

人工智能生成内容 (AIGC) 白皮书 (2022 年)

中国信息通信研究院

京东探索研究院

2022年9月

版权声明

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



前 言

习近平总书记曾指出，“数字技术正以新理念、新业态、新模式全面融入人类经济、政治、文化、社会、生态文明建设各领域和全过程”。在当前数字世界和物理世界加速融合的大背景下，人工智能生成内容（Artificial Intelligence Generated Content，简称 AIGC）正在悄然引导着一场深刻的变革，重塑甚至颠覆数字内容的生产方式和消费模式，将极大地丰富人们的数字生活，是未来全面迈向数字文明新时代不可或缺的支撑力量。

本白皮书重点从 AIGC 技术、应用和治理等维度进行了阐述。在技术层面，梳理提出了 AIGC 技术体系，既涵盖了对现实世界各种内容的数字化呈现和增强，也包括了基于人工智能的自主内容创作。在应用层面，重点分析了 AIGC 在传媒、电商、影视等行业和场景的应用情况，探讨了以虚拟数字人、写作机器人等为代表的新业态和新应用。在治理层面，从政策监管、技术能力、企业应用等视角，分析了 AIGC 所暴露出的版权纠纷、虚假信息传播等各种问题。最后，从政府、行业、企业、社会等层面，给出了 AIGC 发展和治理建议。由于人工智能仍处于飞速发展阶段，我们对 AIGC 的认识还有待进一步深化，白皮书中存在不足之处，敬请大家批评指正。

目 录

一、人工智能生成内容的发展历程与概念.....	1
(一) AIGC 历史沿革.....	1
(二) AIGC 的概念与内涵.....	4
二、人工智能生成内容的技术体系及其演进方向.....	7
(一) AIGC 技术升级步入深化阶段.....	7
(二) AIGC 大模型架构潜力凸显.....	10
(三) AIGC 技术演化出三大前沿能力.....	18
三、人工智能生成内容的应用场景.....	26
(一) AIGC+传媒：人机协同生产，推动媒体融合.....	27
(二) AIGC+电商：推进虚实交融，营造沉浸体验.....	29
(三) AIGC+影视：拓展创作空间，提升作品质量.....	32
(四) AIGC+娱乐：扩展辐射边界，获得发展动能.....	35
(五) AIGC+其他：推进数实融合，加快产业升级.....	37
四、人工智能生成内容发展面临的问题.....	38
五、发展建议与展望.....	43
(一) 发展建议.....	43
(二) 未来展望.....	48

图目录

图 1 AIGC 发展历程	4
图 2 AIGC 多模态大模型生成结果图	17
图 3 OpenAI AIGC 多模态大模型 DALL E 2 生成结果图	18
图 4 AIGC 的三大前沿能力	19
图 5 AIGC 应用视图	27



一、人工智能生成内容的发展历程与概念

1950 年，艾伦·图灵（Alan Turing）在其论文《计算机器与智能（Computing Machinery and Intelligence）》中提出了著名的“图灵测试”，给出了判定机器是否具有“智能”的试验方法，即机器是否能够模仿人类的思维方式来“生成”内容继而与人交互。某种程度上来说，人工智能从那时起就被寄予了用于内容创造的期许。经过半个多世纪的发展，随着数据快速积累、算力性能提升和算法效力增强，今天的人工智能不仅能够与人类进行互动，还可以进行写作、编曲、绘画、视频制作等创意工作。2018 年，人工智能生成的画作在佳士得拍卖行以 43.25 万美元成交，成为世界上首个出售的人工智能艺术品，引发各界关注。随着人工智能越来越多地被应用于内容创作，人工智能生成内容（Artificial Intelligence Generated Content，简称 AIGC）的概念悄然兴起。

（一）AIGC 历史沿革

结合人工智能的演进历程，AIGC 的发展大致可以分为三个阶段，即：早期萌芽阶段（20 世纪 50 年代至 90 年代中期）、沉淀积累阶段（20 世纪 90 年代中期至 21 世纪 10 年代中期），以及快速发展阶段（21 世纪 10 年代中期至今）。

