、立体式对外传播格局。加快构建中国话语和中国叙事体系，全面提升国际传播效能。建设全球文明倡议践行机制。推动走出去、请进来管理便利化，扩大国际人文交流合作。”这些部署为传媒行业的数字化转型和国际化发展指明了方向，也为数据在传媒领域的深度应用提供了政策支持。

传媒行业与数据紧密相连。以用户数据为内核的个性化分发，开创了短视频、直播等新兴媒体的爆炸式发展。数据为传媒行业提供了全新的视角和技术手段，推动了行业的创新与发展。大模型时代到来后，文本、图像、视频等传媒语料数据成为人工智能发展的重要基础。随着 AIGC 内容生成的不断进步，传媒内容生产和用户体验发生了革命性的变化，同时传媒数据在规模上迅速扩大，数据类型和复杂度也显著增长，业界对数据管理、数据应用、数据流通、数据安全以及隐私保护等提出了更高的要求，迫切需要建立一套科学、合理的标准体系，以指导传媒行业的数据发展，提高数据应用水平。通过制定数据标准，可以保证数据的一致性和可靠性，增强数据安全与隐私保护，促进数据流通与合

作，推动数据高水平应用，为行业发展提供科学决策依据。

2024年7月，中央广播电视总台联合中国电子技术标准化研究院、中国移动咪咕公司、阿里云计算有限公司、腾讯云计算（北京）有限责任公司、江苏省广播电视总台、贵州广播电视台等业内产学研单位，开始《传媒数据标准化研究报告》的研究和编制，旨在洞察数据技术发展的最新趋势以及在传媒行业的应用实践，规划传媒行业数据标准体系，并提出传媒行业数据标准化工作指南和建议。

1.2 研究目标及意义

本研究报告全面分析了传媒数据的技术动态和行业应用，以及国内外数据标准化的发展和现状，并参考《国家数据标准体系建设指南》，综合传媒行业特有的内容、传播、用户等数据的特点，提出了传媒数据标准体系的总体框架和主要标准的建议。

从传媒行业角度看，传媒数据除用户行为、系统日志等结构化数据外，还包括大量视频、音频、图片等非结构化数据。随着大模型技术的不断发展，未来将生成大量AIGC的内容数据，如基于AI生成的微短剧、媒体广告图片生成、新闻媒体文本生成等。本研究报告结合国家数据要素流通的战略规划和人工智能时代传媒行业的发展趋势，从基础、数据资源、数据流通、技术、评估、安全以及应用等关键领域提出传媒行业数据的标准规划，和各项标准的建设内容和适用范围，为开展传媒数据的标准化建设提供依据，助力传媒行业数据业务化、资产化、价值化，促进

传媒数据要素流通，提升传媒行业数据资产的价值。

除了技术和行业发展方面的重要意义，传媒数据标准化工作还具有深远的社会文化价值。从社会文化传承和公众素养提升的角度看，传媒数据标准化工作不仅有助于社会文化的高质量传播与传承，还能通过规范信息传播环境，提升公众信息素养，促进社会文化健康发展。这一工作将为构建更加规范、高效、可信的传媒生态提供坚实基础，推动传媒行业在数字化时代的可持续发展。

二、传媒数据发展与应用

2.1 传媒数据技术发展

2.1.1 发展现状

传媒数据的发展呈现出数据来源多样化、存储计算多元化、分析挖掘智能化、学习训练多模化以及数据安全复杂化等特点，并逐渐从单一数据技术向多模态数据融合与多元化技术方向拓展。随着大模型人工智能的快速发展，多模态语料数据处理成为行业发展的一项关键要素。

1. 数据采集技术

传媒数据采集呈现出多种技术相互结合的趋势。通过人工智能技术与数据采集的结合，可以从图文、视音频中自动识别和提取有价值的信息，数据采集的效率和质量大大提高。网络爬虫技术与自然语言处理技术的结合，能够快速从互联网上抓取大量的文本、图像、视频等数据，并进行初步的筛选和整理，大大提高

了数据采集的效率，节省了人力和时间成本。目前，社交媒体平台依然是重要的数据来源，对社交媒体数据的采集和分析技术不断升级，能够实时监测热点话题、公众情绪和舆论走向，为新闻报道和舆情分析提供有力支持。

此外，传统数据采集技术与传感器技术相融合，使得传媒机构能够从现实场景中直接获取一手数据，例如一些新闻现场，记者可以利用配备传感器的设备快速收集环境信息等多源数据。智能化、可穿戴设备产生的大量数据也正在成为传媒数据的重要组成部分，如智能摄像头拍摄的视频数据、智能穿戴设备记录的用户行为数据等，这些数据为传媒行业提供了全新的数据来源。

在媒体制作、播出分发、传输交换和管理过程中产生的数据也是传媒数据的重要组成部分。这些数据贯穿传媒数据的整个生命周期，具有很高的价值。在内容制作过程中，智能化编辑系统记录创作修改、素材使用及创作者行为数据，优化创作流程并为内容推荐和个性化定制提供依据。播出分发过程中，实时监控系統收集用户观看行为、互动数据及内容传播效果，分析用户偏好和内容表现。传输交换过程中，区块链技术记录数据流转路径，确保透明性和可追溯性。管理过程中，媒体数据管理系统整合多源数据，实现统一存储、分析和应用，支持运营决策。这些数据的挖掘应用提升了内容生产效率和质量，为媒体行业的数字化转型和智能化升级打下了基础。

高性能分布式技术的应用使得数据采集速度大幅提升。一些

先进的数据采集系统能够同时对多个数据源进行快速采集和处理，短时间内可以获取海量的数据。例如，一些新闻机构采用的新型数据采集工具，能够在几分钟内抓取大量的网页新闻和用户评论等信息，为新闻报道的实时性提供了保障。同时，数据采集的自动化程度越来越高，减少了人工干预，降低了错误率和成本。

2. 数据存储技术

随着数据清洗和预处理技术的不断提高，数据存储前自动去除噪声数据、重复数据和无效数据，提高了数据存储的效率和可用性。同时，数据验证和监测技术也得到加强，确保采集到的数据符合质量标准和业务需求。

在存储介质方面，传统的硬盘存储仍然占据重要地位。但固态硬盘的应用不断扩大，凭借高速读写、低能耗、抗震性强等优势，在传媒数据的实时存储和快速处理上发挥着重要作用，如在视频编辑、直播等对数据读写速度要求高的场景中得到广泛应用。同时，光存储技术也在不断发展，因其具备大容量、长寿命、数据安全性高的特点，适合对数据长期保存和归档的需求，被一些传媒机构用于重要的历史资料、素材等备份。

从存储架构来看，分布式存储架构成为主流。传媒行业数据量的爆发式增长，使得集中式存储难以满足需求，分布式存储将数据分散存储在多个节点上，不仅提高了存储容量和扩展性，还增强了数据的可靠性和可用性。一些大型传媒集团搭建了自己的分布式存储系统，以便更好地管理和存储海量的新闻稿件、图片、

视频等数据。而且，这种架构还便于数据的并行处理和分析，为传媒数据的深度挖掘和应用提供了基础。

随着云存储应用的日益普及，越来越多的传媒机构将数据存储于云平台上，借助云服务提供商的专业存储能力和强大计算资源。云存储不仅具备灵活的扩展能力和较低的成本，能够根据传媒业务的需求动态调整存储容量；同时，其便捷的远程访问与数据共享机制进一步促进了传媒行业的协同工作和跨地域合作。

此外，高效的压缩算法可以在不影响数据质量的前提下，大幅减少数据占用的存储空间，降低存储成本，对于传媒数据的存储具有重要意义。同时，智能存储管理系统能够根据数据的访问频率、重要性等因素，自动对数据进行分类存储和管理，提高存储效率。

3. 数据分析与挖掘技术

随着人工智能技术的广泛应用，借助先进的 AI 算法，数据分析系统能够自动识别新闻文本、图像、视频等多种媒体形式中的关键信息、主题和情感倾向。例如，在新闻内容推荐中，基于用户的历史浏览行为和兴趣偏好，利用机器学习算法可以精准地为用户推送符合其口味的新闻内容，大大提高了用户的阅读体验和粘性。同时，深度学习技术在图像和视频数据的分析挖掘上取得了显著突破，能够对复杂的视觉信息进行高效处理和理解，为传媒行业的多媒体内容分析提供强大的技术支持。

自然语言处理技术不断进步，使得对海量文本数据的分析更

加准确和高效。新闻媒体可以快速对大量的新闻稿件、社交媒体评论、用户反馈等文本信息进行语义分析、情感分析和话题聚类，从而更好地把握公众的舆论走向和热点话题。并且，自然语言处理技术还能辅助新闻编辑进行文本创作和编辑，提高内容生产的效率和质量。

大模型到来后，借助其强大的理解能力，能够挖掘出更深层次的内容信息，为新闻报道、舆情监测、舆情引导等提供深度支持。例如，在突发事件发生时，媒体可以通过实时数据分析迅速了解事件的发展动态和公众的反应，帮助制定有效的沟通策略，及时发布准确的新闻报道，有效管理舆情走向，促进正面舆论的形成，并减少负面影响。

