WORLD ARTIFICIAL INTELLIGENCE CONFERENCE
WAIC | 2019世界人工智能大会

人工智能时代

数字内容治理的机遇与挑战

腾讯安全战略研究中心
Tencent Security and Strategy Research Center

SIGSI
赛博研究院

版权声明

出品方：

腾讯安全战略研究中心

上海赛博网络安全产业创新研究院

总顾问：

马　利　中国互联网发展基金会理事长

谢　呼　腾讯公司副总裁

报告编写组：

韩李云　腾讯安全战略研究中心高级研究员

赵玉现　腾讯安全战略研究中心高级研究员

刘巧霞　腾讯安全战略研究中心高级研究员

唐巧盈　上海赛博网络安全产业创新研究院高级研究员

石英村　上海赛博网络安全产业创新研究院研究员

咨询专家:

惠志斌　上海社会科学院互联网研究中心主任

戴丽娜　上海社会科学院新闻研究所副所长

田　丽　北京大学互联网发展研究中心主任

王　蔚　上海社会科学院新闻研究所新媒体研究中心主任

方师师　上海社会科学院新闻研究所互联网治理研究中心主任

周　斌　腾讯产业安全平台部总监

鞠　奇　腾讯信息安全部技术总监

刘　鑫　腾讯安全管理部高级专家

钟广君　腾讯微信安全风控副总监

曹建峰　腾讯研究院法律研究中心高级研究员

樊晓芳　机器之心产业服务负责人

李　雯　机器之心智慧文娱产业分析师

本报告中技术应用和案例部分得到腾讯公司安全管理部、信息安全部、微信安全风控中心、腾讯安全联合实验室、优图实验室、腾讯安全天御团队、守护者计划安全团队、安全标准团队、网络安全与犯罪基地等多个安全团队的鼎力支持，特别鸣谢赵文俊、王京婕、杨晓光、王永霞、杨晶、左书臣、姚晓波、王翔、郭佳楠、康晓辉等诸位专家、研究员对本报告编写给予的宝贵建议。

目 录

前言

人工智能作为一项具备颠覆性变革潜力的使能技术，引领着新一代信息技术的集合发展，助力传统行业升级，也创造着新的业态。本报告用“人工智能”来命名互联网新启时代，以体现其新科技生产力的本质。在这一时代背景下，蓬勃发展的数字内容产业是前沿科技与创意文化高度融合的产物，天然兼具科技与文化两种属性，是互联网时代人类文化的主要载体和社交传播的内核，甚至凝聚各个共同体的理念价值,将创造新的个体精神世界。人工智能与数字内容发展相辅相成，人工智能促进了数字内容产业迭代演进，数字内容产业也成为人工智能技术发挥其作用的主要场景之一，甚至成为信息技术研发应用的重要导向。

然而，任何新技术的变革必然伴生着风险，出现一些始料未及的问题，但同时也存在新的机遇。本报告重点研判人工智能时代数字内容产业在内容生产、传播、消费、载体、质量、安全等六方面的发展新趋势，剖析新形势下数字内容治理正在或即将面临的内容爆炸、算法弊端、内容造假、版权保护、信息茧房、隐私泄露、数字鸿沟七类风险问题，梳理总结美、欧、英、日、新、中六个主要国家和地区围绕人工智能时代数字内容治理的政策创新。最后，报告从安全应用的视角重点考察国内外业界应用人工智能促进数字内容治理的技术落地与产业实践，相关案例涵盖了内容审核、事实核查、版权保护、打击诈骗、舆情监测、破除信息茧房等多种场景。基于此，报告认为数字内容治理应成为鼓励和推动人工智能技术应用与产业发展的重要领域，并提出前瞻性建议，以期各方能够成功应对挑战，把握人工智能时代数字内容治理的机遇。

1 人工智能时代数字内容产业繁荣发展

1.1 人工智能时代

人工智能（Artificial Intelligence，AI）概念自提出以来，经历了长期而又波折的算法演进和应用检验。直至5G、大数据、云计算、物联网等新一代信息技术的飞速发展，人工智能得到了超强算力、优质算法、海量数据和广泛连接的支持，逐渐演化为融合信息技术、机械、生物等一系列现代科学技术的集成体系，成为推动经济社会发展的新引擎。近年来，美国、中国、欧盟、日本、新加坡、印度等主要国家和地区不断加强战略布局，大力推动人工智能技术在各行业的普及应用。**一个集聚庞大的数据流、信息流、技术流和基于万物互联、跨界融合、人机共生的人工智能时代已经到来。**

1.1.1 新兴技术集合驱动人工智能普及落地

自1950年阿兰·图灵（Alan Turing）预言智能机器的出现并提出判断标准以来，人工智能六十多年的发展和突破多集中于概念和科研领域。2016年，以围棋机器人“AlphaGo”战胜世界顶尖围棋选手李世石为标志，人工智能在产业落地和商业应用上呈现出爆发式发展，广泛应用于汽车、医疗、金融、家居、教育等各垂直行业，人工智能时代正式拉开帷幕。

强大的算力、精进的算法、海量的数据和广泛的连接是人工智能技术转化落地的主要驱动力。

算力层面，云计算架构的广泛部署和硬件芯片水平的提升，极大减少了人工智能模型计算的时间和成本；算法层面，以深度学习为代表的基于大量数据分析和自我训练的机器学习模型日益成熟，成为人工智能最重要的技术方向；数据层面，5G为代表的新一代通讯技术将创造更加海量的异构数据，大幅提升传输速度，大数据技术的不断优化，能够将海量数据充分转化成人工智能可用训练数据的效能；连接层面，物联网极大地拓展了互联网连接广域和深度，重塑现实与网络关系，人工智能技术将更加深刻地影响人类社会。

以数字内容发展为例，基于专家系统（Expert System，ES）的“媒体大脑”等应用，能够对海量异构数据进行高效的预处理和标准化管理，管理内容资源库，充分挖掘内容素材和文化创意；自然语言处理（Natural Language Processing，NLP）和计算机视觉（Computer Vision，CV）能够完成对各种形式和场景下数字内容含义的辨识理解，实现内容的精准匹配和个性化定制；机器学习（Machine Learning，ML）技术在对内容主体进行要素画像和标签设定方面表现突出，使得数字内容能够突破媒介限制，充分延伸到消费者和其他产业。

1.1.2 政策引导促进人工智能产业融合应用

2017年，我国相继发布《新一代人工智能发展规划》《促进新一代人工智能产业发展三年行动计划(2018-2020年)》，制定了到2030年我国人工智能“三步走”的战略目标。2018年，中共中央总书记习近平在中共中央政治局第九次集体学习时进一步强调，人工智能是引领这一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的“头雁”效应[1]。

国际咨询公司埃森哲研究了人工智能在12个发达经济体产生的影响，预测到2035年，人工智能对经济总量增加值的额外贡献最高接近40%，可使年度经济增长率提高一倍，并有潜力将中国的劳动力生产率提升27%，以及推动中国经济预期增长率提升1.6%[2]。

当前，世界各国正在紧锣密鼓地布局其人工智能战略，推进人工智能产业融合应用。据不完全统计，2016年以来各主要国家和地区发布的人工智能战略或政策文件超过40份（见附录表1）。各国政策共性表现为，一是加大人工智能产业政策扶持和创新平台建设力度，竞相争夺全球人工智能市场与人才，加快人工智能的应用化、产业化与融合化，产生显著融合反应的产业之一就是数字内容产业；二是关注人工智能伦理和安全问题，积极探索应对之法。

1.2 人工智能时代数字内容产业发展概况

数字内容（Digital Content），是指以数字形式产生、存储、流通且具备一定意义的信息存在，包括文本、图像、影音等各种载体。由此，**数字内容产业（Digital Content Industry）**就是指以提供数字内容产品或服务的产业门类。目前，国际上尚无数字内容产业的通用定义，我国国民经济分类中亦没有单独划分出数字内容产业，其分散包含在“电信和其他信息传输服务业”“新闻出版业”“广播、电视、电影和音像业”“文化艺术业”等相关行业中。我国于2006年首次在国家文件《国民经济和社会发展第十一个五年规划纲要》中使用“数字内容产业”的概念。2016年12月，国务院颁布了《“十三五”国家战略性新兴产业发展规划》，将数字创意产业纳入五大新型支柱产业。

1 习近平：推动我国新一代人工智能健康发展[EB/OL]. http://www.xinhuanet.com/2018-10/31/c_1123643321.htm.
2 Accenture. artificial intelligence is the future of growth [EB/OL]. https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth.

本报告所探讨的数字内容及其产业，是指依托各类数字化信息技术、数字基础设施和数字化网络平台，向用户提供数字化图像、字符、影像、语音等形式载体的内容信息、产品与服务，并由此形成的新兴产业集群。其核心特点表现为文化内容创意与数字化信息技术高度融合，产品/服务以多种内容载体形式呈现，基于各类数字化平台广泛传播，并不断延伸到其他垂直行业。概念所指既涉及到数字出版、数字广播、数字电视、数字电影、数字音乐、数字动漫、数字游戏等多个细分行业，也关联到数字内容产品与服务的生产（UGC、PGC等）、研发、经营、传输、技术支持等产业综合体的多项复杂环节。

国际广电和媒体技术供应商贸易协会（IABM）最新的用户调查数据显示，全球数字内容产业中仅有8%的企业部署并使用了人工智能技术应用，处于技术应用曲线的起点[3]，但是人工智能对于数字内容产业的影响和促进作用已不容忽视并且前景可观。随着人工智能时代的推进，AI、5G、大数据、云计算、物联网等各类新技术交织融合，新应用突破发展，以全程媒体、全息媒体、全员媒体、全效媒体[4]为代表的全媒体将崛起发展。未来，万物互联、万物皆媒，万物成为数字内容的载体和出口，数字内容产业必将产生巨大变化。

1.2.1 技术持续演进推动数字内容产业业态变革

以阿帕网建立为标志开启的早期互联网时代（1969-20世纪90年代），数字技术以大型计算机为核心，数字内容的主要载体形态有限，电视新闻、电影、广播音频、单机游戏等是这一阶段数字内容产业形态的主要构成。

在PC互联网时代（20世纪90年代-2008年），特别是WEB2.0革新和一系列门户网站、论坛等的建立推动数字内容实现了质的升级和量的拓宽。人们通过门户网站与搜索链接来阅读新闻、欣赏电影、聆听音乐等，依托论坛、博客等开放性网络介质生产用户的原创数字内容。

移动互联网时代（2008年以来），3G、4G信息基础设施迭代升级，Apple Inc、Google等公司推出智能手机与应用市场，为社交网络、手机游戏、网络音视频等数字内容产品的蓬勃发展奠定了良好的软硬件基础，数字内容产业的消费者与生产者边界不在。

3 IABM：The future is artificial: AI adoption in broadcast and media [EB/OL]. https://www.ibc.org/tech-advances/the-future-is-artificial-ai-adoption-in-broadcast-and-media/2549.article.
4 习近平：推动媒体融合向纵深发展 巩固全党全国人民共同思想基础[EB/OL].http://www.xinhuanet.com/politics/2019-01/25/c_1124044208.htm.