早期萌芽阶段（1950s-1990s），受限于当时的科技水平，AIGC 仅限于小范围实验。1957 年，莱杰伦·希勒（Lejaren Hiller）和伦纳

德·艾萨克森（Leonard Isaacson）通过将计算机程序中的控制变量换成音符完成了历史上第一支由计算机创作的音乐作品——弦乐四重奏《依利亚克组曲（Illiac Suite）》。1966年，约瑟夫·魏岑鲍姆（Joseph Weizenbaum）和肯尼斯·科尔比（Kenneth Colby）共同开发了世界第一款可人机对话的机器人“伊莉莎（Eliza）”，其通过关键字扫描和重组完成交互任务。80年代中期，IBM基于隐形马尔科夫链模型（Hidden Markov Model, HMM）创造了语音控制打字机“坦戈拉（Tangora）”，能够处理约20000个单词。80年代末至90年代中，由于高昂的系统成本无法带来可观的商业变现，各国政府纷纷减少了在人工智能领域的投入，AIGC没有取得重大突破。

沉淀积累阶段（1990s-2010s），AIGC从实验性向实用性逐渐转变。2006年，深度学习算法取得重大突破，同时期图形处理器（Graphics Processing Unit, GPU）、张量处理器（Tensor Processing Unit, TPU）等算力设备性能不断提升，互联网使数据规模快速膨胀并为各类人工智能算法提供了海量训练数据，使人工智能发展取得了显著的进步。但是AIGC依然受限于算法瓶颈，无法较好地完成创作任务，应用仍然有限，效果有待提升。2007年，纽约大学人工智能研究员罗斯·古德温装配的人工智能系统通过对公路旅行中的一切所见所闻进行记录和感知，撰写出小说《1 The Road》。作为世界第一部完全由人工智能创作的小说，其象征意义远大于实际意义，整体可读性不强，拼写错误、辞藻空洞、缺乏逻辑等缺点明显。2012年，微软公开展示

了一个全自动同声传译系统，基于深层神经网络（Deep Neural Network, DNN）可以自动将英文演讲者的内容通过语音识别、语言翻译、语音合成等技术生成中文语音。

快速发展阶段（2010s-至今），自 2014 年起，随着以生成式对抗网络（Generative Adversarial Network, GAN）为代表的深度学习算法的提出和迭代更新，AIGC 迎来了新时代，生成内容百花齐放，效果逐渐逼真直至人类难以分辨。2017 年，微软人工智能少女“小冰”推出了世界首部 100% 由人工智能创作的诗集《阳光失了玻璃窗》。2018 年，英伟达发布的 StyleGAN 模型可以自动生成图片，目前已升级到第四代模型 StyleGAN-XL，其生成的高分辨率图片人眼难以分辨真假。2019 年，DeepMind 发布了 DVD-GAN 模型用以生成连续视频，在草地、广场等明确场景下表现突出。2021 年，OpenAI 推出了 DALL-E 并于一年后推出了升级版本 DALL-E-2，主要应用于文本与图像的交互生成内容，用户只需输入简短的描述性文字，DALL-E-2 即可创作出相应极高质量的卡通、写实、抽象等风格的绘画作品。



图 1 AIGC 发展历程

（二）AIGC 的概念与内涵

目前，对 AIGC 这一概念的界定，尚无统一规范的定义。国内产学研各界对于 AIGC 的理解是“继专业生成内容（Professional Generated Content, PGC）和用户生成内容（User Generated Content, UGC）之后，利用人工智能技术自动生成内容的新型生产方式”。在国际上对应的术语是“人工智能合成媒体（AI-generated Media 或 Synthetic Media）”¹，其定义是“通过人工智能算法对数据或媒体进行生产、操控和修改的统称”。综上所述，我们认为 AIGC 既是从内容生产者视角进行分类的一类内容，又是一种内容生产方式，还是用于内容自动化生成的一类技术集合。本白皮书主要聚焦于 AIGC 含义