4. 数据安全与隐私保护技术

随着同态加密等新型加密算法的发展，数据在加密状态下仍能进行计算和分析，既保障了数据的安全性，又不影响数据的正常使用和处理，为传媒数据在多方共享和协作场景下的安全提供了有力支持。同时，区块链技术也开始应用于数据安全领域，其分布式账本和不可篡改的特性，能够确保数据的真实性和完整性，在传媒行业的版权保护、内容溯源等方面具有广阔的应用前景。

在隐私保护方面，差分隐私、k 匿名等匿名化技术得到了更广泛的关注和应用。这些技术可以在不泄露用户敏感信息的前提下，对数据进行分析和利用，满足了传媒企业在数据挖掘和分析过程中对用户隐私保护的需求。此外，数据脱敏技术也日益成熟，

根据不同的业务场景和需求，采用加密、失真、位置变换等多种方式对数据进行脱敏处理，有效降低了数据泄露的风险。

5. 大模型语料数据构建技术

以传媒数据为支撑的大模型语料数据构建技术正在呈现出多维度、全方位的积极发展态势。智能化数据采集与筛选技术，通过结合自然语言处理和机器学习等先进手段，能够实现对数据的精准获取与高效分类排序，不仅提升了数据采集的质量和效率，也为后续的分析 and 处理提供了更加可靠的基础。这一技术的创新应用极大地增强了数据在多变环境中的适应性，为大规模数据处理和分析任务的顺利进行奠定了基础。

在此基础上，多模态数据融合技术突破了传统单一模态的局限，成功整合了文本、图像、音频、视频等多种形式的的数据，极大地提升了大模型对传媒内容的理解能力。通过多模态数据的深度融合，大模型能够更全面、精准地捕捉和分析复杂信息和知识，推动智能化分析与自动化决策的进一步发展。这一技术的进步不仅促进了对海量数据的高效利用，也为跨领域的智能应用提供了强大的技术支撑。

与此同时，多种数据处理技术的发展提高了数据处理能力。数据增强技术的应用也在不断拓展和深化。通过数据扩充、数据合成等手段，数据增强技术不仅提升了语料的多样性和数量，还显著增强了大模型的泛化能力和鲁棒性。尤其在处理复杂场景或缺乏标注数据的情况下，数据增强技术能够有效弥补数据的不足，

确保大模型在各种环境下都能保持稳定的性能。这为传媒数据领域中的智能算法提供了强大的数据支撑，使得相关应用在面对复杂任务时更加可靠。语料数据标注技术借助预训练模型和多元化的标注方式，标注精准度和可靠性得到了显著提升。通过引入智能标注、半自动标注等创新技术，标注过程更加高效和精准，极大地缩短了数据准备的时间周期，同时保证了标注结果的高质量。这为大规模训练和模型迭代提供了源源不断的高质量数据，推动了大模型在传媒数据领域中的应用落地。

此外，行业合作与共享不断深化，促进了技术交流、资源整合与合作创新，推动了大模型语料数据构建技术的快速发展。同时，数据交易市场的兴起，为语料数据的流通共享提供了新的机遇，促进了数据的价值挖掘和利用。开源语料库也在这一过程中发挥着关键作用，不仅降低了技术门槛，还促进了社区合作和技术创新，为大模型的研究和应用提供了丰富的资源。

2.1.2 面临挑战

传媒数据在数据质量、数据管理、数据安全及法律法规与伦理道德等方面面临诸多挑战。低质量数据和虚假信息对分析准确性和内容可信度构成威胁，数据孤岛限制了数据的高效利用与价值挖掘，而隐私泄露和再识别风险加剧了数据安全管理的复杂性，法律法规的滞后性和伦理问题则为数据利用设置了新的约束。上述问题不仅制约了传媒机构的数据潜能释放，同时也对 AIGC 内容生成和智能分发的健康发展提出了严峻考验。

1. 数据质量

传媒数据的来源广泛，包括社交媒体、新闻网站、短视频平台、广播电视等。不同来源的数据格式、标准和质量参差不齐，存在大量的噪声、重复和缺失值，其中不乏虚假、错误或误导性信息，甚至包含违法违规内容和暴力低俗内容。在发布与转发环节，若 AIGC 依据这些低质量数据生成内容并传播，不仅会影响内容的可信度和媒体的声誉，还可能加剧社会不良风气，助长违法和不道德行为。

在信息爆炸的时代，虚假新闻、谣言等虚假信息在传媒领域屡见不鲜。这些虚假信息将对数据分析结果产生干扰，导致错误的结论和决策。此外，违法违规内容和暴力低俗内容也会污染数据池，进一步扭曲分析结果，影响公众舆论和社会稳定。如何准确识别和过滤虚假信息，确保数据的真实性和可靠性，并有效排除违法违规及暴力低俗内容，是一个亟待解决的问题。

同时在数据采集过程中，可能由于各种原因导致数据缺失，如网络故障、设备故障、用户未完整填写信息等。而大模型分析往往需要大量完整且有质量的数据才能得出准确的结论，数据缺失会影响分析结果的可靠性和完整性，甚至可能导致分析结果出现偏差。

2. 数据管理

传媒机构内部数据往往相互独立，数据无法共享和整合，形成数据孤岛，对数据价值的充分发挥造成严重影响。例如，新闻

编辑部门和广告营销部门的数据可能分别存储在不同的系统中，无法进行有效的协同分析。

为了实现数据可用不可见，确保数据在隐私处理后的完整性也是一个挑战。在传媒数据的传输和存储过程中，数据可能会受到损坏或丢失。当数据被加密并以不可见的形式存储或共享时，验证数据的完整性变得更加复杂。例如，在跨媒体平台的数据共享合作中，对于加密后的新闻素材、广告创意等数据，由于无法直接查看数据内容，很难及时发现其在传输过程中是否被篡改。

3. 数据安全

传媒数据中包含大量的用户个人信息，如浏览记录、兴趣爱好、地理位置等。这些信息如果被不当使用或泄露，将严重侵犯用户的隐私权。随着数据泄露事件的频繁发生，用户对数据隐私的关注度越来越高，传媒机构需要加强数据安全的管理，采取严格的加密、授权等措施，保护用户的隐私。

在 AIGC 模型训练过程中，数据会在不同的计算节点和存储设备之间传输和处理，如果数据加密和访问控制措施不到位，就可能导致数据泄露。此外，模型训练可能会使用到第三方的数据集或工具，如果这些数据集或工具存在安全漏洞，也可能导致数据被泄露给第三方。

大模型训练后的输出可能会导致数据的再识别风险。即使在训练数据中对个人信息进行了匿名化处理，大模型也可能通过其他关联信息重新识别出个人，这种再识别风险在跨媒体数据整合

的场景中尤为突出。比如，当把新闻媒体数据和广告媒体数据结合起来训练大模型时，模型可能会利用广告数据中的设备标识符和新闻数据中的浏览内容，识别出特定用户，并关联其不同媒体平台上的个人信息。

4. 法律法规与伦理道德

与数据技术发展相比，法律法规相对滞后。在传媒领域，对于数据的收集、使用、共享等方面的规定尚不明确，导致一些传媒机构在使用数据技术时存在法律风险。在收集传媒数据作为AIGC训练数据时，可能存在合法性问题。例如，一些数据可能是通过网络爬虫等技术手段获取的，但在抓取过程中可能未遵循网站的使用条款，导致数据收集的合法性受到质疑。

在多源数据融合过程中，传媒数据通常来源于多个渠道，包括用户生成内容（如社交媒体帖子、评论）、媒体机构自身采集的数据（如新闻报道素材、观众收视率数据）以及第三方数据提供商（如市场调研数据）。

随着人工智能的快速发展，AIGC内容生成和智能分发越来越广泛，同时也带来伦理道德方面的尖锐问题，如算法偏见、信息茧房等。算法推荐可能会一味向用户推送某类内容，导致用户陷入信息茧房，无法接触到不同的观点和信息；算法的不透明性也可能导致算法偏见，影响信息的公平性和客观性。

2.2 传媒数据行业应用

2.2.1 发展现状

数据技术在传媒行业的应用日益广泛。在广播电视和网络视听领域，通过分析用户数据来优化节目内容和播出时间，增强节目效果；在社交媒体和短视频领域，通过个性化推荐显著提升了用户触达和体验，创新了信息传播模式；在电影和娱乐领域，不仅优化了电影的营销策略，还提高了制作与市场的精准决策能力。同时，传媒数据正在成为大模型的核心驱动力，依托多模态、高质量的传媒语料数据，AIGC能够自动创造文本、图像、音频、视频等多种媒体内容，从而引发传媒内容生产模式的巨大变革。

1. 广播电视与网络视听领域

中国视听大数据（CVB）系统汇聚了有线电视、IPTV及互联网电视的用户收视数据，并按照相关制度要求积极推进互联网视听收视数据的对接与整合。作为广电行业内覆盖面最广、规模最大、应用最为广泛的收视分析与综合评价平台，它不仅服务于宣传管理，为提升舆论引导能力、推动新时代精品内容创作及优化管理决策提供了坚实的数据支撑，同时还为行业发展赋能，助力电视制作、播出以及广告投放、运营等各类机构在创新实践中实现突破与进步。

2. 社交媒体与短视频领域

在社交媒体领域，数据应用主要体现在用户行为分析、品牌监测以及广告投放优化等方面。例如，通过数据分析工具，可以实时监测社交媒体上关于品牌的讨论和评论，以了解消费者对产品的看法，并识别出用户对品牌的情感倾向。帮助企业及时发现