随着人工智能时代（2016年以来）的到来，终端不断升级，能提供更加便利的移动网络接入，任何智能设备都有可能成为信源或信息终端，即将迎来万物皆媒的新传播景象。信息的及时性、创新性和多态性在新技术的加持下得到增强，写作机器人、虚拟主播出现在日常新闻内容中，以AR、VR为代表的虚拟现实、增强现实技术成为新兴的内容业态而且新技术应用门槛和成本的降低将极大助推用户生成内容（UGC）的激增，加之云技术泛化和5G商用普及的步伐加快，数字内容的生成和传播触手可及，无处不在，这些又将刺激人们对数字内容精准化和个性化服务的消费热情，最终回归到对人工智能创造崭新精神文明的期许上。

表1 互联网时代与人工智能时代数字内容产业比较

时代	核心技术	典型应用	信息/数字内容主要特点	数字内容产业主要形态
早期互联网时代（1969-20世纪90年代）	1G、主机技术等	电子邮件	信息单向、单点传播，开启人-网连接	电视新闻、广播、单机游戏等
Pc互联网时代（20世纪90年代-2008年	2G、Web技术等	门户网站、搜索引擎	信息单向、多点传播，信息获取渠道拓宽，开启人-网络-人连接	以pc端为主的门户新闻、PC游戏等
移动互联网时代（2008年以来）	3G、4G等	移动客户端、社交媒体	信息多向、多点传播，互联网成为平台，开启人-平台-人连接。	基于移动端的移动新闻、手机游戏、移动音/视频、问答社区、社交应用等
人工智能时代（2016年以来）	人工智能、5G、大数据、云计算、物联网等	AI+场景应用	信息多向、多点、多屏、实时传播，开启万物互联，趋向跨界融合、人机共生。	云+屏幕、万物皆媒下的短视频、AI直播、VR、AR、自动化生成内容等

1.2.2 人工智能时代数字内容产业规模庞大、向垂直领域纵深

人工智能时代，数字内容产业规模获得突破性发展。新兴技术与先进网络基础设施为数字内容应用与服务提供了创新扩容支持。美国标普公司的报告显示，全球数字内容产业的直接总产值在2018年达到1896亿美元，预计到2025年底将达到3438亿美元[5]。我国国家统计局数据显示，2017年，以“互联网+”为主要形式的文化信息传输服务业营业收入7990亿元，增长34.6%[6]。另据我国工信部统计，2017年我国以IP为核心的网络文学、动漫、影视、游戏和音乐等泛娱乐产业约为5484亿元，同比增长32%，占数字经济的比重超过20%[7]。综上可见，数字内容产业规模增速极快，涉及的相关行业众多，是数字经济的核心组成部分，涉及到科技转化、国民就业、产业结构等国家经济重要领域，也是全球贸易中主要的文化产品服务类型。

5 Global Digital Content Market Size, Status and Forecast 2018–2025 [EB/OL]. https://www.wiseguyreports.com/reports/3435933-global-digital-content-market-size-status-and-forecast-2018-2025.
6 国家统计局.2017年全国规模以上文化及相关产业，企业营业收入增长10.8% [EB/OL].http://www.gov.cn/xinwen/2018-01/31/content_5262448.htm.
7 工信部：2018年中国泛娱乐产业白皮书[EB/OL]. https://www.sohu.com/a/225112130_502878.

人工智能时代，数字内容细分领域呈现蓬勃发展态势。互联网用户规模特别是移动用户的快速增长为数字内容产业的发展提供了核心动力，催生了短视频、AI直播、VR、AR等新兴业态，以及基于算法推荐的个性化信息服务商业模式，共同促进数字内容产业繁荣发展。

从用户规模看，我国网民使用率最高的TOP10互联网应用中，80%以上的应用与数字内容产业直接相关，平均用户规模超过6亿[8]。从具体行业看，网络视频、网络游戏已成为数字内容产业中的热点领域；社交媒体继续保有数字内容产业产值和用户数量领先的地位；短视频、网络直播和垂直知识平台增长迅猛，投资火热；数字音乐市场快速增长，全球付费订阅音乐服务的营收也增长到了34.98亿美元[9]；虚拟现实和增强现实（VR/AR）技术改变了数字内容产业的产品形态和传播方式，在新闻、旅游、娱乐等领域纷纷落地应用。

表2 2017.12-2018.12 我国网民使用率最高的TOP10互联网应用[10]

	2017.12			2018.12	
应用	用户规模（万人）	网民使用率	年增长率	用户规模（万人）	网民使用率
即时通讯	72023	93.30%	9.90%	79172	95.60%
搜索引擎	63956	82.80%	6.50%	68132	82.20%
网络新闻	64689	83.80%	4.30%	67473	81.40%
网络视频	57892	75.00%	5.70%	61201	73.90%
网络购物	53332	69.10%	14.40%	61011	73.60%
网上支付	53110	68.80%	13.00%	60040	72.50%
网络音乐	54809	71.00%	5.00%	57560	69.50%
网络游戏	44161	57.20%	9.60%	48384	58.40%
网络文学	37774	48.90%	14.40%	43201	52.10%
网上银行	39911	51.70%	5.20%	41980	50.70%

1.3 人工智能时代数字内容产业新趋势

数字内容产业整体发展趋势体现为纵向延伸、垂直整合、跨界布局、生态融合，行业边界逐渐消弭，数字内容产业不断分化新的社会化分工[11]，孵化出一个新的产业价值体系，在内容生产、传播、消费、载体、质量、安全等各环节或方面都呈现出新的趋势。

8 CNNIC.第43次中国互联网发展报告[EB/OL]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201902/P020190318523029756345.pdf.
9 数据来自市场研究机构MIDiA Research公布的一份调查报告.
10 数据来自CNNIC的《第43次中国互联网发展报告》.
11 叶蓁蓁：人工智能催生“内容科技”[EB/OL].http://media.people.com.cn/n1/2019/0813/c14677-31291269.html?from=timeline&isappinstalled=0.

1.3.1 内容生产：“人-机”协同生产模式将进一步解放生产力

内容生产的决策和流程日趋智能化。数字内容生产正在形成一种互补性的“人机协同”关系，机器将替代信息挖掘、数据检索、媒体资源管理等基础性机械劳动，进一步得到劳动力解放的人类将更能在文化创意革新中发挥更大优势。正如近年来新闻写作机器人正在成为数字新闻智能生产的应用方向。

> 美国《纽约时报》的Blossom、路透社的OpenCalais、英国《卫报》的Open001、新华社的“媒体大脑”都是算法用于新闻生产的典型应用。其中，“媒体大脑”是由新华智云自主研发的国内首个媒体人工智能平台，为媒体机构提供线索发现、素材采集、编辑生产、分发传播、反馈监测等服务，其业务流程囊括线上线下业务环节，记者则在业务各关键节点做出价值判断和决策，形成“人机”协调的智能生产模式。

1.3.2 内容传播：算法精准推荐占据信息流分发主导地位

算法能够实现分类、过滤、搜索、优先、推荐、判定等功能，可根据用户的属性、行为、偏好等个性化特征和大数据画像进行数字内容聚合和精准推荐，快速匹配信息与人，实现“信息找人”“千人千面”。比如，智能机器人可以用于跨语言、跨业态的沟通和搜索信息，情感识别技术可帮助改变僵硬固化的标签设置，使定推更加感性和灵活，等等。根据 CNNIC 发布的《2016年中国互联网新闻市场研究报告》数据显示，算法分发逐渐超越编辑分发，成为网络新闻主要的分发方式[12]。随着人工智能技术的进一步成熟运用，这种模式的主导性地位将愈发明显。

1.3.3 内容消费：沉浸式体验与交互式反馈提供个性化选择

人工智能时代的人们对于精细分工和个性化服务的需求大幅增长，用户将在数字内容消费中扮演更为主动的角色，泛娱乐业与媒体行业势必为满足用户日益增长的多元信息需求而纷纷大显神通。预计未来五年，全球个性化支出平均每年增长4.3%，2023年创造的收入将达到2.6万亿美元[13]。

12 CNNIC.2016 年中国互联网新闻市场研究报告[EB.OL].http://www.199it.com/archives/558868.html.
13 PwC. getting personal−putting the me in entertainment [EB.OL].
https://www.pwc.com/gx/en/entertainment−media/outlook−2019/entertainment−and−media−outlook−perspectives−2019−2023.pdf.

数字内容产业形式将从音频、视频向更具真实性的场景沉浸式体验融合发展。具体来看，内容消费将不仅限于传统的文本、图片、音乐、视频形式，VR、AR、MR、立体影像、智能音响等搭载智能交互功能后形成的沉浸式体验产品/服务，将成为数字内容产业未来融合发展的主要业态，预测将在游戏、社交、影音、教育行业大放异彩。

数字内容消费反馈的实时性随之增强，生产者与用户间信息交互无处不在。交互将成为数字内容产业的基本要素，人工智能文本和音频转换、自动翻译等技术的精进可帮助用户与内容生产者、用户与用户间的互动体验更上层楼。用户将拥有更多机会并具备能力更深层地参与到数字内容的文化创意、内容生产过程中，单向的信息流动链条将转变为充分交互的内容社区。

1.3.4 内容载体：万物皆媒环境中数字内容嵌入各行业场景

新技术普及应用将带来巨大的数据量和广泛的连接终端，人工智能保证了海量异构数据的高效和有效转化。万物互联、万物皆媒的未来，金融、家居、汽车、教育、医疗等各类场景均正在或将成为数字内容新的载体、传播渠道和信息终端，数字内容产业正在发展成与其他各类产业深度融合的横向结合体，未来的数字世界里，“只要有屏，就有数字内容产业”。如AI+直播领域，“克拉克拉”直播平台结合表情驱动算法和原生3D技术在移动端推出“捏脸”直播、虚拟形象定制等服务；而在AI+游戏领域，法国游戏公司Ubisoft不仅开发了智能AI游戏助手Sam，能为玩家在角色互动、游戏攻略等方面提供帮助，还通过人工智能技术在游戏中虚拟建构巴黎圣母院场景，甚至能为其损毁后的现实复原提供支持。此外，人工智能在知识产权保护、数字营销、广告宣传等各领域均有结合应用，极大地拓展了数字内容的外延。

1.3.5 内容质量：“内容为王”成为数字内容产业核心竞争力

用户面对爆炸式增长的海量信息难免无从甄别，各类软文广告、虚假新闻和不良信息不断侵害用户的时间线；数字内容的生产者与用户间的界限日渐模糊为治理带来更多复杂性和不确定性；基于人工智能的智能生产和精准匹配仍存很大提高空间，多种因素导致当下数字内容质量良莠不齐。

在此背景下，人们的消费数字内容的核心需求将升级为对优质内容的深度体验，更具创意的成熟内容以及更为合理的商业模式，成为未来数字内容产业的核心竞争力。正如全球范围内以知乎、Quora为代表的知识性平台正迅猛发展，知识消费也在不断增长，知乎2018上半年商业广告营收额相比去年同期增长340%,注册用户数量年增长达到95.12%，知乎大学的付费人次达到600万[14]，而Quora平台在2017年完成D轮8500万美元融资后，估值达到18亿美金，2018年9月平均每月用户突破3亿[15]。

1.3.6 内容安全：各类治理难题凸显下机遇与挑战长期并存

以人工智能为代表的新兴技术集合不断推动数字内容产业的繁荣发展的同时，也带来了一系列治理挑战。一方面，固有顽疾与新生挑战交织，风险复杂泛化，传统监管模式滞后于产业新业态，各类风险不断冲击着现有法律政策体系和公众道德认知，很大程度制约了数字内容产业的优质发展。另一方面，风险挑战本身亦是安全应用的广阔发展前景，对人工智能技术本身的优化，以及对其在治理各环节应用的积极探索，能够为治理机制的完善提供技术支撑，保障数字内容产业健康高质量发展。

14 数据来源：搜狐科技，http://www.sohu.com/a/245978157_116132.
15 数据来源：ET rise:https://economictimes.indiatimes.com/small-biz/money/quora-raises-85-million-the-qa-platform-is-now-a-unicorn/articleshow/58313275.cms.