¹ 维基百科：“人工智能合成媒体（AI-generated Media 或 Synthetic Media）”
https://en.wikipedia.org/wiki/Synthetic_media

中的技术部分。

为了帮助不同领域的受众群体更好的理解 AIGC，我们从发展背景、技术能力、应用价值三个方面对其概念进行深入剖析。

从发展背景方面来看，AIGC 的兴起源于深度学习技术的快速突破和日益增长的数字内容供给需求。一方面，技术进步驱动 AIGC 可用性不断增强。在人工智能发展初期，虽然对 AIGC 进行了一些初步尝试，但受限各种因素，相关算法多基于预先定义的规则或者模板，还远远算不上是智能创作内容的程度。近年来，基于深度学习算法的 AIGC 技术快速迭代，彻底打破了原先模板化、公式化、小范围的局限，可以快速、灵活地生成不同模态的数据内容。另一方面，海量需求牵引 AIGC 应用落地。随着数字经济与实体经济融合程度不断加深，以及 Meta、微软、字节跳动等平台型巨头的数字化场景向元宇宙转型，人类对数字内容总量和丰富程度的整体需求不断提高。数字内容的生产取决于想象能力、制造能力和知识水平；传统内容生产手段受限于人力有限的制造能力，逐渐无法满足消费者对于数字内容的消费需求，供给侧产能瓶颈日益凸显。基于以上原因，AIGC 在各行业中得到越来越广泛的应用，市场潜力逐渐显现。

从技术能力方面来看，AIGC 根据面向对象、实现功能的不同可分为三个层次。一是智能数字内容孪生，其主要目标是建立现实世界到数字世界的映射，将现实世界中的物理属性（如物体的大小、纹理、颜色等）和社会属性（如主体行为、主体关系等）高效、可感知地进

行数字化。**二是智能数字内容编辑**，其主要目的是建立数字世界与现实世界的双向交互。在数字内容孪生的基础上，从现实世界实现对虚拟数字世界中内容的控制和修改，同时利用数字世界高效率仿真和低成本试错的优势，为现实世界的应用提供快速迭代能力。**三是智能数字内容创作**，其主要目标是让人工智能算法具备内容创作和自我演化的能力，形成的 AIGC 产品具备类似甚至超越人的创作能力。以上三个层面的能力共同构成 AIGC 的能力闭环。

从应用价值方面来看，AIGC 将有望成为数字内容创新发展的新引擎，为数字经济发展注入全新动能。一方面，AIGC 能够以优于人类的制造能力和知识水平承担信息挖掘、素材调用、复刻编辑等基础性机械劳动，从技术层面实现以低边际成本、高效率的方式满足海量个性化需求；同时能够创新内容生产的流程和范式，为更具想象力的内容、更加多样化的传播方式提供可能性，推动内容生产向更有创造力的方向发展。另一方面，AIGC 能够通过支持数字内容与其他产业的多维互动、融合渗透从而孕育新业态新模式，打造经济发展新增长点，为千行百业发展提供新动能。此外，2021 年以来，“元宇宙”呈现出超出想象的发展爆发力；作为数实融合的“终极”数字载体，元宇宙将具备持续性、实时性、可创造性等特征，也将通过 AIGC 加速复刻物理世界、进行无限内容创作，从而实现自发有机生长。

二、人工智能生成内容的技术体系及其演进方向

AIGC 作为人工智能技术和产业应用的要素之一，随着技术能力的不断迭代升级，正在降低内容创作门槛、释放创作能力，未来将推动数实融合趋势下内容创作的范式转变。探讨其能力体系的构成，即赋能内容创作的技术路径，对制定领域内标准、建立行业生态、争取更加广泛的开发者和应用场景具有十分重要的意义。