潜在的品牌危机，并采取相应的措施。此外，社交媒体平台通过分析用户的行为数据，优化广告投放，提高了广告的点击率和转化率。

在短视频领域，则更加注重内容推荐算法的优化和用户体验的提升。例如快手、抖音等短视频平台，通过深度学习和 AI 技术，提高了内容推荐的精准度，增强了用户参与度和互动性。此外，这些技术还为内容创作者和平台提供了精准的数据支持，优化了内容分发机制。

3. 电影和娱乐领域

在电影和娱乐领域，数据技术的应用不仅优化了电影的营销策略，还提高了制作与市场的精准决策能力。在内容推荐方面，通过分析用户的点播记录和搜索历史，娱乐平台能够提供个性化的电影或音乐推荐，增强用户体验。票房预测也得益于数据分析，可以根据导演声誉、演员影响力、内容类型等因素预测电影的票房表现。

在娱乐领域，数据帮助娱乐公司衡量观众对预告片、海报等促销活动的反应，评估不同渠道的传播效果，优化内容和营销策略。此外，数据技术的应用还扩展到了游戏设计和音乐市场分析。通过收集分析玩家模式和使用反馈，调整游戏机制，提高玩家满意度和留存率；或被用来分析听众偏好、流媒体习惯，追踪音乐流行趋势，以优化用户体验。

4. AIGC 内容生成

随着大模型到来，文本、音乐、图像、视频等各类 AIGC 内容不断涌现，吸引了众多关注者。在这一过程中，高质量语料数据作为大模型训练的“燃料”，催生了多模态 AIGC 的快速发展，重要性愈发凸显。以 DeepSeek 为代表的多模态大模型，通过高效的多模态数据处理和生成能力，展现了在文本、图像甚至跨模态内容生成中的卓越表现，进一步推动了 AIGC 技术的普及和应用。

高质量传媒数据是新质生产力的核心要素，也是人工智能发展的关键。数据作为知识与信息的载体，扮演着“原材料”的角色，高质量的数据集意味着更加准确、全面、有效的信息输入，对于大模型的训练、优化至关重要。例如 DeepSeek 在训练过程中，特别注重数据的多样性和质量，通过多源数据融合和精细化标注，显著提升了生成内容的准确性和多样性，为 AIGC 领域树立了新的标杆。

人类智慧需要丰富的知识和经验积累，人工智能系统的水平同样依赖于大量高质量数据的“喂养”，以实现在学习过程中逐渐逼近或达到期望的性能指标。在传媒领域，高质量数据集的应用正在重新定义生产要素，推动经济模式从传统的劳动密集型、资本密集型向数据密集型转变。

数据作为一种新型生产资料，其价值挖掘和利用能力是企业竞争力的核心体现，促进了数字经济的蓬勃发展，催生了数据交易市场、数据服务新业态，为经济增长开辟了新的路径。传媒数

据正在赋能内容生产进入新时代，颠覆人机交互模式。

2.2.2 面临挑战

1. 内容生产能力评价模型

在传媒行业，内容生产能力是核心竞争力所在。无论是广播电视、网络视听，还是社交媒体、新闻媒体、出版等领域，内容就如同灵魂，主导着传媒产品的价值。优秀的内容能引发情感共鸣、传递信息、塑造文化，为行业赢得声誉和经济收益，是吸引受众、维持行业发展的关键。同时，内容生产者作为内容的创造者，其能力水平直接决定了内容质量的高低。内容质量和内容生产者紧密关联，共同构成了传媒行业发展的基石，对整个行业的繁荣至关重要。

当前传媒行业中，内容质量和内容生产者素质两个方面面临着诸多问题和挑战。一方面，内容质量不稳定。在广播电视领域，部分节目为追求高收视率，过度依赖数据热点，导致内容同质化严重，形式与题材重复、创新不足。在网络视听领域，数据推送虽精准，但内容质量良莠不齐，流量至上催生了大量噱头和标题党现象，甚至存在违法违规内容和暴力低俗内容的传播，严重影响用户体验和社会风气。在社交媒体领域，信息过载且虚假信息较多，算法推荐加剧了虚假消息的传播，还可能放大违法违规和暴力低俗内容的影响，对内容质量造成冲击。此外，这些不良内容还可能引发社会负面情绪，破坏网络环境的健康和谐。在新闻媒体领域，新闻生产面临着及时性和准确性的矛盾，部分媒体凭

数据热点仓促发布新闻，导致虚假新闻的出现，影响了新闻质量。同时，违法违规内容和暴力低俗内容也可能混入新闻报道中，进一步损害媒体的公信力和专业形象。在出版行业，数据使得出版内容生产面临迎合市场和传承文化价值的矛盾，大量低质量的跟风书籍充斥市场。

另一方面，内容生产者素质参差不齐。在广播电视领域，部分生产者专业能力不足，编导把控节目节奏欠佳、部分摄像师拍摄的画面存在画质模糊及构图混乱等质量差的问题。创新意识淡薄，模仿成风，导致新节目缺乏吸引力。个别从业者还存在职业操守问题，为追求收视率而操纵节目内容，破坏了行业公正性。在网络视听行业，生产者专业素养差异大，自媒体创作者居多且大多未经专业培训，导致在拍摄、编剧、表演等方面存在不足，作品质量较低，通过夸张的标题和标签吸睛，忽视了内容质量，且版权意识缺乏，创作者随意抄袭改编，侵权行为频发。在社交媒体行业，生产者普遍缺乏信息核实能力，发布信息多靠听说和猜测，不考虑信息真实性。在专业领域讨论中，非专业生产者随意发表意见，易误导用户；且言论责任意识淡薄，不考虑对他人和社会的影响。在新闻媒体行业，生产者质量不一，记者职业素养问题较为突出，如在采访中存在诱导、歪曲等行为，在灾难或犯罪报道中不顾及受害者家属情感。编辑把关不严，未能及时核实新闻稿件数据、事实和语法错误。在新兴领域报道时，记者和编辑专业能力也存在不足。在出版行业，作者水平差异较大，新

手作者语言、错别字、情节构思等方面存在较多问题。编辑质量把控不力，学术著作审核不严。在商业利益驱动下，出版机构降低了门槛，导致作品质量整体下滑。

以上问题严重影响了传媒行业的健康发展，迫切需要建立基于数据的科学化内容生产能力评价模型，规范和引导内容生产。通过对内容生产的准确性、原创性、时效性、深度与广度、观赏性等维度进行评价，规范内容生产流程。同时，对内容生产者的专业素养、创作效率、创作稳定性、合规性、信誉和口碑等维度加以评价，引导内容生产者在数据时代更好地平衡数据导向与内容价值，进而提升整体的内容生产质量，推动传媒行业的健康发展。

2. 传播效果评价模型

在传媒行业，内容传播是其生命线。有效的传播能打破时空枷锁，将内容广泛传播给受众，充分挖掘其价值。广播电视依靠传播将节目呈现给观众；网络视听内容依赖传播获取流量；社交媒体上的信息通过传播成为热点，影响大众认知；新闻媒体通过传播履行职责；出版行业则因书籍传播彰显文化和商业价值。

当前传媒行业在传播效果评价方面面临各种障碍。在广播电视领域，传统收视率调查法存在局限。观众观看习惯多样，抽样收视率数据无法全面反映真实情况，网络平台观众可能被遗漏，也无法衡量观众情感反馈与参与度，广告插播效果评估也变复杂。网络视听行业同样困境重重，内容大量涌现，分散了用户的注意

力，数据分析受隐私政策的限制，平台间数据壁垒使得跨平台观看行为分析碎片化。用户互动行为（如弹幕、评论）的质量和情感倾向分析难度大，影响了传播效果的准确评估。社交媒体平台作为重要传播渠道，评估传播效果也面临难题。数据更新快，信息呈病毒式扩散，但算法变化让内容曝光不确定，热门话题易被取代；用户行为受多种因素影响，点赞和转发不一定代表认可；争议话题的转发可能伴随负面情绪，从而损害品牌形象。新闻媒体领域面临信息过载和受众注意力分散问题。新闻大量涌现，传统指标无法衡量新闻对受众思想和社会价值的影响，且不同平台传播效果差异大，难以整合分析。出版行业在数字化转型中也面临挑战。纸质书和电子书销售数据只是传播效果的一部分，读者的阅读体验、对书籍内容的理解和传播等行为数据难以获取，会员和借阅模式也使得评估更加复杂，不能仅依阅读量来判断书籍影响力。

为推动传媒行业良性发展，依托数据建立全面的传播效果评价模型势在必行。通过制定一系列评价指标，如受众覆盖指标（到达率、曝光量等）、受众参与指标（点击率、播放量、互动率等）、受众忠诚度指标（观看/阅读时长、留存率等）、传播广度指标（地域覆盖范围、年龄层次分布、社会群体渗透等）、传播深度指标（传播层级、内容记忆度等），能够更精准地评价内容在不同平台、不同受众群体中的真实传播效果，为广播电视节目制作、网络视听内容创作、社交媒体运营、新闻报道以及出版策划等提