2 人工智能时代数字内容治理面临的挑战

人工智能助推数字内容产业颠覆性发展，难免伴随着现有治理模式及监管体系与新业态的不匹配：既存在现行治理手段滞后带来的治理失效和力不从心，也体现为数字内容治理问题与其他社会问题及风险的交织泛化，具体包括内容爆炸、算法风险、内容造假、版权保护、信息茧房、隐私泄露、数字鸿沟等各类技术问题、信息问题和社会问题。

2.1 数字内容极大丰富，内容审核能力不足矛盾凸显

人工智能技术极大地促进数字内容产业的繁荣。据不完全统计，截止2018年，每天约有2.5万亿字节的数据被创建，过去两年里生成的数据占到了全球总数据的90%。预计到 2022年，全球互联网流量将达到每秒 7.2 PB[16]，未来数字内容必将继续呈指数级增长。

> Facebook每天上传的照片超过3亿张，每分钟发布51万条评论，30万条新状态；Instagram每天分享的照片和视频量达到9500万次；《2018微信年度数据报告》显示2018年，微信每月有10.82亿用户保持活跃，每天有450亿次信息发送出去，每天有4.1亿音视频呼叫成功；2017年头条号账号总数超120万，平均每天发布50万条内容，创造内容消费达48亿次。以每个账号每天投稿5条内容保守估计，仅头条每天的投稿内容就高达600万条[17]。

然而，面对庞大的数字内容洪流和各国日趋严格的内容审核政策，试图依靠传统审核模式实现内容含义的准确判断并及时应对信息爆炸引发的各类问题，越发捉襟见肘。内容审核能力不足的问题将会更加凸显，**集中表现为人工审核难以应对海量负面信息和算法审核机制尚不完善。**

内容审核因工作量繁重，工资偏低，海量负面内容轰炸以及高准确率的内容审核要求，对从业人员产生不小的精神以及身体创伤。

> Facebook在印度的内容审查外包公司Genpact合同工，平均每分钟需审核4篇贴文，日审核量2000篇以上。一名前Facebook合同工Selena Scola曾在2018年9月向美国加州法院提起诉讼，指控因长期审查平台可怕图片后而导致精神创伤，却未曾受到公司适当保护[18]。

16 数据来源：http://www.doudouhong.com/news/show/102.html.
17 数据来源：https://www.ofweek.com/ai/2018-04/ART-201717-8120-30220783.html.
18 Reuters:Some Facebook content reviewers in India complain of low pay, high pressure[EB/OL] https://www.reuters.com/article/us-facebook-content-india-feature/some-facebook-content-reviewers-in-india-complain-of-low-pay-high-pressure-idUSKCN1QH15I.

尽管机器审核能够有效缓解人工审核压力，却可能存在因算法审核机制不完善而导致的结果歧视。比如，审核算法样本数据的偏差可能使不同语言内容在同一平台出现不同审核结果。NLP工具通常用于解析英语文本，在解析非英语文本则表现出明显差异，包括流程更长、精度降低或标准苛刻，甚至由此带来文化或地域歧视[19]。

因此，如何建立良好、协调、及时和高效的审核机制，尚需时间探索。其中具体包括人工审核与机器审核的关系、跨国公司审核标准和业务统一规范，内容审核从业者工作环境优化，公司审核成本与从业者工资矛盾等问题。

2.2 “算法偏见”与“算法黑箱”影响数字内容公正性，技术攻关存难度

作为人工智能核心之一的算法，“偏见”和“黑箱”是普遍存在的问题。训练数据不完全、算法本身设计不健全、技术开发或是与人类交互过程中存在偏见观念等问题因素，误导或投射到算法输出结果中，使之带有歧视性，即为**“算法偏见（Algorithm of Prejudice）”。算法在数字内容领域应用风险之一体现为算法推荐的歧视结果导致用户接收不公正、片面加强或偏差的信息。**

2015年，Google因图片搜索识别结果将黑人错误标记为大猩猩而备受批评[20]，其后更被质疑在检索与黑人有关姓名时，结果呈现出较多的犯罪记录检测广告或枪支广告[21]；Google推送招聘广告服务中，男性与女性、白人与黑人的招聘广告也存在职业、薪水上的明显差距[22]。再比如，据非营利组织ProPublica研究透露，亚马逊公司购物推荐系统一直偏袒其自己及合作伙伴的商品，通过隐瞒商品运费来误导消费者，使其在购物比价服务中得到错误的比价结果[23]。

19 Center for Democracy & Technology :Mixed Messages? The Limits of Automated Social Media Content Analysis[EB/OL] https://cdt.org/files/2017/12/FAT-conference-draft-2018.pdf.
20 BBC:Google apologises for Photos app's racist blunder[EB/OL].https://www.bbc.com/news/technology-33347866.
21 BBC:Google searches expose racial bias, says study of names[EB/OL].https://www.bbc.com/news/technology-21322183.
22 The Guardian：Women less likely to be shown ads for high-paid jobs on Google, study shows [EB/OL]. https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study.
23 Propublica:MACHINE BIAS Investigating Algorithmic Injustice[EB/OL]. https://www.propublica.org/series/machine-bias.

“算法黑箱”（Algorithm Black-Box）或算法不透明的产生既有商业机密保护的原因，也有因技术门槛高导致普通用户难以理解算法的情况,还存在由于算法模型依赖的组件、模块和数据库过于复杂，而导致开发人员也很难说清楚每个决策的影响因子和判定依据。**算法在数字内容产业各环节的普遍应用，不可避免的会遭遇算法黑箱风险：**一方面表现为公众难以或无法理解内容生产、推送和传播背后的流程逻辑，无从对推荐内容做出反馈和干预，甚至对推荐内容产生反感、质疑情绪；另一方面，受决策不可解释的局限，政府部门针对算法输出的错误结果归因问责困难，从而采取更加严苛的数字内容监管措施。

“算法偏见”和“算法黑箱”问题是人工智能技术体系内嵌的安全问题，解决其带来的数字内容治理风险，不仅需要考虑算法设计层面的公正性，还需要充分认识到技术的现实局限性，从算法攻关上寻求突破。

2.3 内容造假滋生灰色产业，深度伪造威胁国家社会稳定

人工智能生成内容取得显著进步的同时，事实上也为滥用或恶意利用人工智能进行内容造假提供了温床，以炮制虚假新闻（Fake News）和深度伪造（Deepfake）为典型风险。政治或经济利益驱动下的内容造假，大范围地广泛传播或是针对特定人群的精准推送，将恶化数字内容质量，破坏信息传播秩序，误导用户判断，甚至催生或煽动社会极化情绪,诱发各种社会问题。

恶意利用人工智能技术炮制虚假新闻，威胁新闻真实性和中立性。路透研究院发布的2019年全球新闻行业报告显示，公众对新闻的总体信任程度下降至42%，过半数受访者忧虑自己对网络上真/假新闻的甄别能力，32%的人表示他们因此正在主动拒绝阅读新闻[24]。麻省理工学院传媒实验室针对虚假新闻在Twitter上的传播情况展开长达12年的调查研究，研究表明假新闻往往更具惊奇性和爆炸性，因此传播得更远、更快、更深、更广，传播速度是真实信息的6倍[25]。

2019年，OpenAI为实现“语言建模”任务开发了一种算法，该算法能够根据已有单词预测接下来的文本内容，类似于输入法中的自动补充文本信息，可用于执行翻译、开放式问答任务，检查语法错误，帮助作家寻找创意或者生成对话，为企业或者政府决策者提炼总结性文本。然而，类似于此的人工智能算法，一旦被恶意利用，只需为其提供只言片语的信息，就能编写逼真的虚假新闻，大大降低虚假新闻制作门槛，且逼真度极高。

24 Nic Newman etc:Reuters Institute Digital News Report 2019[EB/OL]. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_1.pdf.
25 Katie Langin etc:Fake news spreads faster than true news on Twitter—thanks to people, not bots[EB/OL]. https://www.sciencemag.org/news/2018/03/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots.

虚假新闻精准投放，人为操纵网络舆论，影响用户观念。美国伊隆大学数据科学家奥尔布赖特（Jonathan Albright）的研究显示，人工智能技术在美国2016年总统大选中的应用几乎达到了操纵局势的级别[26]，美左右翼势力均试图使用一种基于数据驱动算法模型和人工智能技术训练的“微宣传机器网络”（Micro-Propaganda Machine Network）[27]以此影响选举。具体做法包括通过数据心理分析模型和广告定位算法对网民画像，基于政治话题和新闻事件大规模制造具备误导性、有偏见甚至虚假不实的热点信息，再依靠精准推送和分发系统，达到深度影响选民观念的政治目的。

> 2017年1月，美国国家情报总监办公室公布的《评估俄罗斯在近期美国总统大选中的活动与意图》[28]报告披露了在国家机器掌控下的兼具微观瞄准和情绪操纵的政治舆论操控机制：敏感信息披露、虚假不实信息制造、热点设置和舆论引导、精准定向分发等多种方式融合，依托主流媒体、新闻门户网站、第三方转载网站及应用、社交平台、视频直播平台、搜索引擎以及电子邮件组等渠道“组织”全方位宣传轰炸，以达到国家级别的政治目的。

机器人水军刷单评论，滋生灰色产业链，破坏网络社群规则。另外，内容造假亦可利用机器在社交平台建立大量的虚假账号，进行自动化造假，使得虚假新闻生产产业化。据网络安全公司Trend Micro的数据显示，能对大选结果产生影响的虚假新闻全程服务费用在40万美元左右，能煽动民众上街游行的虚假新闻服务费用在20万美元左右。另据美国《纽约时报》2018年1月28日报道，美国Devumi公司专门在Facebook、Twitter等社交网络上制造“机器人水军”帐号，再卖给电视演员、企业家、运动员等“想出名或者想在互联网上施加影响力的人。”

26 Jonathan Albright：Welcome to Fake News[EB/OL].https://medium.com/@d1gi/election2016-fakenews-compilation-455870d04bb.
27 陈慧慧《“人工智能技术操纵美国大选”研究报告的评述分析》[J].信息安全与通信保密，2017（07）：40-47.
28 政策文本链接：https://www.dni.gov/files/documents/ICA_2017_01.pdf.