本部分从技术驱动的视角出发，对 AIGC 的能力体系进行归纳和推理，展示现有技术应用和其背后技术演化整体进程。第一节首先从技术趋势的角度，提出 AIGC 的技术创新已经完成由传统方法向深度学习过渡的应用创新阶段，并逐步深化到学习范式和网络结构方面的理论创新阶段。第二节则重点分析前沿理论多模态大模型方面的突破，让 AIGC 进行跨模态融合性创新成为可能，也给予了 AIGC 前所未有的产业空间与实践潜力。第三节进一步归纳总结在前沿技术驱动下，AIGC 赋能内容创作的三大能力，并对三大能力的技术演化路径进行展望。

（一）AIGC 技术升级步入深化阶段

人工智能算法的不断迭代是 AIGC 发展进步的源动力，从技术演进的角度出发，可将 AIGC 技术可大致划分为传统基于模板或规则的前深度学习阶段和深度神经网络快速发展的深度学习阶段。

早期的 AIGC 技术主要依据事先指定的模板或者规则，进行简单的内容制作与输出，与灵活且真实的内容生成有较大的差距。该时期

的人工智能算法并不具备强大的学习能力，而是大多依赖于预先定义的统计模型或专家系统执行特定的任务。通过巧妙地规则设计，早期 AIGC 技术可以完成简单线条、文本和旋律的生成。例如，通过定义复杂的函数方程组，计算机所绘出的函数曲线具备某种美学图样；通过记录大量的问答文本，在面对新的问题时，计算机可以通过检索和匹配的方式生成简单的答案，甚至于改写故事。但是由于缺乏对客观世界的深入感知和对人类语言文字等知识的认知能力，早期的 AIGC 技术普遍面临所生成的内容空洞、刻板、文不对题等问题。参考人类的内容创作过程，研究人员们提出，理想的 AIGC 算法需要具备对数据内容的学习能力，在理解数据的基础上进行知识与分布的学习，最终实现高质量的内容创作。

神经网络在学习范式²和网络结构上的不断迭代极大的提升了人工智能算法的学习能力，从而推动了 AIGC 技术的快速发展。不同于传统人工智能算法，深度学习中的损失函数和梯度下降算法可以灵活快速的调整神经网络中的参数，从而实现从数据中进行学习功能。2012 年，卷积神经网络 AlexNet^[1]凭借优秀的学习能力，在当年的 ImageNet 大规模视觉识别挑战赛中一举夺魁，比第二名传统机器学习算法的错误率提升 10.8 个百分点，开启了深度学习时代的序幕。就在紧随其后的 2013 年，深度变分自编码器^[2]的提出让 AIGC 技术能力有了极大的进步。对于给定的神经网络，深度变分自编码器要

² 人工智能的学习范式是指人工智能模型从数据中进行学习的方法。

求网络的输出是对于输入内容的重建，通过重参数化等技巧，网络在重建过程中学习训练数据的统计分布。在测试阶段，变分自编码器通过在学习到的统计分布中进行采样，首次能比稳定的生成从未观测过的低分辨率图像。2014年，一种新的博弈学习范式伴随着生成对抗网络^[3]被提出。生成对抗网络由一个生成器和一个判别器组成，判别器致力于不断寻找生成数据和真实数据间的不同，生成器根据判别器的反馈不断完善自身，以求生成真假难辨的内容。得益于双方博弈的学习策略，生成内容的真实性和清晰度都得到了极大的提升，生成对抗网络也被应用于很多内容生成的具体应用。除了变分自编码器和生成对抗网络，强化学习^[4]、流模型^[5]、扩散模型^[6]等学习范式均取得了喜人的进展，这些模型范式在不同场景中各有优势，让 AIGC 技术可以快速地应用到不同的场景和任务中。

深度神经网络的结构升级是推动 AIGC 快速发展的另一主要因素。一方面，实验证明，深度学习神经网络的学习能力和模型大小呈正相关，伴随着模型参数量的增加，相对应深度学习神经网络的能力一般会取得大幅提升。但是，随意地增加神经网络规模是行不通的，越大规模神经网络往往意味着更大的训练难度，因此深度神经网络的结构设计显得尤为关键。从早期的玻尔兹曼机，多层感知机，卷积神经网络，到深度残差网络和 Transformer 大模型，网络结构进化带来了深度学习模型参数量从几万到数千亿跃升，模型层数也从开始的个位数逐步发展到成百上千。深度学习模型规模上的量变引起了 AIGC 技术能力