供有力决策依据,推动传媒行业朝着更高效、更精准的方向发展。

3. 行业知识图谱

传媒行业知识涵盖新闻、广告、娱乐、文化、教育等多个领域,当前各领域拥有相对独立的知识体系,领域之间的联系不够紧密,欠缺有效的知识表示方式来完整地涵盖这些领域的知识。传媒行业的知识数据来源广泛,包括社交媒体平台、新闻机构、市场调研机构等,这些数据分散在不同的系统和平台中,基于商业利益等因素的考量,各平台之间的信息相对封闭,形成了不少信息孤岛,阻碍了知识的整合和共享。此外,传媒知识数据具有丰富的语义信息,机器在处理这些数据时,倘若缺少文化背景知识,极易产生歧义。

利用数据技术构建传媒行业知识图谱,是解决上述问题的有效手段。在内容生产方面,知识图谱可以为传媒行业创作者提供丰富的创作素材和背景信息,提升生产内容的深度。在内容传播方面,知识图谱可以帮助深刻理解用户意图,提供精准的信息检索结果,结合知识图谱中的内容关联关系,为用户提供个性化的内容推荐,提高用户对媒体平台的黏性和满意度。在内容管理整合方面,知识图谱可以将跨平台、多模态、异构的媒资内容进行整合和关联,形成统一知识体系,方便对各种媒资内容进行管理。通过对知识图谱数据的分析和挖掘,还可以发现一些潜在的价值信息和趋势。在舆情监测方面,知识图谱可用于关键舆情实体的提取,及时发现舆情热点,并对舆情进行深度分析,为相关的媒

体机构、企业、政府部门制定舆情应对策略提供指导。

尽管应用数据技术构建传媒行业知识图谱对行业发展意义重大，但当前依然面临着诸多挑战。首先，构建传媒行业知识图谱通常需要融合多平台数据，跨平台异构数据极大增加数据处理的复杂度。其次，不同来源的数据质量参差不齐，构建知识图谱前必须尽量去除噪声，这使得数据清洗和矫正的难度大幅度增加。再者，数据技术应用于传媒知识图谱的过程中，经常涉及用户个人信息和行为数据，数据安全方面的挑战尤为突出，如如何防止数据泄露和滥用等问题。此外，传媒知识图谱需要处理大量多模态数据，图像、音频、视频数据处理本身就需要消耗更大的算力以及占据更大的存储空间，随着数据量的不断增长，为保证快速地构建和更新知识图谱，数据处理和存储效率必将面临巨大挑战。最后，传媒领域信息更新速度极快，传媒行业知识图谱必定规模巨大，如何实现知识图谱的实时更新也是整个行业必须解决的难题。

4. AIGC 数据评估模型

当传媒行业享受 AIGC 带来的红利同时，也必须看到整个行业面临着巨大挑战。首先，AIGC 虽然带来了创作效率提升，但保证生成内容质量的难度也随之加剧。在传媒行业，生产高质量内容至关重要，一旦不良内容得以传播，可能会误导公众，造成恶劣的社会影响。其次，原创性是传媒领域内容核心价值之一，而 AIGC 生产的内容在原创性方面容易存在巨大争议。因为其生

成过程是基于对大量现有数据的学习和模仿，生成的内容可能与现有作品存在相似性，难以清晰界定版权归属，极易引发抄袭和侵权问题风险。此外，传媒行业利用 AIGC 进行个性化内容推荐时，容易导致用户陷入信息茧房，忽视其他不同的观点和信息，进而加剧社会的意见分化和对立，不利于多元文化的交流和传播。最后，保持时效性是传媒行业的关键要素，基于 AIGC 生成内容只有及时更新和发布，才能够满足受众对新鲜信息的需求，优秀的生成内容应该及时涵盖当前热点、匹配流行趋势、过滤过期资讯，排除过时内容。

综上所述，应用数据技术建立 AIGC 数据评估模型是应对诸多挑战的有效方法。基于数据采集、预处理、多模态内容理解、相似度匹配等技术，有效利用用户反馈信息，充分发挥实时处理能力，建立 AIGC 数据评估模型，从内容准确性、原创性、多样性、时效性等多个方面对 AIGC 生成内容进行全方位综合评估，才能真正让 AIGC 为整个传媒行业带来巨大变革。

2.3 传媒数据热点方向

传媒数据正处于蓬勃发展的阶段，传媒行业正从“数据驱动”向“智能驱动”过渡，智能化和个性化日趋凸显，AIGC 智能生产、智能分发、数据开放和共享、国际传播应用逐步成为传媒数据的热点方向。

1. AIGC 智能生产

AIGC 人工智能生成内容，正在以前所未有的速度重塑传媒

行业的内容创作与传播生态。AIGC 依托深度学习、自然语言处理（NLP）、计算机视觉等人工智能技术，能够自动化生成高质量的文本、图像、音频、视频甚至多模态内容，同时支持跨模态创作，例如将新闻文本自动转化为视频或音频内容，彻底改变了传统内容生产的方式、流程和效率。在新闻领域，AIGC 可以实时抓取和分析海量数据，自动生成新闻稿件，快速响应热点事件；在节目制作中，AI 可以辅助节目创作，甚至生成虚拟演员和场景，大幅降低制作成本；在广告营销中，AIGC 能够根据用户画像和行为数据，动态生成个性化广告内容，提升营销精准度和转化率。

此外，AIGC 与数据分析的深度融合，为传媒行业带来了更深层的变革。通过对用户行为、内容表现和市场趋势的多维度数据分析，AIGC 可以精准预测受众偏好，优化内容分发策略，实现从内容创作到传播的全链条智能化。例如，在短视频平台，AI 可以根据用户的观看习惯实时生成推荐内容，增强用户粘性；在广播和电视领域，AIGC 可以结合收视数据，自动调整节目编排和内容形式，提升观众满意度。

更重要的是，AIGC 正在推动传媒行业向“千人千面”的个性化时代迈进。无论是新闻推送、影视推荐，还是广告投放，AIGC 都能根据用户的实时需求生成定制化内容，极大地提升了用户体验。同时，AIGC 还为内容创作者提供了强大的辅助工具，帮助他们突破创意瓶颈，探索更多可能性。

然而，AIGC 的快速发展也带来了新的挑战，如内容真实性、版权归属和伦理问题等。如何在技术创新与行业规范之间找到平衡，将成为传媒行业未来发展的关键课题。总体而言，AIGC 智能生产作为传媒数据的重要应用方向，正在引领行业迈向智能化、个性化和高效化的新阶段，成为推动传媒生态变革的核心驱动力。

2. 智能分发

智能分发技术通过结合数据分析与人工智能算法，正在彻底改变内容传播的方式与效率。智能分发系统能够实时采集和分析用户的行为数据、兴趣偏好以及内容互动情况，构建精准的用户画像，并基于这些数据实现内容的个性化推荐与精准触达。例如，在新闻客户端、短视频平台和流媒体服务中，智能分发技术可以根据用户的阅读习惯、观看时长和互动反馈，动态调整内容推送策略，确保用户接收到最相关、最感兴趣的信息，从而显著提升用户参与度和留存率。

此外，智能分发还优化了内容传播的时空效率。通过对传播路径和效果的数据分析，系统能够预测内容的传播趋势，选择最佳的分发渠道和时间节点，最大化内容的覆盖范围和影响力。例如，在热点事件发生时，智能分发系统可以快速识别潜在的高关注度内容，并通过多渠道协同推送，实现信息的快速扩散。

智能分发不仅提升了用户体验，也为内容创作者和平台运营者提供了科学的数据支持。通过分析分发效果，创作者可以更好地了解受众需求，优化内容生产；平台则可以通过数据驱动的分

发策略，提高运营效率和商业变现能力。然而，智能分发也面临着数据隐私、算法偏见和信息茧房等挑战，如何在技术创新的同时保障用户权益和信息多样性，是行业需要持续探索的问题。总体而言，智能分发作为传媒数据的重要应用方向，正在推动内容传播向更智能、更精准的方向发展，成为传媒行业数字化转型的核心动力之一。

3. 数据开放和共享

数据开放与共享正在成为传媒行业数字化转型的重要推动力。随着数据技术的快速发展，传媒机构积累了海量的用户行为数据、内容表现数据以及市场趋势数据。通过开放和共享这些数据，传媒行业能够打破信息孤岛，促进跨平台、跨领域的协同创新。例如，新闻机构可以通过开放数据接口，与科研机构、政府部门合作，挖掘数据背后的社会价值，推动公共决策的科学化；广告平台可以通过数据共享，实现跨渠道的用户画像整合，提升广告投放的精准度和效果。

此外，数据开放与共享还为内容创作者提供了丰富的素材和灵感来源。通过接入开放数据平台，创作者可以获取实时的热点话题、用户兴趣趋势以及市场反馈，从而优化内容创作策略，提升作品的传播力和影响力。例如，在纪录片制作中，创作者可以利用开放的公共数据资源，增强内容的真实性和权威性；在影视剧本创作中，数据分析可以帮助创作者捕捉观众的情感共鸣点，提升作品的吸引力。

然而，数据开放与共享也面临着数据安全、隐私保护和版权归属等挑战。如何在保障数据安全的前提下实现数据的最大化利用，是传媒行业需要解决的关键问题。总体而言，数据开放与共享作为传媒数据的重要方向，正在推动行业向更加开放、协作和创新的方向发展，成为构建智慧传媒生态的重要基石。