深度伪造是内容造假在人工智能时代的升级版，其更加逼真且难以辨识的特点引起了各国政府的高度重视。这种基于生成对抗网络（GAN）的深度伪造技术应用在视频制作中，能够“换脸”“换声”和模仿行为举止，未来甚至能虚构环境场景，其与虚假新闻最大区别在于，人们基本无法从内容本身做出真伪判断。据研究称，三名以色列研究人员利用深度伪造技术通过医院放射科的网络篡改甚至生成伪造的MRI/CT图像，对从业几十年以上的专业医生的欺骗率超过95%，意味着他们可以利用深度伪造远程操控病人的生命[29]。这种具有高度欺骗性的技术目前在色情视频中被广泛应用，对诸多公众人物的形象造成损害，比如2017年人工智能“换脸术”曾将《神奇女侠》女主角盖尔·加朵的脸嫁接到一个成人电影女星的身上，从而引起社会轰动。如果该项技术一旦被应用于政治选举或经济竞争，后果不堪设想。

2.4 智能化内容生产权责归属困难，版权保护亟待健全法规体系

人工智能普遍应用于数字内容生产环节，那么，由人工智能技术创造的内容，根据现行法律体系判定其法律权属普遍缺乏依据，在此背景下的数字内容著作权保护政策亟待完善。

人工智能所生产内容的版权问题包括可版权性和版权归属两个问题，其实质是人工智能是否应当以及能够具备明确的法律主体地位，这是人工智能技术应用对现行法律体系革命性冲击的又一表现。

> 2019年4月26日，北京互联网法院对全国首例计算机软件智能生成内容著作权纠纷案(即菲林律所起诉百度网讯公司侵害署名权、保护作品完整权、信息网络传播权纠纷一案)进行了一审宣判，首次对人工智能软件自动生成内容的属性及其权益归属作出司法回应[30]。判决认定计算机软件智能生成的涉案文章内容不构成作品，但同时指出其相关内容亦不能自由使用，百度网讯公司未经许可使用涉案文章内容构成侵权。

目前，人工智能生产的数字内容版权保护仍普遍面临法律滞后和监管缺失，如此极有可能令数字内容产业陷入恶性竞争和混乱秩序。类似问题还存在于采用人工智能技术分发内容过程中出现问题之后平台与生产者之间责任划分等方面。

29 Yisroel Mirsky etc:CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning[J].https://arxiv.org/pdf/1901.03597.pdf.
30 判决书原文链接：https://www.bjinternetcourt.gov.cn/cac/zw/1556272978673.html.

2.5 “信息茧房”循环强化，或引发网络社群“部落化”“极群化”

算法主导下的内容分发模式，会放大和加强被标签了的信息输出和推送，由此引发“自我封闭”的危险。“信息茧房（Information Cocoons）”[31]并非人工智能时代的产物，但在算法的助推下，传播中“回音室效应”将被放大，人们极易过滤和忽视那些自己不熟悉、不喜欢、不认同的信息，只会看到和听到自己希望看到和听到的内容，在不知不觉地长期重复和自我证实中塑造出桎梏自身观念的“茧房”。

一旦身处其中，就再难接受异质化的信息和不同的观点，甚至在不同群体、代际间竖起阻碍沟通的高墙[32]。相同观念的人们在各类议题热点下逐渐聚集，对所处社群观念不断强化认知，加速网络社群的“部落化”，甚至最终走向不同社群间观念极化对立和舆论失衡的极端。

2.6 无序数据挖掘泄露个人隐私，跃升数字内容治理突出问题

人工智能技术对数据的依赖程度极高，由此加剧的个人隐私泄露问题，跃升为数字内容治理的突出风险。

万物互联、万物皆媒的人工智能时代，任何智能终端都有可能成为内容的信源和接收的窗口，存储着大量可供挖掘的数据。相较于传统互联网应用主要采集用户上网习惯、消费记录等信息，人工智能应用可采集的信息更加丰富多样，包括用户人脸、指纹、声纹、虹膜、心跳、基因等具有强个人属性的生物特征信息。这些信息具有私密性、唯一性和不变性，一旦被泄露或滥用将对公民权益造成严重影响[33]。国内外媒体不乏曝光人工智能应用泄露个人隐私数据的案例，所涉用户高达几百万，恶意收集并泄露的隐私数据量以百万计，其中不乏个人身份信息、人脸识别图像和 GPS 位置记录等。

此外，若利用人工智能技术对公开合法手段所收集的非敏感信息进行综合关联分析，同样存在推测出敏感个人信息的风险，而各种匿名化技术使得对数据泄露的溯源追踪更加困难，增加了个人信息保护和维权的门槛。

2.7 信息资源竞争催生新数字鸿沟，将成数字内容治理新难题

人工智能时代，包括国家、地区、行业、企业、社区在内的信息活动主体，由于对信息、技术的拥有程度、应用程度及创新能力存在差别，或将造成扩大的信息落差，导致贫富进一步分化。

31 Cass R. Sunstein：Infotopia:How Many Minds Produce Knowledge[M].Oxford University Press
32 2017年人民网二评算法推荐的文章：《别被算法困在“信息茧房”》.
33 中国信通院《人工智能数据安全白皮书（2019年）》.

信息社会的竞争目前已逐渐演变为对信息资源的争夺，谁拥有信息资源，谁才能有效地使用信息资源。人工智能的价值分配会使一部分群体受益，那些拥有信息及强大数据处理能力的企业或个体通过人工智能技术获得更多信息占比，从而获取更多优势，或可能产生新的数字鸿沟和机会差距。

在个人层面，一些占有或获得信息较多的人，逐渐从他们原来从事的职业领域中分化出来，从事崭新的职业或更高层次的职业，进而获得更高的收入。而那些被自动化所取代的就业者不得不寻求新的就业机会，即便找到新的工作，也往往是低附加值的，且工作报酬更低，这就有可能进一步加剧个人间的财富分配差距。

在国家层面上，发达经济体掌握成熟的智能技术，持续更新升级产业形态，将直接冲击发展中国家人力资源等比较优势，使许多发展中国家再次被锁定在资源供应国的位置上。

同时，人工智能技术较高的迭代速度下，极易形成“马太效应”，使具有先发优势的国家强者更强，后发国家越来越难以追赶，造成国际社会的阶层固化。

另外，人工智能时代各国数字内容产业发展的背后，是本国几乎全部文化信息的“数字化曝光”，传统时代难以获取的社会情报将在人工智能时代公开化，数字内容强国的“战略传播”将更为便利。

上述新产生的信息分化和数字鸿沟，对人类社会的潜在冲击是十分深远和不容小觑的。

3 人工智能时代数字内容治理的政策与举措

不同国家的数字内容产业有着不同的发展路径，所呈现的治理问题在种类、风险和危害程度上亦有差异，因此各国治理政策各具特色，本节主要考察了代表性较强的美国、欧盟、英国、新加坡、日本、中国六个国家和地区，比较并总结其人工智能时代数字内容治理的政策发展情况。

3.1 美国：重点治理内容造假，多手段规制算法

美国是目前全球数字内容产业规模最大的国家，2018年，数字内容产业占美国GDP比例高达18%～25%[34]。人工智能时代下，美国政府面临的数字内容治理的突出问题，一是国内外政治势力利用内容造假与精准推送操控政治舆论，冲击美国政治生态；二是数字内容产业应用人工智能技术过程中产生的歧视和不透明现象。为应对上述挑战，美国政府不断强化其出于国家安全目标考虑的网络内容监管审查。

针对第一类问题，美国会参众两院高度重视，通过召开听证会或提出法案等尝试将深度伪造问题纳入立法程序实施规制。

2019年6月13日，美国众议院情报委员会召开了关于人工智能深度伪造的听证会，公开谈论了深度伪造技术对个体名誉与隐私权损害、企业信誉及经济损失、国家/公共安全、新闻媒体行业的社会信任衰退等带来的风险及应对措施，这是美国众议院首次举办专门讨论深度伪造及其他类型的AI合成技术的听证会。

2019年6月28日美国国会参众两院议员共同提出《深度伪造报告法案》（Deepfake Report Act of 2019）[35]，要求美国国土安全部长定期评估：（1）深度伪造所使用的具体技术工具；（2）国内外网络攻击者、色情爱好者、新闻媒体所使用的深度伪造工具类型；（3）外国政府及其代理者如何使用该技术危害美国的国家安全；（4）有哪些方法可以用于判断一段内容属于深度伪造；（5）有哪些可用于反击深度伪造的技术措施等。德克萨斯州也于同一时期通过了《关于制作欺骗性视频意图影响选举结果的刑事犯罪法案》（An Act relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election）[36]，旨在对抗虚假新闻和深度造假对选举安全的破坏。

34 赵春华:《我国数字内容产业政策的演变与评估》[D]2018.06.山西大学经济管理学院：12.
35 法案文本链接：https://www.congress.gov/bill/116th-congress/house-bill/3600/.
36 法案文本链接：https://legiscan.com/TX/text/SB751/id/1902830.

针对第二类问题，政策举措侧重在制定算法标准、设立法律规范以及采取技术手段。

2017年1月12日，美国计算机协会下属美国公共政策委员会发布文件，阐述了关于人工智能算法透明化和可责性七条原则，具体包括意识原则、准入和补救原则、问责原则、透明原则、数据来源原则、可审计性原则、验证和测试原则。

2017年12月11日美国纽约市议会通过《算法问责法案》[37]，提出成立一个由自动化决策系统专家和受自动化决策系统影响的公民组织代表组成的工作组，专门监督市政机构使用自动决策算法的公平性、问责性和透明度。

2019年4月19日众议院提案《2019年算法问责法》（Algorithmic Accountability Act of 2019）[38]将赋予美国联邦贸易委员会（FTC）新的权力，迫使企业研究审查其技术应用中是否存在种族、性别或其它方面的偏见。

总体而言，美国人工智能时代数字内容治理政策主要呈现三个突出特点：第一，结合对外政治价值和文化输出影响，加强对传统大众传媒产业的分级审核；第二，基于国家意识形态安全和政治生态稳定的考虑，对互联网内容产业实施日趋严格的监管；第三，极为重视人工智能算法问题；第四，相关立法层级和执行部门均相当分散，依旧强调行业机构与企业媒体自律，公众自我约束和社会多元参与。

3.2 欧盟：多方应对歧视言论，强调伦理与隐私

欧盟是最早提出“数字内容产业”概念的政治体，以数字音乐、电子游戏为代表的传统数字内容产业总体产值巨大，但在以搜索引擎和社交平台为代表的新兴数字内容产业发展中缺失了先机，市场份额基本被美国公司占据。人工智能时代下，欧盟的数字内容产业治理问题突出表现为本土数字内容产业发展缺失带来的经济损失,种族歧视、仇恨言论问题较为严重,以及算法本身的公正性和过度挖掘数据带来的问题。

欧盟近来因“难民问题”和极右翼势力抬头导致社交平台上出现较多种族歧视、仇恨的言论，为此各成员国不断推进立法规制。

37 法案文本链接：https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0.
38 法案文本链接：https://www.congress.gov/bill/116th-congress/house-bill/2231.