的质变，在新型网络结构的加持下，上述的生成对抗网络等算法开始能生成超高清晰度的视频，高质量的文本段落和优美灵动的乐曲旋律。另一方面，研究者在深度神经网络结构的设计中引入包含语义的隐式表达和物理知识，以降低模型的训练难度、增强生成内容的丰富程度。例如，研究者发现通过在神经网络的每一层引入隐式表达，能够极大地提升内容生成算法的可控性和生成效果^[7]。另外，在三维数据的生成任务中，神经辐射场^[8]在网络结构设计时充分考虑了物理世界的固有约束，极大提升了三维渲染效率和效果。

AIGC 要真正发挥对不同行业的驱动作用，需要与各行各业的特异性场景深度融合。在处理这些实际应用中，深度学习算法在感知、认知、模仿、生成等方向的基础能力决定了 AIGC 技术所能创作的生产力。近些年中，这些算法技术齐头并进、百花齐放，并最终形成了 AIGC 应用于不同场景的底层支撑。通过人工智能支撑技术的不断升级，AIGC 技术将持续赋能各类文化创意、生产生活、科学发现^[9,10]等各种场景。

（二）AIGC 大模型架构潜力凸显

超级深度学习近年来的快速发展带来了神经网络技术在大模型和多模态两个方向上的不断突破，并为 AIGC 技术能力的升级提供了强力的支撑和全新的可能性。当前 AIGC 技术已经从最初追求生成内容的真实性的基本要求，发展到满足生成内容多样性、可控性的进阶需求，并开始追求生成内容的组合性。数字内容的组合性一方面

关注复杂场景、长文本等内容中各个元素的组合，例如虚拟数字世界中人、物和环境间的交互并组合生成为整体场景；长篇文字内容用词、语句、段落间的相互呼应和组合。另一方面，组合性追求概念、规则等抽象表达的组合，以此完成更加丰富和生动的数字内容生成，这些新出现的需求对传统单一模态的人工智能算法框架提出了新的挑战。近年来，研究界在大规模深度网络、多模态人工智能方面的探索表明大模型具备易扩展性，能够实现跨模态的知识沉淀，以大模型为基础模型，通过大模型小型化技术使得人工智能在小数据集场景下也能具备优秀的理解、生成和泛化能力，具有超大规模、超多参数量的多模态大型神经网络将引领 AIGC 技术升级正在成为学界、产业界共识³。

1. 视觉大模型提升 AIGC 感知能力

以图像、视频为代表的视觉数据是互联网时代信息的主要载体之一，这些视觉信息时刻记录着物理世界的状态，并在不断传播和再创作的过程中，反映人的想法、观念和价值主张。赋以人工智能模型感知并理解这些海量的视觉数据的能力^[1]，是实现人工智能生成数字内容、数字孪生的基础；感知能力的提升，是实现生成视觉内容语义明确、内涵丰富、效果逼真的前提。

针对视觉信息的感知研究，在传统机器学习时代主要基于科研人员手动建模的特征和基于统计学习理论构建的朴素分类器，例如支持

³ 百度文心大模型：<https://wenxin.baidu.com/>；OpenAI DALL-E 2 大模型：<https://openai.com/dall-e-2/>；智源研究院大模型：<https://mp.weixin.qq.com/s/j8q018Lck1TWHO3NxQDiJQ>

向量机模型（SVM），其能完成的任务类型和感知能力都非常有限；在深度学习时代，主要基于深度神经网络模型，例如深度残差网络（ResNet），其数据驱动的端到端学习范式使得模型的感知能力有了显著提升，在工业界也得到广泛的应用。但是，这类模型往往针对单一感知任务进行设计，很难同时完成多种视觉感知任务。如何解决不同场景、环境和条件下的视觉感知问题，并实现鲁棒、准确、高效的视觉理解，是 AIGC 技术必须要解决的挑战。