4. 国际传播应用

国际传播应用是传媒数据在全球范围内发挥影响力的重要领域。随着全球化进程的加速，传媒机构需要通过数据技术实现跨文化、跨语言的内容传播，以提升国际传播的精准度和有效性。例如，通过自然语言处理（NLP）和机器翻译技术，传媒机构可以实时将新闻内容转化为多种语言，满足不同地区受众的需求；通过情感分析和舆情监测技术，可以精准捕捉国际受众对特定事件的态度和反应，从而调整传播策略，增强传播效果。

此外，数据技术还为国际传播提供了科学的数据支持。通过对全球用户行为数据的分析，传媒机构可以了解不同文化背景下的受众偏好，优化内容生产和分发策略。例如，在影视作品出口中，数据分析可以帮助制作方了解目标市场的文化特点和观众口味，从而进行本地化改编；在国际新闻报道中，数据驱动的传播策略可以帮助媒体快速响应全球热点事件，提升国际话语权和影响力。

然而，国际传播应用也面临着文化差异、数据主权和技术壁垒等挑战。如何在尊重文化多样性的同时实现高效传播，是传媒

行业需要持续探索的问题。总体而言，国际传播应用作为传媒数据的重要方向，正在推动全球信息流动向更加智能化、精准化的方向发展，成为提升国家软实力和全球影响力的重要工具。

三、传媒数据标准现状

3.1 国际标准化现状

传媒数据的国际标准化工作涉及多方面内容，包括数据描述、数据交换、数据存储、数据安全以及数据分析等关键领域。主要涉及以下国际标准化组织：

1) 国际标准化组织/国际电工委员会（ISO/IEC）

国际标准化组织（ISO）和国际电工委员会（IEC）都是国际标准的制定机构，ISO 负责制定各行业的国际标准，促进全球贸易和技术交流，而 IEC 专注于电工和电子技术标准的制定。通过合作设立的信息技术联合技术委员会（ISO/IEC JTC 1），ISO 与 IEC 共同制定了涵盖媒体相关的数据格式、数据管理与数据交换等方面的标准。其中，SC29（多媒体编码分委员会）专注于多媒体数据的描述和管理，制定了如多媒体内容描述接口和 MPEG-21 多媒体框架等标准，为媒体数据的互操作性和智能化应用提供了技术支持。SC32（数据管理和交换分委员会）负责本地和分布式信息系统环境内和它们之间的数据管理标准的研制，为跨特定行业领域协调数据管理能力提供支持性技术。SC42（人工智能分委员会）则关注数据治理、参考架构和隐私保护，为传媒行业数据的处理和应用提供了框架支持，详细情况见表 1。

表 1 ISO/IEC 传媒数据相关标准

序号	标准编号和名称	简介
1	ISO/IEC 15938 《信息技术 多媒体内容描述接口》	该标准对多媒体内容进行描述，使得多媒体数据能够更有效地被检索、管理和交互。
2	ISO/IEC 20546 《信息技术 大数据 术语和定义》	该标准提供了大数据相关的术语和定义，以及大数据领域的概念性概述，包括其与其他技术领域和标准工作的关系等。
3	ISO/IEC 20547 《信息技术 大数据参考架构》	该标准为大数据系统的架构设计、组件之间的交互以及大数据解决方案的实现提供了一个通用的框架和指导原则。
4	ISO/IEC 21000 《信息技术 多媒体框架（MPEG-21）》	该系列标准为多媒体内容的创建、分发、消费和保护等方面提供一个综合性的框架，以实现多媒体资源在不同设备和网络之间的有效管理和交互。
5	ISO/IEC 29100 《信息技术 安全技术 隐私框架》	该标准提供了一个通用的隐私框架，用于帮助组织在处理个人信息时，保护个人隐私。
6	ISO/IEC 38505 《信息技术 IT 治理 数据治理》	该系列标准数据治理管理体系标准，为数据治理提供了实施方法、框架和方向。
7	ISO/IEC 5259 《人工智能 分析和机器学习的数据质量》	该系列标准提供通用的方法和指南，以确保分析和机器学习训练和评估的数据质量，进而提高人工智能系统生成内容的质量和可信度。

2) 国际电信联盟（ITU）

国际电信联盟（ITU）是联合国的一个专门机构，致力于信息与通信技术（ICT）的全球标准化。ITU 在传媒数据领域重点关注媒体数据的传输、存储、处理与分析技术。ITU-T（电信标准化部门）制定了一系列涉及数据的标准，定义了数据生态系统的功能框架和应用场景，涵盖了媒体数据的采集、存储、处理和分析，详细情况见表 2。

表 2 ITU-T 传媒数据相关标准

序号	标准编号和名称	简介
1	ITU-T Y.3600《大数据 基于云计算的要求及能力》	该标准规定了大数据在云计算环境下的要求与应具备的能力。
2	ITU-T Y.3601《大数据 数据交换的框架和要求》	该标准提供大数据生态系统中的数据交换框架，大数据交换用于交换数据源与数据目的地之间多种不同类型和多种不同格式的数据。
3	ITU-T Y.3602《大数据 数据溯源的功能要求》	该标准描述了大数据溯源的模型和操作，提供了大数据服务提供商用于管理大数据溯源的功能要求。
4	ITU-T Y.3603《大数据 数据目录元数据的要求和概念模型》	该标准描述了元数据的一般概念及其在大数据生态系统中的应用。
5	ITU-T Y.3604《大数据 数据保存的概述和要求》	该标准概要介绍了大数据的保存及从相应使用案例中得出的要求。
6	ITU-T Y.3605《大数据 参考架构》	该标准定义了一个大数据参考架构，该架构可作为大数据标准化的基本参考点，并为大数据的基本概念和原则提供一个总体框架。
7	ITU-T Y.3607《大数据 数据溯源的功能架构》	该标准介绍了大数据溯源的功能架构，根据ITU-T Y.3602 建议书中确定的功能要求和逻辑组件定义了大数据溯源的功能。

3) 万维网联盟 (W3C)

W3C 专注于推动 Web 环境下媒体数据描述与共享相关的标准化工作开展，旨在构建统一、规范的技术准则与框架。其中，资源描述框架 (RDF) 和 Web 本体语言 (OWL) 这两项标准，从语义表达和概念关联的角度出发，为媒体内容进行语义化描述以及建立关联关系奠定了技术基础，有助于提升对媒体内容理解与整合的深度与精准度。上述标准在传媒数据领域已获得了广泛且深入的应用，具体涵盖知识图谱构建、语义搜索以及跨平台数据共享等关键应用场景，在助力传媒数据发挥更大价值方面起到了不可或缺的作用。此外，W3C 持续推动 JSON-LD 等标准的应

用落地，通过优化数据结构、提升传输效率等方式，切实为媒体数据的轻量化描述以及分布式存储提供了稳固且可靠的技术支持，进一步拓展了媒体数据应用的灵活性与适应性，详细情况见表3。

表3 W3C 传媒数据相关标准

序号	名称	介绍
1	资源描述框架 (RDF)	用于描述互联网信息语义、实现信息推理的标准化框架，通过对象-属性-值的三元组来描述资源之间的相互关系，为语义网提供了基础的数据模型。
2	Web本体语言 (OWL)	在RDF基础上发展起来的一种更强大的本体语言，用于明确地表示词汇表中术语的含义以及这些术语之间的关系，可用于构建更复杂的语义模型和知识图谱。
3	JavaScript 对象表示法-关联数据 (JSON-LD)	基于JSON格式，用于在Web上表示和交换关联数据，通过简单的键值对方式来描述数据，将数据与其他数据进行链接和集成，为媒体数据的轻量化描述和分布式存储提供支持。
4	SPARQL查询语言	用于查询和操作RDF数据的标准语言，允许用户从RDF图中检索、更新和操作数据，用于语义网中的数据查询和处理的工具。
5	数据目录词汇表 (DCAT)	用于描述和交换数据目录中的元数据，帮助组织和发现媒体数据资源，定义了数据目录、数据集、数据服务等元数据属性和关系，促进了媒体数据的共享和互操作性。
6	超文本标记语言第五版 (HTML5)	用于创建Web页面的标记语言，规定了网页的结构和内容的语义，可用于在网页中嵌入和展示各种媒体内容，如图片、视频、音频等。
7	层叠样式表第三版 (CSS3)	用于描述网页样式和布局的样式表语言，可与HTML配合使用，对网页中的媒体内容进行样式设计和布局排版，增强媒体内容的展示效果。
8	可扩展标记语言 (XML)	用于标记电子文件使其具有结构性的标记语言，它可以用来标记数据、定义数据类型，是一种允许用户对自己的标记语言进行定义的源语言，在数据存储、数据交换、内容管理等方面有着广泛应用。

3.2 国家标准化现状

我国在数据标准方面已开展了不少工作并取得成果，以全国数据标准化技术委员会、全国信息技术标准化技术委员会和全国网络安全标准化技术委员会为主。目前已经发布了若干数据基础