2016年11月23日，欧洲议会通过了《欧盟反击第三方宣传的战略传播》（EU strategic communication to counteract propaganda against it by third parties）决议案[39]，呼吁打击恐怖主义、犯罪组织和国外势力利用新闻媒体特别是社交媒体和各类数字平台扩散宣传和制造虚假信息。2018年9月26日，欧洲主要的线上平台、社交媒体巨头、广告商和广告经营者代表齐聚布鲁塞尔，联合发布了欧盟史上首份《反虚假信息行为准则》（Code of Practice on Disinformation，CPD），以应对日益严峻的假新闻与网络虚假信息肆虐的局面。这是全球范围内企业界主动自愿联合，共同应对虚假信息挑战的首次尝试，在全球互联网信息治理历史上具有标志性意义。

从欧盟接连发布的人工智能战略和准则不难看出，欧盟普遍重视技术伦理、算法黑箱、数据歧视等人工智能技术的内嵌安全问题。

2019年4月，欧盟委员会发布《欧盟人工智能伦理准则》[40]，强调了人工智能应用的安全性、隐私数据管理、透明度和问责机制等原则，倡导将伦理和法律纳入人工智能算法设计。对于算法应用带来的数据过度挖掘问题，以及由此加剧的个人隐私泄露风险，欧盟是在数据安全议题下统一规制的。

2018年5月生效的《一般数据保护条例》（GDPR）作为全球最严格的数据保护法律之一，对算法采集、使用、加工个人数据也做出了诸多规定。

总体而言，欧盟成员国针对人工智能时代数字内容治理做出的政策创新，一是在打击有关意识形态、恐怖主义、种族仇恨等言论和虚假新闻上比较容易达成共识；二是普遍关注人工智能技术伦理、算法等内嵌安全问题，且有关倡议和规则对此的影响力也相对较大；三是将因算法应用产生的个人隐私风险放在数据安全规范中统筹考虑，严格规制。

39 法案文本链接：http://www.europarl.europa.eu/doceo/document/A-8-2016-0290_EN.html?redirect.
40 准则文本链接：https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

3.3 英国：重视技术赋能作用，鼓励企业自律

英国将数字内容产业定义为“创意产业”，游戏、音乐、电影、广告、出版、设计等产业规模和竞争力都位居世界前列，是世界上最早进行国家政策性推动创意产业发展和最早制定国家级人工智能战略的国家。

2017年10月，英国政府发布《英国人工智能发展报告》（Growing The Artificial Intelligence Industry In The UK）[41]，明确提出提升对数据开发的信任，提升信息数据共享性，支持文本和数据挖掘，并将其视为一种研究的标准和不可或缺的工具，特别是提出将数字营销和文化创意产业列为人工智能发展的重点应用场景。

2018年4月，英国议会下属的人工智能特别委员会发布了《英国人工智能发展计划、能力与志向》（AI in the UK: ready, willing and able?）[42]报告，强调支持英国实现利用人工智能技术赋予对社会和经济更大潜力，包括发挥其对提升文化创意产业生产力的作用；认为当前不需要对人工智能进行专门监管，各个行业的监管机构完全可以根据实际情况对具体问题和监管做出适应性调整；英国政府应当通过制定国家层面的人工智能准则、伦理原则和相关标准，促进行业自律。

据此，英国政府设立了数据伦理与创新中心（CDEI），于2019年3月发布研究战略和年度工作计划[43]，称将对算法决策（algorithmic decision-making）引发的社会偏见开展研究，计划2019年完成在线消息定位审核、偏见审查、机会和风险预测等方面的研究，并形成报告。此外，CDEI还将与种族差异审计部门（RDU）合作，联合调研刑事司法、金融服务、招聘和地方政府管理中算法决策应用的潜在歧视问题。机构总体目标是解决现有系统中潜在的偏见，保障更公平的决策，确保数据驱动型技术和人工智能被用来造福社会。但CDEI只是专家组成的独立咨询机构，BBC、《卫报》、“第四新闻频道”等各类媒体仍为自主审核的主体。

总体而言，英国重视人工智能对数字内容产业的赋能作用，在政策发展上体现为，一是鼓励利用人工智能促进数字内容产业发展，增进海外竞争力的作用；二是针对虚假信息问题，目前依旧沿袭其传统媒体治理的中立理念和保守政策，但未来或可能推进立法政策。

41 法案文本链接：https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.
42 法案文本链接：https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.
43 计划文本链接：https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation.

3.4 新加坡：加强媒体内容治理，战略布局产业

新加坡对数字内容产业的界定通常采用“文化创意产业”概念。“城市国家”的独特定位使该国十分重视以电子商务、文化旅游、建筑设计为代表的数字内容产业发展；出于其多民族、多语言聚居和高度国际化的国情特点，新加坡强调媒体治理的重要性。人工智能时代下新加坡面临的数字内容治理风险，一方面关切算法公平等技术问题、虚假新闻和歧视言论等社会政治问题，而另一方面，受限于国家体量，同时存在对可能因加强技术监管而导致本国丧失数字内容产业竞争力的担忧。

2017年5月，新加坡推出为期五年，总投入1.5亿新元的《新加坡人工智能战略》[44]，其中将文化创意和数字营销列为城市公共管理的重点应用行业。2019年1月，新加坡个人数据保护委员会（PDPC）与信息媒体发展局（IMDA）正式提出“人工智能治理框架建议模型”[45]，旨在促进人工智能技术采用，建立消费者信心与信任，为人工智能提供训练数据。

新加坡国会于2019年5月8日通过的《防止网络假信息和网络操纵法案》（Protection from Online Falsehoods and Manipulation Bill）[46]规定，政府有权要求个人或网络平台更正或撤下对公共利益造成负面影响的虚假信息；允许政府下令媒体平台关闭传播虚假信息的账户或机器人，不遵守规定的网络平台最高可被判罚约合500万元人民币的罚款。

此外，新加坡通讯及信息部下属的资讯通信媒体发展局将专设办事处，监督科技公司对国家政策的落实，为政府应对虚假信息提供技术咨询，并与企业合作制定相关行业守则，包括要求网络平台进行身份验证，防止人们滥用账号，确保政治广告来源透明以及优先推送可信内容。

总体而言，新加坡更倾向于采取强化媒体内容监管审核与大力促进数字内容产业双管齐下的治理思路，前者主要出于维护国家稳定和安全的考虑，后者则是与新加坡从国家战略层面布局文化创意产业保持一致的体现。

44 战略文本链接：https://www.nrf.gov.sg/programmes/artificial-intelligence-r-d-programme
45 政策文本链接：https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf.
46 法案文本链接：https://sso.agc.gov.sg/Bills-Supp/10-2019/Published/20190401?DocDate=20190401.

3.5 日本：关注算法公平，探索知识产权保护

在日本数字内容产业中动漫、电影、音乐、游戏在日本数字经济发展中占有重要地位，人工智能时代下，日本的数字内容治理政策主要关注人工智能算法和内容智能生产的知识产权归属等问题。

日本在《第五期科学技术基本计划》中提出“Society 5.0”发展目标后，内阁府于2018年12月27日发布了《以人类为中心的AI社会原则》，从宏观和伦理角度阐明了日本政府的态度，并提出“AI-Ready社会”的7项基本原则：人类中心原则，教育应用原则，保护隐私原则，安全保障原则，公平竞争原则，公平、说明责任及透明原则，创新原则。

日本知识产权战略总部早在2016年5月就宣布着手保护人工智能(AI)创作小说、音乐等的知识产权，由于AI作品不属于现行日本《著作权法》的适用对象，将为此研究制定新法[47]。

2019年6月21日，该部门发布《知识产权推进计划2019》，明确强调要建立相关制度以促进大数据和人工智能的合理利用，并将在专利制度中重点讨论如何区分仅由人工智能完成的发明和有发明人参与完成的发明[48]，但目前尚未有较为明确详尽的法案提出。

3.6 中国：侧重垂直行业监管，探索应对新兴问题

我国数字内容产业虽起步较晚，却在互联网企业的创新推动和国家政策的扶持鼓励下，发展颇为迅速。

2016年12月19日，国务院颁布的《“十三五”国家战略性新兴产业发展规划》将数字创意产业纳入了五大新型支柱产业，并提出了创新技术装备、丰富内容和形式、提升设计水平、推进与其它行业的合作发展四点要求。随着我国国力日益提升，文化自信不断增强，在视频直播、影视动漫、电子游戏等各个产业均呈现出可喜的发展势头，在搜索引擎、社交媒体、电子商务、移动支付等行业均有具备全球竞争力的本土企业，海外市场拓展前景明朗。

47 新华网：日本政府决定着手保护人工智能作品知识产权[EB/OL].http://www.xinhuanet.com/world/2016-05/09/c_128971675.htm.
48 日本发布《知识产权推进计划2019》[EB/OL].http://www.worldip.cn/index.php?m=content&c=index&a=show&catid=64&id=1050.

在人工智能时代，我国面临的数字内容治理问题主要包括数字内容产业新业态与现有治理模式不匹配，算法歧视和黑箱问题、个人隐私泄露的风险加剧等。

目前，我国没有针对数字内容治理的专项法律，而是在既有法律体系中，通过健全完善数字内容产业各领域行政法规、部门规章、部门工作性文件、地方性法规等开展治理，新闻信息服务、社交群组、视频、直播、电子游戏是我国数字内容治理中的政策焦点[49]。这种模式能够保持政策的灵活性、连续性、及时性，但同时也存在主管部门分散、政策文件繁杂交叉、强制性政策较多、不同政策协调性较低等缺陷。

据统计，自2000年《互联网信息服务管理办法》发布以来，中央党政机关及其职能部门出台的涉及“网络媒体”“互联网媒体”的政策文本高达95份，强制型政策工具共计307条，占全部政策文本条款的68.5%[50]。

针对人工智能应用于数字内容产业中带来的风险问题，政府关注并积极开展相关调研，并尝试倡导和推进算法公约，呼吁各企业加强自治。

在算法歧视问题上，中国计算机协会在2017年出版了七项算法透明度和问责制的原则，包括意识、获得和纠正、问责、解释、数据来源、可审核性、验证和测试。

2019年6月，国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》，强调和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理等八项原则。

2019年7月上海市人工智能产业安全专家咨询委员发布的《人工智能安全发展上海倡议》中强调“面向未来、以人为本、责任明晰、隐私保护、算法公正、透明监管、和平利用、开放合作”等八大目标原则及其基本要求。

49 谢新洲，李佳伦《中国互联网内容管理宏观政策与基本制度发展简史》
50 王国华，李文娟. 政策工具视角下我国网络媒体政策分析——基于2000—2018年的国家政策文本[J/OL]. 情报杂志. http://kns.cnki.net/kcms/detail/61.1167.G3.20190712.1431.013.html

2019年8月13日，由中国科技部、中央宣传部、中央网信办、财政部、文化和旅游部和广播电视总局联合发布了《关于促进文化和科技深度融合的指导意见》，基于对“文化和科技深度融合仍面临许多新的挑战，科技对文化建设支撑作用的潜力还没有充分释放，相关部门和地方对文化和科技融合的重要性和紧迫性的认识尚需进一步提高”的认知。提出了加强文化共性关键技术研发、完善文化科技创新体系建设、加快文化科技成果产业化推广、加强文化大数据体系建设、推动媒体融合向纵深发展、促进内容生产和传播手段现代化、提升文化装备技术水平、强化文化技术标准研制与推广等六项重点任务。强调要推动人工智能技术在文化领域的深度应用和创新发展，在文化领域建设人工智能公共服务平台，建立“智能+文化”开源技术开发社区……强化文化领域新一代人工智能技术的有效供给。由此可见，我国已经逐渐意识到未来数字内容发展的重要性，并开始逐步加强顶层设计和统一布局，相信不久的未来，数字内容治理问题亦将提上战略日程。