以视觉 Transformer（ViT，一种神经网络模型）^[12] 为代表的新型神经网络，因其优异的性能、模型的易扩展性、计算的高并行性，正在成为视觉领域的基础网络架构，并且逐渐发展出来十亿甚至百亿参数规模的模型。在过去的 2-3 年间，视觉感知和理解技术正迎来突飞猛进的发展。无监督学习技术，包括对比式自监督学习（例如 SimCLR 和 MoCo 系列技术）和生成式自监督学习（例如 MAE 技术），能够大幅降低训练模型所需的有标注数据的数量。经过无监督预训练的神经网络模型，仅需要在少量的有标注样本上经过微调学习，即可在多种场景，线上线下均取得优异的性能。近年来基于 Transformer 衍生出来一系列网络结构，例如 Swin Transformer^[13]、ViTAE Transformer^[14,15]。通过将人类先验知识引入网络结构设计，使得这些模型具有了更快的收敛速度、更低的计算代价、更多的特征尺度、更强的泛化能力，从而能更好地学习和编码海量数据中蕴含的知识。这些新型的大模型架构，通过无监督预训练和微调学习的范式，

在图像分类、目标检测、语义分割、姿态估计、图像编辑以及遥感图像解译等多个感知任务上取得了相比于过去精心设计的多种算法模型更加优异的性能和表现^[16,17]，有望成为基础视觉模型（Foundation Vision Model），显著提升场景感知能力，助力 AIGC 领域的发展。

基于视觉 Transformer 完成多种感知任务的联合学习是目前的研究热点。通过探索不同任务关联关系，挖掘丰富的监督信号，能够促使模型学习到更具泛化能力和可被理解的特征表示。此外，联合文本、语音等不同模态数据进行联合学习，探索不同模态数据的语义关联和信息互补，也是训练视觉大模型的重要路径。由此得到的视觉基础大模型在环境感知、内容检索、语义理解、模态对齐等任务上具备先天的优势，对于提升 AIGC 基础环境孪生能力、丰富 AIGC 应用场景具有重要价值。

2. 语言大模型增强 AIGC 认知能力

作为人类文明的重要记录方式，语言和文字记录了人类社会的历史变迁、科学技术和知识文化等。利用人工智能技术对海量语言、文本数据进行信息挖掘和内容理解是 AIGC 技术的关键一环。一方面，语言模型的训练和学习是进行文本生成的核心基础；另一方面，学习并理解人类语言将大幅丰富数字内容的生产能力，创新、丰富数字内容的生产方式，例如构建低门槛创作工具，使用户通过语言描述就能完成例如语言定位、语言编辑等高阶编辑操作。

在如今信息复杂的场景中，数据质量参差不齐、任务种类多，导

致数据孤岛和模型孤岛的存在，传统自然语言处理技术的不足尤为明显：模型设计、部署困难；数据难以复用；难以学习海量无标签数据挖掘、知识提取的共性能力。

对于传统自然语言处理技术的普遍问题，基于语言的大模型技术可以充分利用海量无标注文本进行预训练，从而赋予文本大模型在小数据集、零数据集场景下的理解和生成能力。基于大规模预训练的言模型不仅能够在情感分析、语音识别、信息抽取、阅读理解等文本理解场景中表现出色，而且同样适用于图片描述生成、广告生成、书稿生成、对话生成等文本生成场景。这些复杂的功能往往只需要通过简单的无标注文本数据收集，训练部署一个通用的大规模预训练模型即可实现。研究者们相信基于语言的认知智能可以更快的加速通用人工智能的到来。例如，谷歌和 OpenAI 分别提出大规模预训练模型 BERT^[18] 和 GPT^[19]，在诸多自然语言理解和生成任务上取得了突破性的性能提升，验证了大模型在零资源、小样本、中低资源场景的优越性。紧随其后，国内外知名企业和高校均投入非常大的人力、算力、数据于自然语言处理大模型的研发，包括谷歌、微软、Meta、清华大学、斯坦福大学、华盛顿大学、卡内基·梅隆大学、京东、华为、百度等等。模型参数量也从最初的千万级发展到了千亿级别^[20]，训练代价也从数十天增长到了不容忽视的几十万天（按在单张 V100 上计算）。