标准，为传媒数据标准化提供了有力支持，详细情况见表 4。

表 4 数据相关国家标准

序号	标准编号	标准名称
1	GB/T 20093-2013	中文新闻信息分类与代码
2	GB/T 35295-2017	信息技术 大数据 术语
3	GB/T 35589-2017	信息技术 大数据 技术参考模型
4	GB/T 34952-2017	多媒体数据语义描述要求
5	GB/T 34945-2017	信息技术 数据溯源描述模型
6	GB/T 36073-2018	数据管理能力成熟度评估模型
7	GB/T 36343-2018	信息技术 数据交易服务平台 交易数据描述
8	GB/T 36344-2018	信息技术 数据质量评价指标
9	GB/T 36345-2018	信息技术 通用数据导入接口规范
10	GB/T 37721-2019	信息技术 大数据分析系统功能要求
11	GB/T 37722-2019	信息技术 大数据存储与处理系统功能要求
12	GB/T 37728-2019	信息技术 数据交易服务平台 通用功能要求
13	GB/T 38633-2020	信息技术 大数据 系统运维和管理功能要求
14	GB/T 38643-2020	信息技术 大数据 分析系统功能测试要求
15	GB/T 38667-2020	信息技术 大数据 数据分类指南
16	GB/T 38672-2020	信息技术 大数据 接口基本要求
17	GB/T 38673-2020	信息技术 大数据 大数据系统基本要求
18	GB/T 38675-2020	信息技术 大数据 计算系统通用要求
19	GB/T 38676-2020	信息技术 大数据 存储与处理系统功能 测试要求
20	GB/T 40685-2021	信息技术服务 数据资产 管理要求
21	GB/T 41818-2022	信息技术 大数据 面向分析的数据存储与检索技术要求
22	GB/T 42129-2022	数据管理能力成熟度评估方法
23	GB/T 42381.62-2023	数据质量 第 62 部分：数据质量管理：组织过程成熟度评估：过程评估相关标准的应用
24	GB/T 42381.63-2023	数据质量 第 63 部分：数据质量管理：过程测量
25	GB/T 42381.120-2023	数据质量 第 120 部分：主数据：特征数据交换：溯源性
26	GB/T 42381.130-2023	数据质量 第 130 部分：主数据：特征数据交换：准确性
27	GB/T 42381.140-2023	数据质量 第 140 部分：主数据：特征数据交换：完整性
28	GB/T 42450-2023	信息技术 大数据 数据资源规划
29	GB/T 44109-2024	信息技术 大数据 数据治理实施指南
30	GB/T 44216-2024	信息技术 大数据 批流融合计算技术要求
31	GB/T 37973-2019	信息安全技术 大数据安全管理指南
32	GB/T 37964-2019	信息安全技术 个人信息去标识化指南
33	GB/T 37988-2019	信息安全技术 数据安全能力成熟度模型

序号	标准编号	标准名称
34	GB/T 37932-2019	信息安全技术 数据交易服务安全要求
35	GB/T 35273-2020	信息安全技术 个人信息安全规范
36	GB/T 39335-2020	信息安全技术 个人信息安全影响评估指南
37	GB/T 42016-2022	信息安全技术 网络音视频服务数据安全要求
38	GB/T 35274-2023	信息安全技术 大数据服务安全能力要求
39	GB/T 20092.1-2022	中文新闻信息内容 第1部分：概念模型
40	GB/T 20092.2-2022	中文新闻信息内容 第2部分：新闻元数据
41	GB/T 20092.3-2022	中文新闻信息内容 第3部分：数据交换的XML格式

随着全国数据标准化技术委员会的成立，2025年规划拟制定一批与传媒数据有关的数据领域国家标准，详细情况见表5。

表5 第一批数据领域国家标准（拟制定）

序号	标准名称
1	数据 基础术语
2	数据产品 质量评价通用要求
3	数据产品 描述要求
4	数据要素从业人员能力要求
5	公共数据资源登记 实施指南
6	公共数据资源授权运营 监测评估指南
7	城市全域数字化转型 城市数据有效利用水平评估模型
8	数字化转型 区域数字化综合发展水平评价方法
9	高质量数据集 格式要求
10	高质量数据集 分类指南
11	高质量数据集 质量评测规范
12	高质量数据集 数据标注要求
13	数据服务能力评估 第2部分：评估要素
14	数据服务能力评估 第3部分：实施指南
15	数据利用管理技术要求
16	数据匿名化流通实施及评估指南
17	全国一体化算力网 公共传输通道网络传输服务能力基本要求
18	全国一体化算力网 监测调度平台建设指南
19	数据基础设施 参考架构
20	数据基础设施 互联互通基本要求
21	数据基础设施 用户身份管理和接入要求
22	数据基础设施 标识要求
23	数据基础设施 连接器技术要求
24	数据基础设施 数据目录描述要求

3.3 行业标准化现状

为发挥标准在广播电视和网络视听领域数据体系构建中的引领和规范作用，推动广播电视和网络视听高质量创新性发展，国家广播电视总局制定了一系列与传媒数据有关的行业标准，详细情况见表 6。

表 6 广播电视行业传媒数据标准

序号	标准编号	标准名称
1	GY/T 202.1-2004	广播电视音像资料编目规范 第 1 部分：电视资料
2	GY/T 202.2-2016	广播电视音像资料编目规范 第 2 部分：音频资料
3	GY/T 259-2012	下一代广播电视网（NGB）视频点播系统元数据规范
4	GY/T 261-2012	广播电视数字版权管理元数据规范
5	GY/T 301-2016	视频节目对白字幕数据格式规范
6	GD/J 074-2018	电视收视数据元素集规范
7	GD/J 075-2018	电视收视数据交换接口规范
8	GY/T 332-2020	互联网互动视频数据格式规范
9	GY/T 339.1-2020	有线电视网络大数据技术规范 第 1 部分：通用要求
10	GY/T 339.2-2020	有线电视网络大数据技术规范 第 2 部分：平台要求
11	GY/T 339.3-2020	有线电视网络大数据技术规范 第 3 部分：数据规则
12	GY/T 350.1-2021	网络视听收视大数据技术规范 第 1 部分：总体要求
13	GY/T 350.2-2021	网络视听收视大数据技术规范 第 2 部分：数据元素集
14	GY/T 350.3-2021	网络视听收视大数据技术规范 第 3 部分：接口
15	GY/T 351-2021	广播电视和网络视听收视综合评价数据脱敏规则
16	GY/T 360-2022	广播电视和网络视听节目内容标识标签规范
17	GY/T 418-2024	广播电视和网络视听收视大数据清洗规范

四、传媒数据标准体系

4.1 基本原则

根据中国国家标准化管理委员会发布的《标准体系构建原则和要求》（GB/T 13016-2018），构建标准体系是运用系统论指导标准化工作的一种方法。构建标准体系是开展标准体系建设的基础和前提，也是编制标准制、修订规划和计划的依据之一，是在一定范围内包括现有、应有和预计制定标准的蓝图，是一种标准体系模型。

本研究报告在已有国际国内数据标准的基础上，根据传媒行业的实践经验和需求，按照下述原则构建传媒数据标准体系。

1) 全生命周期覆盖

应覆盖数据“采集、生产、传输、存储、处理、交换和销毁”的全生命周期，以指导和规范数据在整个生命周期中的一致性、安全性和可追溯性。

2) 互操作性

应考虑数据在不同平台、系统和设备之间的互操作性，以促进数据共享和高效流通，避免信息孤岛。

3) 开放性与灵活性

应具有充分的开放性和灵活性，以适应数据技术（包括大模型的应用）的快速发展，并与国际标准良好兼容。

4) 安全与隐私保护

应将数据安全和隐私保护作为重要目标，通过数据管理和数据安全等标准，指导规范数据的合法使用。

5) 面向人工智能应用

应充分考虑面向大模型 AI 的标准应用，在标准体系中积极引入多模态语料数据及智能化应用相关标准。通过制定统一的标注、语义和评估标准，助力传媒数据赋能大模型技术的快速发展。

6) 行业协同

应整合行业内外资源，形成跨领域协同机制，联合技术开发、媒体应用和标准化组织等多方力量，共同推动传媒数据标准制定

和实施。

7) 长期演进与持续评估

应随着技术和行业的快速发展持续更新，并定期进行评估和改进，确保标准的适用性。

传媒数据标准体系应综合考虑数据全生命周期覆盖、数据互操作性、开放性与灵活性、数据安全与隐私保护等基本原则，建立兼具国际化视野和本地化适应性的标准框架，为传媒行业的数字化转型和规范发展奠定坚实基础。

4.2 指导思想

本研究报告的标准体系以落实党的二十届三中全会精神，服务传媒行业发展和产业应用为核心目标，结合传媒数据技术发展和业界最佳实践，并充分考虑大模型 AI 发展的最新趋势，提出传媒数据标准体系框架及关键标准的建设内容。

1) 服务传媒行业数字化转型

围绕推动传媒行业数字化转型这一核心目标，规范数据的采集、存储、处理和分析，提升行业的数据利用效率。通过标准化推动建立统一的数据管理和分析框架，为传统媒体向数字化、智能化转型提供技术支撑。

2) 聚焦行业特色与应用需求

为传媒行业的新技术应用和智能化发展提供支持。通过规范化的数据描述和处理框架，为内容语料、个性化推荐、用户画像等智能化业务提供数据基础，促进大模型等前沿技术在传媒行业