目前，我国数字内容治理政策沿袭传统模式，仍以内容垂直领域监管为主，对于人工智能时代的算法问题、内容造假、知识产权等新兴问题总体处于行业探索和部分试行阶段。

3.7 小结

总体而言，各国在应对人工智能时代新型、复杂的治理挑战时，**各国现行数字内容治理政策显得较为滞后和无力，但都结合各国数字内容发展情况作出各有侧重的政策创新和尝试。**

美国强调对内容造假和算法歧视冲击国内政治生态的问题治理。

欧盟强化算法伦理规范，对数据泄露和歧视言论采取较为严格的规制。

英国侧重发挥人工智能对数字内容产业的促进作用，针对负面问题则更多采取企业自律的传统路径。

新加坡和日本除了同样重视人工智能技术的赋能作用以外，就内容造假和知识产权问题开启多项立法进程。

中国主要通过健全完善数字内容垂直领域的政策规制展开治理，集中关注新闻信息服务、社交群组、泛视频、直播和电子游戏。

具体而言，首先数字内容与各国政治、经济、文化、历史等国情结合紧密，表现出较大的国际和地域差异，产业界定和发展历程各不相同，**各国在数字内容治理的理念、主导部门、政策目的、措施手段区分明显，但当与人工智能结合时面临的风险存在一定共性，反映到政策中也有相似之处，如对传统治理路径的依赖，关注算法歧视、内容造假、隐私泄露、产权保护等。**

其次，由于受不同政治体系和社会环境影响，各国所面临的风险种类、表现、危害程度有所不同，**各国治理政策重点会因此各有倾斜，但总体仍呈现出对新兴问题采取应激性举措的特点。**

最后，人工智能技术尚处于技术成熟早期，在数字内容产业的应用尽管遍地开花，但总的来看还只是初露峥嵘，大量普及和深度应用刚刚起步，相应的风险处于初现阶段，各国此类问题的认识和研究有限，**因此，普遍以人工智能科技和数字内容文化两种分野明显的视域践行治理政策，缺少针对二者结合产业的本质特点，对人工智能时代下数字内容治理问题制定的顶层规划和统筹应对。**

4 人工智能时代数字内容治理的企业实践与探索

作为多技术集合体，人工智能在数字内容治理方面有着极大的应用前景，数据挖掘、归因分类、机器学习、自然语言处理、计算机视觉、模式识别等各种技术方向均会在数字内容治理的不同场景中发挥重要作用。本节以场景为导向，介绍人工智能技术在内容审核、事实核查、版权保护、打击诈骗、舆情治理和破除信息茧房等场景领域下的技术应用思路和产业探索实践。

4.1 内容审核

4.1.1 安全风险

以淫秽色情、垃圾信息、暴力低俗等为代表的违法与不良信息是数字内容治理的固有顽疾，随着承载并允许用户自主生产和发布内容的社交媒体平台的繁荣，负面信息的扩散传播带来的社会影响越来越引起各方关注，而随着内容数量指数级的增长，文本、图片、音频和视频多种内容形式复杂变化使传统人工内容审核愈发低效和艰难。

4.1.2 AI+

人工智能技术发展与应用，极大地提高了不良信息识别发现、审核判别、处置处罚等治理效率，有效节省人力、物力成本。在信息识别发现环节，利用人工智能技术可加强对不良信息的特征识别，提高筛选素材的效率和质量，为高效治理奠定基础；在审核判别环节，传统的人工审核机制需要配备大量审核人员，对于违规特征不明显、不易识别的恶意信息，此审核方式效率极低，且不同审核人员存在标准认知差异，审核输出质量偶有不同，人工智能技术的应用使机器审核在准确率、效率和标准化方面均能得到保障；在处置处罚环节，人工智能技术加大了审核后处置方式选择、处置生效效率，避免人工操作失误，实现机器审核、机器处置的闭环，对人工审核、处置的治理方式形成补充。

国内案例：

基于腾讯云图像分析等人工智能技术，微信推出“珊瑚内容安全助手”小程序，为广大小程序开发者提供风险自测、内容鉴别、行业动态三个维度的内容安全能力。如在文字鉴别、图片鉴别模块，“珊瑚”小程序可对用户上传内容中是否含有违法有害信息给出鉴别结果，同时提供能力介绍和接入引导文档，为开发运营者从“体验”、“理解”到“接入”提供一站式安全能力服务。同时，“珊瑚”小程序持续运营、定期更新行业内容安全动态信息，帮助开发运营者了解最新的行业情况，共建健康繁荣的小程序生态。

腾讯云、信安团队提供的内容安全解决方案和冰鉴引擎，在内容审核上能实现风险的精准识别、风险主动感知、风险实时决策、模型快速运营，综合运用Hybrid Neural Network、Generative Adversarial Networks、Domain Adaption、Meta Learning等算法模型，以及在OCR图片识别模块中，采用字典识别技术、CNN算法模型对图片进行相似度识别，无需进行强化训练，根据数据来直接升级模型，达到快速生效的目的，识别准确率高达99%，首创通过类似语音关键词唤醒的技术极其快速地检测图片是否包含色情等不良信息中特定的关键词；在暴恐识别方面引入先进的Image Caption技术等，提高有害信息的审核精度。

国际案例：

Facebook公司开发使用eGLYPH工具用于极端主义、歧视言论的内容审核，包括事前审核、事后检测、删除和防止洗稿上传。该工具利用数字哈希技术和NPL创建，基于开发者预设的关键词句特征识别文本和图片内容，并做出相应处理，被判定为不良而删除的信息，会作为种子样本进入训练数据库，继而通过算法模型进行深度学习，实现扩展数据库、软件更新、算法迭代的闭环。

此外，Facebook还推出了DeepText（深度文本）引擎，利用深层神经网络架构理解帖子内容，据称DeepText能够以近乎人脑的精确度，每秒同时理解数千篇文章的文本内容。除了速度更快这一优势外，DeepText能够审核超过20多种语言的文字，甚至能实时通过用户发送的内容分析理解用户行为背后的“想法”，凭借对意图、情绪和实体（人物/地点/事件）的提取技术，结合文本、图片，自动移除垃圾信息。

YouTube开发的名为ContentID的内容审核系统，可有效监测并阻断涉及色情、低俗和暴力等违规内容的传播。以2017年第四季度为例，平台清除了800万条“令人反感”的视频，其中670万条由监测软件自动标记。大约75%被标记的视频，在用户观看之前就被成功拦截。YouTube的内容审核能力有赖于Google的深度学习技术Google Brain的支持。

4.2 事实核查

4.2.1 安全风险

虚假新闻、网络谣言和深度造假等内容造假是目前数字内容治理影响最大的问题之一，除各国日趋重视并立法规制以外，各大媒体和社交平台也在建立或优化事实核查机制打击虚假信息。然而，数字内容信息来源多样，内容复杂，数量庞大，传播速度迅猛，信息误导还可能以标题党、冒名顶替、不相干语境、事实夸大或片面等各种形式存在，对事实核查高效、快速、准确能力的要求越来越高。

4.2.2 AI+

基于人工智能技术的数据挖掘、文本汇聚、深度学习等技术能够有效检索虚假内容的传播源头，构建各类结构数据库和标识体系，帮助核查者对海量资讯进行针对性处理，随着对合成图片、声音和视频的鉴伪技术和溯源技术的研发精进，对虚假信息的识别取得了较大突破。相信未来，人工智能技术能够在内容造假的发现、筛选、分类、核实、处理、澄清等各个方面提供技术支撑，在虚假新闻、网络谣言、深度造假治理方面发挥更大作用。

国内案例：

基于海量数据与人工智能技术，腾讯微信公众平台辟谣中心、微信安全中心、腾讯新闻较真平台、腾讯内容开放平台企鹅号等探索出一系列有效的事实核查措施，如打造辟谣数据库，智能识别处置谣言，借助机器算法触达谣言易感人群，基于阅读或投诉谣言的类型标签进行精准推送辟谣防谣。

此外，腾讯微信公众平台通过机器学习等算法和大数据，分析“标题党”传播特征、手段，提高机器针对“标题党”的识别准度和精度。

在识别出“标题党”文章后，公众平台会对相关文章的标题加注显著标识，并在文章转发传播时以警示处罚覆盖原标题。针对内容恶劣的“标题党”文章，微信公众平台会直接删除该文章并向运营者下发处罚通知，根据帐号违规严重程度，建立了删除文章——帐号禁言——封号的阶梯处罚机制。上述措施效果明显，能较好地对制作不符标题的公众号运营者起到警示作用，大部分问题账号接受处罚后会选择删除“标题党”文章，并在后续的发文中注意措辞。

国际案例：

美国三大主流事实核查机构PolitiFact、FactCheck和FactChecker在事实核查中不断提升人工智能技术的应用水平，以PolitiFact为代表，基于机器学习研发的真实性测量仪“Truth-O-Meter”将信息资讯的真实程度分为“真实”、“基本真实”、“部分真实”、“大部分失实”、“失实”、“完全失实”六个级别，以便根据失真程度做出不同处置。

2016年8月，Facebook宣布针对平台上的新闻资讯内容，已实现全部由机器算法完成核查。同年11月，Facebook与“国际事实审核网络”（International Fact-Checking Network, IFCN）开展合作，邀请ABC新闻、美联社、华盛顿邮报等IFCN缔约成员机构使用自己开发的工具，共同评估新闻的真实准确性。

Google的做法是查询和比对可靠的新闻核查机构和schema.org中“声明核查”（ClaimReview）已标记的检验信息来实现信息真实性的核查。Schema.org是2011年Google、必应、雅虎等搜索引擎巨头共同创建的优化搜索引擎结果的html标记系统。

2017年1月，法国《世界报》推出了一个名为“Décodex”的事实核查数据库，帮助读者识别假冒或不可靠的网站。《世界报》事实核查团队查证了600多个网站，其中包括博客网站、Facebook网页、Twitter账号等，依据它们的可信度和准确性，对这些网站进行分类。此外，使用《世界报》网站搜索引擎，读者可通过颜色编码系统快速判别某网站的可靠性，标记为绿色的网站高度可靠，黄色应谨慎阅读，红色意味着该网站虚假信息威胁度极高；讽刺性网站标记为蓝色，而一些不能被验证的网站则被标记为灰色。用户还可主动标记尚未成为数据库一部分的其它网站，经后台系统审核确认后添加进数据库，《世界报》同时宣布其数据库将保持开源。

初创公司Factmata正在构建的社区驱动事实检查系统，可以通过深度学习技术和自然语言解析技术，对新闻内容进行检索和分析，甚至给新闻内容质量和可行度进行评分，识别可能出现的错误，并在网络上提供更准确的信息。该技术目前已被运用于投资研究评分，对冲基金交易，程序化广告，搜索引擎等领域。

4.3 版权保护

4.3.1安全风险

数字内容的全球化传播和共享，以及各个地区知识产权的政策差异使得内容版权保护成为重要议题，对不同内容形式进行的深度洗稿使得传统版权保护措施难以应对，原创内容中的版权确权成本高、盗版猖獗、维权难、变现模式单一等亦是数字内容产业的普遍痛点。这使得未来的知识产权保护必须做到敏捷、高效、精准、实时，而人工智能技术能够帮助管理主体更好地维持保障创新回报与防止资源配置失衡之间的平衡。