显然，指数级增长的成本换取的微弱增益让人们意识到，如何设计更效率的自监督学习方法、更高参数效用比的模型架构、更绿色

节能的训练框架成为了大模型未来方向之一。在这个方向上，诸多机构开始了高效绿色的大模型探索之路，并且取得了显著的效果，如通用语言理解评估基准（GLUE）目前（2022年6月）在榜第一名的是由京东探索研究院研发的 Vega v1 织女模型⁴，依托于预训练阶段多种文本粒度、语种类型、负采样方式上的自监督学习创新，实现了高效的数据知识提取，并采用了有理论支撑的更快捷的分布式优化器。此外，超级深度学习模型可以通过非常低成本的微调快速适应新的产业、领域、行业，实现跨模态、全链路的知识积累、沉淀、传播、复用。

基于语言的超级深度学习技术的发展趋势主要体现在训练模型的数据量日益增大、数据种类也更加丰富，模型规模增大、参数量以指数倍增加。通过不断构建语义理解能力增强、逻辑知识可抽象学习、同时适用于多种任务的语言大模型，将会对 AIGC 场景中的各项认知应用产生极大价值。

3. 多模态大模型升级 AIGC 内容创作能力

在日常生活中，视觉和语言是最常见且重要的两种模态^[21]，上述的视觉大模型可以构建出人工智能更加强大的环境感知能力，而语言大模型则可以学习到人类文明的抽象概念以及认知的能力。然而 AIGC 技术如果只能生成单一模态的内容，那么 AIGC 的应用场景将极为有限、不足以推动内容生产方式的革新。多模态大模型的出现，

⁴ <https://gluebenchmark.com/leaderboard>

则让融合性创新成为可能，极大丰富了 AIGC 技术可应用的广度。对于包含多个模态的信息，多模态大模型则致力于处理不同模态、不同来源、不同任务的数据和信息，从而满足 AIGC 场景下新的创作需求和应用场景。

多模态大模型拥有两种能力，一个是寻找到不同模态数据之间的对应关系，例如将一段文本和与之对应的图片联系起来；另一个是实现不同模态数据间的相互转化与生成，比如根据一张图片生成对应的语言描述。为了寻找到不同模态数据之间的对应关系，多模态大模型将不同模态的原始数据映射到统一或相似语义空间当中，从而实现不同模态的信号之间的相互理解与对齐，这一能力最常见的例子就是互联网中使用文字搜索与之相关图片的图文搜索引擎。在此基础上，多模态大模型可以进一步实现不同模态数据间的相互转化与生成，这一能力是进行 AIGC 原生创作的关键。



来源：京东探索研究院

图 2 AIGC 多模态大模型生成结果图

如图 2 所示，只需给定用户简单手绘的语义图或是素描图，多模

态大模型学习模型便能够创作出逼真的风景图像，同时，当给定具体文本语义时，图像中的内容也将随之改变，展现出不同的季节亦或是“黄昏时河道干涸”的场景。再以 OpenAI 最新提出的多模态大模型 DALL-E 2 为例，给定一个已有的场景图像，该模型能够在指定位置添加指定的目标主体，如图 3 所示，当要求在沙发上（位置 3 处）添加一只柯基狗时，算法可以在指定位置添加不同形态的真实的柯基；当要求在左侧画框中（位置 1 处）添加一只柯基时，算法先是成功的识别出该位置是一幅画，并创作了符合相应画风的柯基狗⁵。基于多模态大模型，AIGC 具备了更加接近于人类的创作能力，并真正的开始展示出代替人类进行内容创作，进一步解放生产力的潜力。