的深入应用。

3) 促进数据互联互通与资源共享

致力于打破数据孤岛，推动传媒数据的互联互通与共享利用。通过统一的数据描述标准、交换格式和接口协议，实现跨媒体平台、跨地域的数据协同，提升行业资源的整合能力。

4) 推动社会价值实现

传媒行业承担着价值传播和社会责任，标准体系应体现社会价值导向。通过规范数据的使用和管理，推动数据用于正向传播、文化交流和社会治理，为社会健康发展创造积极价值。

4.3 标准体系框架

根据数据国家标准体系建设的指导思想，结合传媒数据标准化需求及未来发展趋势，从促进传媒行业数据技术发展、开放共享、价值实现和可管可控的目标出发提出数据标准体系框架和结构，根据标准的应用领域、目标对象和内容范围，传媒数据标准体系分为基础、数据资源、数据流通、技术、评估、安全、应用共七大类标准。如图 1 所示。

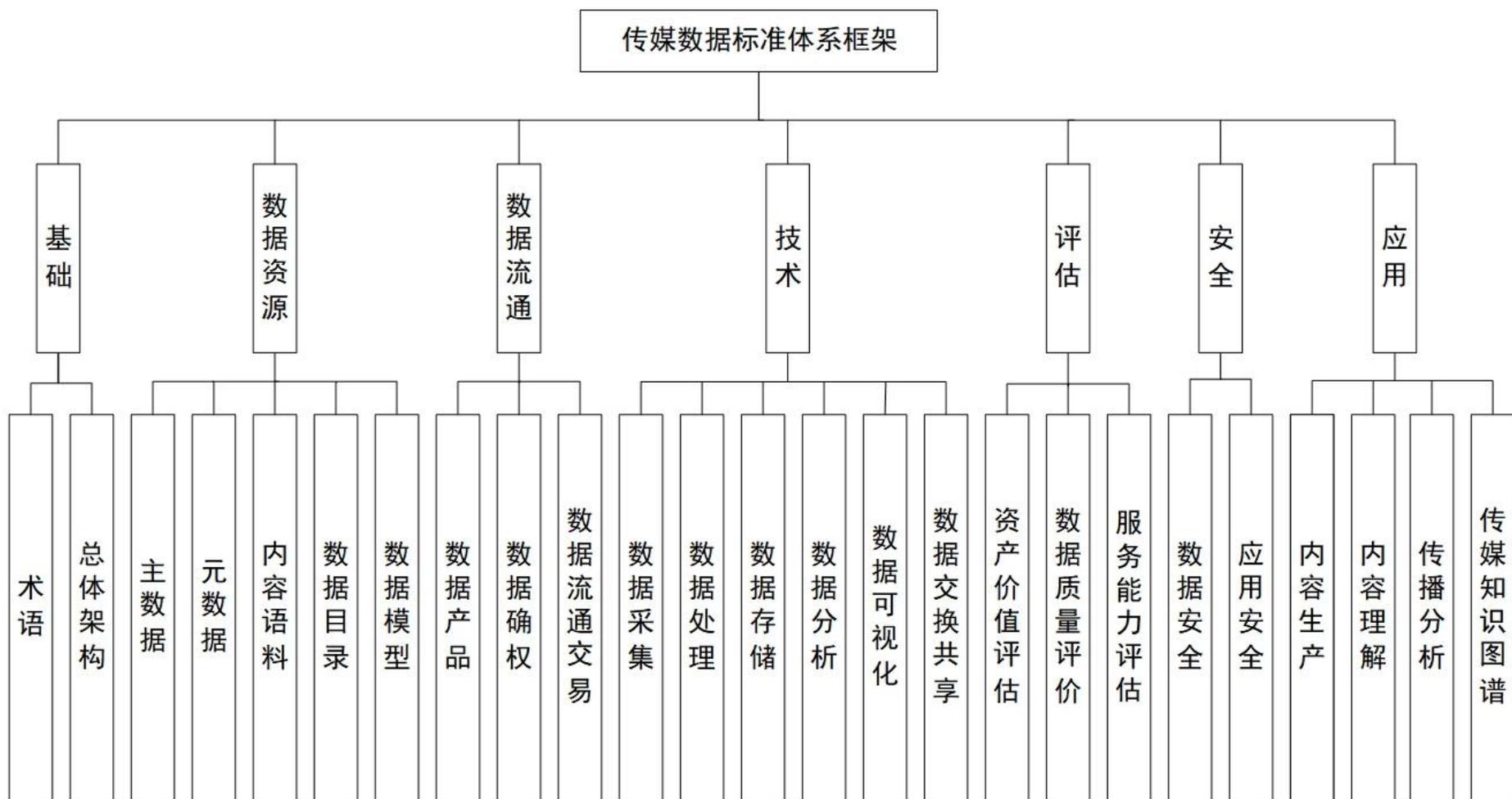


图1 传媒数据标准体系框架

4.3.1 基础标准

基础标准为整个标准体系的共性基础，包括术语及总体架构两种标准，为传媒数据其他部分的标准制定提供基础。

基础标准适用于传媒数据应用的各个领域。

4.3.1.1 术语

术语标准是整个传媒数据标准体系的字典表，明确整个标准体系中使用的术语与缩略语的含义、特征和应用方法，确保其统一性和唯一性。术语标准与国家标准保持一致，对于国家标准已有定义的予以引用，对国家标准未定义的进行规范。该类标准可快速查询定义，确保术语及缩略语的正确使用。

4.3.1.2 总体架构

总体架构标准描述传媒数据的参考架构，包括角色、活动和功能组件之间的相互关系和系统框架，为规划、设计和建设传媒数据标准体系提供指导。

4.3.2 数据资源标准

数据资源标准主要针对传媒数据基础要素进行规范，包括主数据、元数据、内容语料、数据目录、数据模型五类标准。

数据标准适用于传媒数据应用的各个领域。

4.3.2.1 主数据

主数据标准主要规范传媒行业主数据的描述、管理和应用。包括主数据分类、数据标签、管理等标准。

4.3.2.2 元数据

元数据标准规范传媒行业元数据的描述、管理和应用。包括元数据管理、注册、编码要求，以及公共数据字典、基本数据集元数据等标准。

4.3.2.3 内容语料

内容语料标准规范传媒行业大模型学习训练的文本、图像、音频和视频等内容语料，涵盖特征描述、格式要求、标注规则等维度。用于指导和约束内容语料数据的流通和共享。包括内容语料格式、质量要求、标注规则等标准。

4.3.2.4 数据目录

数据目录标准用于指导传媒行业建立统一的数据目录和开展数据目录维护和权限管理。包括数据目录分类标识、编码要求、数据目录维护和权限管理等标准。

4.3.2.5 数据模型

数据模型标准规范传媒数据模型的架构、管理及维护等，为传媒行业数据提供统一的数据模型规范，确保数据结构化及标准化。包括数据模型架构定义、模型管理要求、模型验证与优化等标准。

4.3.3 数据流通标准

数据流通标准旨在规范传媒行业数据资源的合规共享、有序流通，为数据高效流通和数据价值实现提供保障与技术指引，推动数据要素的市场化配置。包括数据产品、数据确权、数据流通交易等标准。

数据流通标准适用于传媒行业数据的共享、交易、流通等环节，为传媒数据要素的市场化配置提供标准化支撑。

4.3.3.1 数据产品

数据产品标准主要聚焦于传媒行业数据资源上升为数据产品的运营、规划及开发。包括数据产品设计规划、数据产品管理、数据产品开发、数据产品应用等标准。

4.3.3.2 数据确权

数据确权标准主要聚焦于传媒行业数据产权的结构性分置方法、数据确权的信息管理、登记程序、平台要求、技术要求等内容。包括数据确权规则、数据确权方法、数据产权流转规范、数据产权登记平台技术要求、数据产权登记指南等标准。

4.3.3.3 数据流通交易

数据流通交易标准主要聚焦于传媒行业数据交易的数据要求、参考架构、交易平台、交易指南等。包括数据流通参考架构、数据交易指南、数据交易平台等标准。

4.3.4 技术标准

技术标准在通用数据技术国家标准的基础上，主要针对传媒数据系统相关技术进行规范，提出传媒数据系统应具备的能力。技术标准包括数据采集、数据处理、数据存储、数据分析、数据可视化、数据交换共享等全流程技术标准。其中数据处理、数据存储、数据分析、数据可视化标准直接引用既有国家数据技术标准，数据采集、数据交换共享领域新增传媒行业特定技术标准。

技术标准适用于传媒数据技术系统的全生命周期，包括设计、开发、建设和运行维护等阶段。

4.3.4.1 数据采集

数据采集标准规范传媒行业数据的采集行为，明确数据获取范围、质量基准及技术实现路径，为后续的数据处理和分析提供高质量的数据基础。包括采集方式与技术要求、采集数据质量控制、采集流程管理等标准。

4.3.4.2 数据交换共享

数据交换共享标准指导和促进不同系统和组织之间的数据开放、共享和交换等，以提升数据利用效率，支持业务创新和协作。本标准包括数据开放要求、数据交换方式等标准。

4.3.5 评估标准

评估标准在已有的数据管理能力成熟度评估标准及传媒数据行业特点的基础上，规范传媒数据全生态体系的数据评估，包括资产价值评估、数据质量评价、服务能力评估等标准。

评估标准适用于传媒数据的规划、管理、应用和流通。

4.3.5.1 资产价值评估

资产价值评估标准规范传媒数据资产的价值评估方法。包括内容数据价值评估、数据创新价值评估、数据利用率评估、数据

变现能力评估等标准。

4.3.5.2 数据质量评价

数据质量评价标准规范传媒数据生命周期各阶段及相关应用的数据质量评价指标和定义，用以提升数据分析、数据服务的准确性和可靠性，提高传媒行业的整体数据质量。包括 AIGC 数据质量评估、数据质量评价框架、评价指标说明及范围等标准。