4.3.2 AI+

基于人工智能技术构建的智能检索、类比分析应用于帮助检测侵权状况和行为，保障内容创作者的知识产权；依托人工智能技术在内容中生成的特定标识可提高侵权成本，从而减少侵权行为；人工智能识别技术还可应用于数字内容或以数字内容认证的其他产品服务的知识产权保护、打击抄袭洗稿伪造等执法活动；而在海量、复杂、异构的知识产权管理工作中，也可利用结构数据库优化管理流程，提高效率。

国内案例：

基于腾讯安全团队开发的神鸮网络侵权监测系统，充分利用了AI优势，包括70亿个点和1000多亿条领先的知识图谱，成功克服并解决了现有市场监管领域“数据、算法、计算力”不足的问题。神鸮采用了基于违法样本挖掘并进行自动化、可视化的方式方法，搭建了从数据源管理到风险展示的系统架构，对侵权行为的关键环节构建针对性的保护措施，形成了“全网发现——线索串并——电子固证——实时拦截/助力抓捕”的网络侵权全流程处置方案，基于全网每天20亿网址识别及全网数据分析建模，可实时发现侵权链接、盗链情况，将侵权盗版态势情况的多维度风险评估结果及时反馈给版权人和相关单位。

另外，腾讯安全能够实时阻断侵权网址链接在微信、QQ内的传播，同时与APPLE、华为、 vivo、OPPO等手机厂商展开合作,对其内置默认浏览器的网络侵权链接进行实时阻断,拦截大约93.6%的侵权网址访问行为。2019年春节期间，《流浪地球》等热门影片广受关注，网络侵权盗版问题突出，腾讯安全神鸮监测数据显示，这些热门影片盗版网站数量达到1.46万个，涉及域名1379个。为阻断侵权行为，腾讯安全神鸮配合国家版权局阻断访问盗版网址总量达到1.9亿次。

百度图腾在百度超级链基础上构建的内容版权链，将版权内容信息如登记确权、维权线索、交易信息等存储于百度分布式存储系统中。平台将通过计算机视觉技术对图片进行识别，并基于智能图片搜索系统和覆盖全网的盗版追踪监测系统，可以第一时间锁定版权图片在互联网中的盗版使用，并自动进行取证。

国际案例：

CSAM检测技术被称为PhotoDNA，最初由微软开发，现已扩展为美国各大互联网巨头（Twitter、Google、Facebook）和执法部门（美国国家失踪和受剥削儿童中心等）等机构广泛使用的基础工具，用于版权保护。PhotoDNA从一个包含现有非法CSAM图像的数据库中生成数字散列，可在几微秒的时间内实现广谱范围内的检测散列，算法所训练的签名数据库也是不断更新的。

YouTube采用PhotoDNA技术创建的ContentID技术，允许用户为其图片、视频内容创建数字散列，保护其免受版权侵犯。一旦创建了这些散列，所有随后上传到YouTube平台的内容都将经过数据库筛选，以识别潜在的侵犯版权行为。目前，YouTube的CSAM数据库存储了超过72万个示例，还将随着侵权内容标记的添加而不断扩展。

以上两种工具对包括调整大小、颜色变化和水印等操作具有特别弹性。

4.4 打击诈骗

4.4.1 安全风险

通过电话、短信、邮件进行骚扰和诈骗一直是互联网时代的难题顽症，其衍生的网络病毒扩散、用户体验下降、重大财产损失使得人们深恶痛绝。

4.4.2 AI+

人工智能技术能够在打击垃圾和诈骗通信上发挥重要作用。对于垃圾邮件和短信而言，人工智能技术能够通过文本分类、词汇处理、数据清洗、向量值确认等实现对垃圾邮件、短信的有效检测、识别和拦截，甚至别出心裁通过智能机器人自动回复来报复对方；在打击诈骗过程中，能够通过海量样本数据的学习实现对诈骗前期发现和预警，基于大数据类比技术和生物识别能够协作甄别“伪基站”。

国内案例：

腾讯公司基于其黑产对抗经验技术和反欺诈AI模型建造能力，打造“宾果反诈骗防控系统”，该系统通过海量学习警情和大数据分析能力，自主提取警情中作案手法、通信行为、网络特征、资金流向等特点规律，实现智能建模、智能运算、智能预警，在诈骗事前、事中、事后等环节起到预警、分析作用。通过对警情、通信、网络、金融等领域大数据的深度学习，宾果系统可自动发现电信网络诈骗犯罪行为并自动预警，自2018年3月21日上线以来（截止2019年8月下旬），宾果反诈系统预警已超过140000起，准确率超过90%，覆盖全国31个省区市，累计为民众避免损失60亿元。其次，宾果系统可实现对电信网络诈骗窝点、人群的智能聚类，为警方开展刑事打击提供线索参考。此外，宾果系统还能够对已发案件进行智能分类、特征刻画、人员扩展和团伙聚类，通过系统机器学习和自动运算，排查诈骗团伙窝点位置、规模、人员、作案手法等，为警方开展针对性刑事打击提供有力参考依据。

国际案例：

Google公司利用人工智能技术在几年前就将Gmail电子邮件服务的垃圾邮件拦截率提升至99.9%、误报率降低至0.05%。Gmail目前拥有高达9亿的全球用户，面对不同地区用户，Gmail垃圾邮件智能过滤器不仅仅是通过预设规则清除垃圾邮件，在运行期间还可根据具体情况自行制定新规则。

4.5 舆情监测

4.5.1 安全风险

各主体在数字内容的生产、传播和交流中的表达情绪和立场汇聚成为网络舆情，对现实社会产生有着极大的影响力和观念塑造能力，不仅仅关系到网络安全和舆论安全，更关系到整个社情民意和国家稳定，是数字内容治理中十分重要的命题。

4.5.2 AI+

基于人工智能构建的情感分析技术、舆论模型、态势感知、可视化呈现、应急处置等机制，能够实现对网络舆论各个主体的标准化信息采集、汇聚、分类，对特定事件和整体舆论环境的实时态势作出理解和评判，对舆情的发展趋势、关切热点、各方态度做出预测和提供处理建议，从而推动舆情治理工作更加准确和有效。

国内案例：

智能舆情平台是蚂蚁金服基于人工智能技术和数据处理能力构建的一站式专业舆情分析服务平台，主要应用于金融行业的舆情检索、事件聚类分析和舆情API服务，能够对与特定公司相关的投资事件进行全网检索和跟踪，并基于企业知识图谱进行多维度的舆情画像，还可以利用OpenAPI提供基于金融实体的金融舆情订阅和深度分析。该平台目前覆盖十万级站点，日更新量千万级，更新频率达到分钟级，并且可以根据业务需要自定义数据源、监控和跟踪实体以及各类事件。

国际案例：

亚马逊公司的AWS cloud，能够为其公共部门客户利用机器学习来“聆听民众的声音”提供支持。通过CloudFormation模板，客户可输入特定热点或政策，收集社交媒体和数字平台中的公开言论，自动生成可视化舆情结构图表和报告，观察民众对于政策的讨论热度和态度变化趋势，实现及时更新。

硅谷大数据独角兽Palantir公司长期接受与中央情报局深度关联的高科技风险投资公司In-Q-Tel的投资，主要利用数据挖掘技术来构建数据库，经过机器学习分析后用直观的可视化方式输出结果，其产品和服务在公共安全、国防安全和海外军事行动中被美国各个机构广泛采用。

4.6 破除信息茧房

4.6.1 安全风险

信息茧房由美国哈佛大学凯斯·桑斯坦（Cass R.Sunstein）提出的，是指公众在海量信息中倾向于关注自我选择或使自身愉悦的信息，这种倾向将在长期下塑造桎梏自身观念的“茧房”，难以接受其他信息和不同观点。在长期接受重复性同质资讯下，相同观念的人们在各类议题热点的讨论中日益集聚，并基于与其他社群不同观念强化自身认知，这将会导致网络社群“部落化”，不同社群之间的观念极化对立和舆论失衡。

4.6.2 AI+

用户接受信息的总量是有限的，“茧房”边界可能始终存在，但是通过算法在“精准化”和“多元化”两个方向的优化，能够有助于破除信息茧房[51]。在算法设计中挖掘用户选择更深层因素的分析，发现特定用户在不同类别内容之间更紧密的联系，就可以做到更精准的推荐。在时间线中，按照特定比例推送给用户不常接触到的信源或内容，提升多元化。算法优化还可以基于对特定用户群体提供不同内容推荐模式，比如基于地理位置和行为的分析，识别出青少年或老年人群体，自动切换推荐模式来保障特定“数字弱势”群体的内容消费安全和质量。

国内案例：

简易信息聚合（Really Simple Syndication，RSS)的各类工具和应用的普及能够帮助用户破除“信息茧房”。其技术原理是将网站、平台、APP上的内容按照用户设定的要求进行整理推送。如Telegram机器人，其基于Linux系统开发，用户可以通过开源代码自主设定需要聚合的网站、筛选和推送规则等。Tiny RSS工具同样支持自定义的内容过滤和推送规则，在订阅源管理和web端设置上十分灵活便捷，目前仍处于比较活跃的开发中。

《华尔街日报》于2016年创设了一个“红推送，蓝推送”（Red Feed, Blue Feed），将Facebook上同类内容的自由倾向、保守倾向的信息并列呈现给用户，以此提醒用户其偏向性，并推荐另一观点相左的内容,帮助用户平衡、多元化其新闻消费。

51破解信息茧房，算法推荐需要引入“父爱式”传播http://news.gmw.cn/xinxi/2019-07/04/content_32973588.htm

4.7 小结

在更广泛的数字内容治理应用中，人工智能技术依旧拥有广阔的潜力和前景。比如，目前涉及内容治理的法律规定、法律文本、裁判文书等法律资料规模宏大，而其组成的规模巨大的数字化法律数据资源已经成为非专业人士法律维权的一大门槛，人工智能技术可以在用户和内容生产者法律咨询、维权方面提供帮助；人工智能技术还能够推动各种文化载体和形式的数字化（文物、建筑、遗迹等），能够在文化存储、继承、保护、传播中发挥重要作用；在网络和实地等各类场景中定位恐怖主义和反动势力标识，为保护国家安全提供情报数据来源。

总体而言，人工智能技术在数字内容治理的内容识别、审核分级、问题感知、违法打击等各个环节下都有广阔的应用前景。企业主体不断探索在数字内容治理中人工智能技术的应用，能够有效节省人力和成本，提高效率和质量，从而增进应用主体的经济和社会效益。数字内容企业自主采用人工智能安全技术，能够清理内容审核中的“死角盲区”，减少人为误差和遗漏；人工智能技术对于简单重复性岗位的取代能够降低成本，提高效率；人类能动创造性与人工智能结合能够确保内容审核质量，建立高效的内容安全管理体系，实现真正的内容安全合规；数字内容管理和防护结合人工智能技术，可提高企业经济效益，也可用来维护自身的品牌形象。

5 总结与展望

5.1 总结

在超强算力、优质算法、海量数据和广泛连接的支持下，人工智能技术落地和产业应用的蓬勃发展，标志着一个集聚庞大的数据流、信息流、技术流和基于万物互联、跨界融合、人机共生的人工智能时代的到来。数字内容产业作为文化内容创意与数字化信息技术高度融合的新兴产业集群，在人工智能技术的助力下，显示出极大的增长空间和潜力，在用户数量、产值规模和细分领域均有不俗的势头。呈现出以内容生产的“人-机”协助、内容传播的精准匹配、内容消费的沉浸式体验、内容载体的多行业场景、内容质量的回归主导为代表的发展趋势。