4.3.5.3 服务能力评估

服务能力评估标准旨在全面规范传媒行业数据服务能力的评估体系，为行业提供科学、系统且可操作性的评估指引。包括数据处理能力评估、数据访问能力评估、数据产品平台能力评估、数据安全保障评估等标准。

4.3.6 安全标准

安全标准是传媒数据安全保障体系的基础，涵盖数据安全和应用安全两个维度，贯穿数据生命周期的各个阶段。

安全标准适用于传媒行业数据的安全设计、系统建设、运行运维和监督管理。

4.3.6.1 数据安全

数据安全标准聚焦传媒行业数据的机密性、完整性与可用性保障，防范数据泄露、篡改、滥用等风险，为传媒行业数据的稳定运行和价值实现提供坚实保障。包括数据分级、数据脱敏规则、内容语料数据安全、数据安全审计等标准。

4.3.6.2 应用安全

应用安全标准主要对传媒数据应用层面的安全问题进行规范，聚焦应用层面的安全风险管控，涵盖业务逻辑安全、接口调用安全等。适用于传媒行业各类数据应用系统，为其开发、运维全生命周期提供安全保障指导。

4.3.7 应用标准

应用标准针对传媒数据的典型应用进行规范，为传媒机构的经营和管理提供更高的数据价值。包括内容生产、内容理解、传播分析、传媒知识图谱等标准。

应用标准适用于传媒数据的内容生产、内容理解、分析评价、运营辅助、决策支持。

4.3.7.1 内容生产

内容生产标准规范了传媒行业利用语料等数据进行内容创

作的全流程，确保内容生产的高效性和质量稳定性，推动内容生产的智能化转型。包括内容质量管控、内容生产能力评价等标准。

4.3.7.2 内容理解

内容理解标准规范了传媒数据（文本、音频、视频、图像）在内容理解类应用中的语义切分、多模态对齐等环节，确保数据能够有效支撑人工智能模型。包括视频场景分割、多模态数据对齐等标准。

4.3.7.3 传播分析

传播分析标准规范了传媒行业利用数据技术进行传播效果评估和传播策略优化的全流程，确保传播活动的精准性、高效性和可衡量性，助力传媒机构提升传播影响力和市场竞争力。包括全媒体传播数据指标、传播效果评估等标准。

4.3.7.4 传媒知识图谱

传媒知识图谱标准规范了传媒行业知识图谱的构建、应用与维护流程，旨在通过结构化的知识表示和关联，提升传媒内容的深度挖掘与智能应用能力，推动传媒行业的智能化发展。该标准涵盖了传媒知识图谱构建、知识抽取及融合、知识查询及推理等

标准。

4.4 标准规划建议

1) 传媒数据分类指南

本标准规定了传媒行业数据的分类原则与分类方法，根据数据业务属性、内容来源等维度建立分类体系。该标准旨在统一行业数据标准，促进数据互联互通，支撑业务协同与安全管控，同时为行业技术演进预留扩展空间。适用于传媒行业全生命周期中数据规划、治理、共享及应用过程中的分类管理。

2) 传媒数据移动互联网信息传播用户标签

本标准规定了移动互联网信息传播用户标签的构建原则、数据来源、标签分类、生成与更新机制、质量评估、安全与隐私保护等方面的要求。适用于移动互联网信息传播领域中用户标签的创建、管理和应用，涵盖各类移动互联网应用（APP）、内容平台等在信息传播过程中对用户标签的使用。

3) 传媒数据元数据

本标准规范了传媒行业数据资源的元数据及标识符，规定了传媒行业元数据的一般要求，确立了具体的元数据表及数据元值

域代码表。包括不限于元数据的中文名称、英文命名、数据类型、约束条件等定义要素。

4) 传媒数据多模态媒体内容语料标注规范

本标准规定了多模态媒体内容语料标注标准的总体要求、标注体系构建原则、语料特征提取与表示方法、标注生成与管理流程等，适用于多模态媒体内容的标注、处理、存储和共享等环节，旨在为多模态媒体内容语料标注的规范化和标准化提供指导，促进不同来源、不同格式的多模态媒体内容语料的整合与互操作。

5) 传媒行业高质量数据集建设指南

本标准提供了传媒行业高质量数据集建设全生命周期的指导和建议，涵盖数据需求、数据规划、数据采集、数据存储与传输、数据预处理、数据标注、模型验证等阶段。适用于传媒行业高质量数据集的规划、建设和维护工作。

6) 传媒数据新媒体数据采集

本标准规定了新媒体平台的数据采集的事件内容、信息属性和适用场景，适用于新媒体平台数据采集的设计、开发、应用和交换。

7) 传媒数据新媒体内容生产能力评价

本标准评估传媒机构的新媒体内容生产能力提供一套系统化的框架，从数据生产能力评价、生产效率评价、系统可靠性与稳定性评价等多个方面定义评价指标体系，准确、客观地评价不同组织的新媒体内容生产能力。

8) 传媒数据全媒体互联网传播数据指标

本标准规范了全媒体互联网传播数据指标的术语定义、分类、计算方式及数据质量要求，涵盖传统媒体与新媒体等多种形式的传播数据衡量与评估方法。适用于媒体、互联网平台、市场研究机构等各类组织在全媒体传播领域进行数据监测、分析和决策。贯穿内容生产、分发、传播、反馈等各环节，关注传播广度、深度及受众互动行为与反馈信息，全面反映传播效果。同时，适用于不同传播场景和业务需求，为多维度分析传播效果提供基础。

五、标准化工作建议

传媒行业数据标准化建设是引领基于互联网、移动互联网的智能传媒高质量发展的战略要素。通过制定传媒行业数据标准，可提高数据质量和一致性，为数据的有效利用奠定基础；促进数

据共享和流通，打破数据壁垒，提升行业整体效率；提升行业创新能力，激发更多创新成果的产生；增强行业竞争力，提升传媒企业的运营管理水平 and 竞争力，推动整个行业的发展；保障数据安全和隐私，建立规范的数据安全和隐私保护标准，降低数据泄露和滥用的风险；提高决策科学性和准确性，为传媒行业的决策提供可靠的数据依据；推动国际交流与合作，有助于与国际传媒行业接轨，促进国内外交流与合作。为了更好地推动传媒行业数据标准化建设，提出以下五点建议：

1. 完善顶层设计，统筹整体工作

传媒数据技术的研究和应用在不断持续发展，传媒数据标准化工作的开展与落实也必然是长期且复杂的系统性工程，需要标准建设单位具备足够的耐力，给予坚持不懈的投入。中央广播电视总台应在《国家数据标准体系建设指南》的指导下，组织相关单位抓紧完成传媒数据标准体系的规划，对传媒数据标准建设的整个过程进行科学合理的统筹安排与规划设计，制定标准总体方案以及阶段性工作任务，以此来引导和推动行业专家、大型传媒企业、研究机构、技术设备供应商、系统开发企业等主体按照明

确的步骤和清晰的目标稳步前行。

2. 坚持关键技术研究，提升行业能力建设

持续深入开展传媒数据关键技术研究，不断深化传媒数据在获取、采集、存储、处理、分析等方面的核心技术创新，为标准化建设铺就坚实而良好的技术基础。数据采集与存储技术、数据处理与分析技术、数据可视化技术、云计算与分布式计算技术、人工智能技术、数据安全与隐私保护技术等将是当前传媒数据的关键技术。根据数据的海量性、多样性、高速性等特征，全面收集各类传媒相关的数据，将分散的、异构的数据进行整合和关联，使其能够相互衔接和融合，以提高数据的一致性和可用性，不断探索和创新，以应对不断变化的传媒环境和数据特点，提高数据采集与获取、集成与贯通的效率和效果。在数据获取与集成工作完成后，需针对各种不同的应用需求，开展数据的处理和管理工

作，选取适宜的数据分析、展示方式，以便达到最大程度的价值创造目标。

3. 统筹国家标准与行业标准建设

传媒数据标准与其他行业数据标准体系都构筑在国家数据

标准基础之上，具备共同的底层技术基础。因此，应在沿用并参考现有技术标准成果的前提下，分析传媒特殊业务要求，合理构建传媒数据标准体系。

4. 加强标准宣贯与应用示范

充分发挥标准的引领作用，形成标准制修订、宣贯应用、咨询服务和执行监督的整体管理体系。依据“试点先行、全面推广”的工作思路，形成应用典型案例并面向社会发布，引起传媒行业上下游共同关注传媒数据标准化建设，共同探索标准的产业应用，以项目为载体加强相关行业、企业交流，深挖潜在商业价值。

5. 构建产业生态体系

以需求为导向，以应用促发展，凝聚传媒行业内优秀的技术、人才、数据等资源，选择具有代表性、创新性且示范效应强的领域，积极开展数据创新应用示范，构建集技术创新与融合、内容创作与传播、用户体验与互动、产业协同与合作、商业模式与盈利、组织架构与人才等多方面于一体的、动态的数据产业生态体系，积极探索传媒行业数据产业新模式、新业态，打造具有广泛适用性和可借鉴性的实践样板。