在人工智能赋能下，数字内容产业在繁荣增长的过程中，同时也面临着技术、发展和社会等相互交织的各类风险。以内容爆炸下审核能力不足、算法歧视与黑箱导致的不公平、内容造假对媒体真实性和政治生态的冲击、智能生产的内容产权地位与归属不明、用户面临更多更深层的个人隐私数据泄露风险、信息茧房引发的舆论失衡和社群极化、数字内容新的竞争引发数字鸿沟等七类挑战为代表的治理难题日益突出，制约着数字内容产业进一步高质量发展。

以美、欧、英、新、日、中为代表的六个国家和地区在人工智能时代下采取了一系列数字内容治理政策，但总体而言，各国普遍表现出对传统治理路径的依赖，对各类新兴问题也以应激性的政策为主，且治理政策基于人工智能科技和数字内容文化两种视域分野明显，缺乏对于人工智能时代下数字内容治理问题的顶层规划和统筹应对，必将导致未来治理出现应对乏力现象。

在产业的安全实践和探索中，人工智能技术可以应用于内容审核、事实核查、版权保护、打击诈骗、舆情治理、破除信息茧房等数字内容治理场景，提高治理效率和质量，保障数字内容产业安全发展。

5.2 展望

人工智能是信息时代未来技术发展的主流方向，而数字内容将成为信息时代人类文化的主要形态，两者的结合是前沿科技与创新文化的深度融合，人工智能在数字内容领域的应用和赋能将成为未来数字经济时代最重要的特征之一，数字内容产业也将成为人工智能时代的核心产业之一。

5.2.1 人工智能技术将继续促进数字内容产业繁荣发展

人工智能技术已显示出其在数字内容生产、传播、市场、审核等各环节中解放生产力和提高产能方面的巨大潜能，随着技术难点的攻克和算法模型不断优化，势必将发挥更大的促进和赋能作用。

内容生产上，人工智能技术将深入到选题、采编、写作、审核等多个流程，实现智能化全链衔接的同时保证内容质量可控；内容传播上，人工智能技术将进一步降低内容消费“噪音”；内容市场上，随着人工智能技术与和行业的深度绑定，将解放更多内容生产者，为内容创新注入更为优质的生产力，大大拓宽数字内容市场容积；内容审核上，人工智能技术将继续保持在海量信息和高速内容传播下高效审核中的优势作用。

5.2.2 人工智能时代下数字内容机遇与挑战将长期并存

未来很长时间，人们将面临数字内容固有顽疾更加复杂的现实，还将应对由人工智能本体安全性和误用滥用产生的新问题，但是，人工智能技术本身的不断成熟优化以及在各个治理场景中的应用潜能，又将会给数字内容治理提供更高效的技术支持和创新思路。因此，人工智能技术集合对于数字内容的赋能作用呈螺旋式发展，机遇与挑战将长期并存。

5.2.3 综合统筹社会协同，形成数字内容治理良好环境

人工智能时代的数字内容治理靠传统治理模式或单一治理手段绝对行不通，应当强化综合统筹治理、协同参与的思路。国家需从战略规划层面高度重视，健全完善垂直领域数字内容政策法规体系、明确各方主体责任；充分调动社会各界力量在产业规划、标准制定、行业自律中发挥作用；促进人工智能与其他学科、行业的跨界沟通和交流，加强科普和舆论引导，推动全民数字内容技能和数字内容素养提升；引领相关企业和科研人员秉持向善之心，践行科技向善理念，追求至善，不仅仅局限于把技术和产品开发出来，而是更多地思考技术和产品开发可能产生的影响，提高风险研判水平，肩负起对科技、对社会的责任，确保未来的正确方向。

5.2.4 大力推动国际合作，构建数字内容治理国际机制

人工智能的技术创新具有改变人类基本规则的可能性，对于全球都有着深刻影响，在其赋能下的数字内容发展问题更是全球性的治理议题，这要求在进行人工智能时代数字内容治理时，必须要大力推进国际合作，构建国际协同治理机制。

目前，部分发达国家处于人工智能和数字内容发展的引领地位，其他国家则被技术发展的趋势所裹挟，各国在数字内容发展中的获益情况具有较大差异。通过加强国际合作，才能够应对数字鸿沟带来的国际不公平现象。第一，多边参与：要基于多边参与的立场下构建数字内容的国际治理机制。第二，明确原则：国际社会要在数字内容治理议题中确定基本的价值原则和规范，明确发达国家和发展中国家不同的责任义务。第三，平衡发展：平衡数字内容全球化发展和各国传统文化传承两条主线，尊重每个文明、每个国家的数字内容发展权利。

5.2.5 探索创新应用落地，形成“人-机”良好协同路径

企业本身将在产业应用和治理实践中直面人工智能技术的不足，同时也必然承担起数字内容产业发展中对应的社会责任和法律义务。这些将对作为数字内容治理实践者的企业提出更高要求，要大胆探索，稳妥推进。

第一，不断积极探索：企业要积极探索布局在各种领域人工智能技术的治理应用。在未来，内容治理合规将成为各国普遍要求，利用人工智能技术提高内容合规的质量和效率，能够在市场竞争中占据主动；第二，寻求技术突破：人工智能企业要主动寻求技术突破，提供高质量的自动化工具和针对性的解决方案。数字内容治理将是人工智能产业未来重大应用领域，主动布局能够形成人工智能产业的先发优势；第三，形成良好的“人-机”协作：目前来看，人工智能技术短期内无法完全替代人类在关键时机进行决策，人工与机器智能将长期并存，良好协调的“人机合作”模式能够最大限度地发挥人类决策和机器智能的相对优势，企业要针对人工智能技术本身的局限，寻求良好协调的“人机合作”模式。

6 附录

6.1 近年来主要国家和地区人工智能战略或政策文件列表

国家	时间	主要战略和政策
美国	2016 年 10 月	《为人工智能的未来做准备》
	2016 年 10 月	《国家人工智能研究和发展战略计划》
	2016 年 12 月	《人工智能、自动化与经济报告》
	2019 年 2 月	《美国人工智能倡议》
	2019 年 2 月	《2018 年国防部人工智能战略摘要——利用人工智能促进安全与繁荣》
中国	2016 年 5 月	《“互联网+人工智能三年行动实施方案》
	2017 年 7 月	《新一代人工智能发展规划》
	2017 年 12 月	《促进新一代人工智能产业发展三年行动计划(2018-2020 年)》
欧盟	2018 年 4 月	《欧盟人工智能》
	2018 年 12 月	《人工智能协调计划》
	2018 年 12 月	《可信人工智能伦理指南（草案）》
英国	2016 年 10 月	《机器人技术和人工智能》
	2016 年 11 月	《人工智能：未来决策的机会与影响》
	2017 年 10 月	《发展英国人工智能产业》
	2018 年 4 月	《人工智能行业新政》
德国	2018 年 7 月	《联邦政府人工智能战略要点》
	2018 年 11 月	《人工智能战略》
法国	2013 年 5 月	《法国机器人发展计划》
	2017 年 3 月	《法国人工智能战略》
	2018 年 5 月	《法国与欧洲人工智能战略研究报告》
西班牙	2019 年 3 月	《西班牙人工智能研究、发展与创新战略》
荷兰	2019 年 3 月	《荷兰人工智能伦理规范》
丹麦	2019 年 3 月	《人工智能国家战略》
芬兰	2017 年 12 月	《芬兰的人工智能时代》
	2018 年 6 月	《四个视角下的人工智能：经济，就业，知识和伦理》
瑞典	2018 年 5 月	《国家重点关注人工智能》
加拿大	2017 年 3 月	《泛加拿大人工智能战略》
日本	2015 年 1 月	《机器人新战略》
	2016 年 7 月	《下一代人工智能促进战略》
	2017 年 3 月	《人工智能技术战略》
韩国	2018 年 5 月	《人工智能研发战略》
新加坡	2017 年 5 月	《新加坡人工智能战略》
	2018 年 6 月	《人工智能治理和道德的三个新倡议》
	2019 年 1 月	《人工智能治理框架》
越南	2018 年 10 月	《决定发布实施“2025 年人工智能研究与开发”的计划》
印度	2018 年 6 月	《国家人工智能战略》
阿联酋	2017 年 10 月	《人工智能战略》
墨西哥	2018 年 6 月	《迈向墨西哥的人工智能战略：利用人工智能革命》

6.2 各主要国家和地区关于数字内容产业界定

【美国】

美国将数字内容产业概念界定为信息产业（Information Industry），在2017年北美产业分类标准(North American Industry Classification System，NAICS)中的信息产业分类中包括了出版业、电影和录音业、广播和传播业、电信、数据处理和相关服务、其他信息服务业（包括新闻集团、网络出版、广播和搜索）。

【欧盟】

1996 年，欧盟在《信息社会2000计划》中提出了“数字内容产业”概念：制造、开发、包装和销售信息产品及其服务的产业，其产品范围包括各种媒介的印刷品（书报杂志等）、电子出版物（联机数据库、音像服务、光盘服务和游戏软件等）和音像传播（影视、录像和广播等）等多个领域。

2002年，爱尔兰政府出台的《爱尔兰数字内容产业发展战略》将数字内容产业定义为创建、设计、管理和销售数字产品和服务以及为上述活动提供技术支持的产业。包括数字游戏，以计算机和网络为基础的学习、虚拟教室、数字化协作的数字学习，与娱乐、预订旅行、财务交易、定位服务有关的商业与客户的电信/无线服务，主要用于科学与工业的高端成像、设计和虚拟现实工具与应用的非媒体应用等。

【英国】

1998年，由英国多个政府部门和产业界代表组成的创意产业工作组在《创意产业专题报告》中，首次提出创意产业（Creative Industries），包括广告、建筑艺术、艺术和古董市场、手工艺品、时尚设计、电影录像、交互式互动软件、音乐、表演艺术、出版业、软件及计算机服务、电视广播和设计。

【日本】

日本将数字内容产业定义为“内容产业”，2004 年正式公布了《内容产业促进法》，2010年《内容产业白皮书》将日本数字内容产业分成图书报刊（图书、报纸、图片、文本等）、影像、音乐、游戏四大类。

【中国】

我国除了“数字内容产业”外，也有“信息内容服务业”“信息内容产业”“文化创意产业”等多种概念。《2008-2009上海数字内容产业白皮书》则建议把数字内容产业细分为网络游戏、数字动漫、数字出版、数字学习、移动内容、数字视听、其他网络服务和内容软件八大类。

《北京市文化创意产业分类标准》界定的文化创意产业包括 9 个大类 88 个小类，主要包括九大行业:文化艺术、新闻出版、广播、电视、电影、软件、网络及计算机服务、广告会展、艺术品交易、设计服务、旅游、休闲娱乐以及其他辅助服务。《国家“十一五”时期文化发展规划纲要》提出的文化创意产业主要包括文化科技、影视制作、音乐制作、时尚设计、艺术创作、工艺美术、广告创意、动漫游戏等。

腾讯安全战略研究中心

Tencent Security and Strategy Research Center