来源：OpenAI

图 3 OpenAI AIGC 多模态大模型 DALL E 2 生成结果图

对于人工智能而言，能够高质量的完成多模态数据的对齐、转换

⁵ <https://openai.com/dall-e-2/>

和生成任务意味着模型对物理世界具备了极为深刻的理解。从某种程度而言，基于多模态大模型的 AIGC 是人工智能算法迈向通用人工智能的重要一步。就好像人类通过不断的对比试错、总结归纳来了解我们身处的物理世界一样，多模态 AIGC 大模型也有希望能够自行总结客观规律，发展出认知与常识，进而帮助人类创造出新的数字世界。

（三）AIGC 技术演化出三大前沿能力

AIGC 技术被广泛应用于音频、文本、视觉等不同模态数据，并构成了丰富多样的技术应用。本节归纳 AIGC 变革内容创作方式的三大前沿能力（如图 4 所示），分别是智能数字内容孪生能力，智能数字内容编辑能力和智能数字内容创作能力。



来源：京东探索研究院

图 4 AIGC 的三大前沿能力

1. 增强与转译构建数字内容孪生能力

内容数字化是现今所有数字系统得以存在和运转的前提，其过程

是指将视觉、声音、文本等信息转化为数字格式。传统的数字化主要关注对传感器所采集数据的客观记录和储存，但容易忽略所记录的内容本身的完整性和相关语义。相比于传统的内容数字化，智能数字内容孪生技术致力于进一步挖掘数据中的有效信息，在深入理解数据内容的基础上，实现一系列高效、准确、智能的数字内容孪生任务。作为传统数字化的扩充和升级，数字内容的孪生技术受到了持续且广泛的研究。

智能数字内容孪生可大致分为智能增强技术和智能转译技术两个主要分支。考虑现实场景中数据采集、传输和储存中可能遇到的多种限制，原始的数字内容经常会存在缺失或者损坏等问题。智能增强技术旨在消除上述过程中的干扰和缺失问题，根据给定的低质量原始数据生成经过增强后的高质量数字内容，力求在数字世界中孪生并重构完整逼真的客观世界。在计算机视觉任务中，智能增强技术多被用于修复并增强由采集设备或环境因素引起的视觉内容受损，例如低分辨率、模糊、像素缺失等。同理，对于有缺陷的文本和音频数据，相关的智能增强技术被用于解决片段缺失、脉冲干扰和音频失真等问题，在实际生产生活中为相关应用生成复原高质量的数字内容。

除了对各种模态数据内容的修复和增强，近年间，数字内容孪生中智能增强技术在三维视觉领域取得了快速地发展。具体来说，数字图像是三维世界在摄影设备上的二维投影，传统的数字化记录了拍摄影像的色彩信息，但却无法保留三维世界中的深度、材质和光照等信

息。现有的数字孪生技术，可以利用对同一场景拍摄的多张照片，重构并生成相应的三维内容。最近，谷歌等多家国内外科技公司正探索使用互联网上商家和用户上传的照片，生成并渲染不同餐厅、街道和景点的三维全景。通过数字内容孪生中的智能增强技术，算法可以过滤剔除不同照片中天气、时间、行人等扰动信息，专注于生成并渲染不同场所的全时间段三维全景^[22]。

数字内容孪生中的智能转译技术是建立在对客观世界内容感知的基础上，进一步理解孪生后的数字内容，从而实现多样化的内容呈现的一类技术集合。现阶段比较成熟的智能转译技术包括给定语音信号进行字幕合成，依据文字进行语音生成等。对于智能转译技术，放在第一位的是生成内容的准确性，无论是语音到文本还是文本生成语音，准确地呈现原始信息是该类技术走向实际应用的基础。在准确的基础上，为应对不同的使用场景，相关算法、工程人员还在不断地提高转译算法的实时性和生成语音的真实性。近些年间，智能转译技术已被越来越多地应用于社交、传媒、协同办公、残疾人辅助等实际场景中，为人们的生成生活带来更多的便利。

相比于较为成熟的语音/字幕合成，视觉内容描述^[23]是近年间学术领域的热点研究课题之一。视觉描述技术致力于生成能够准确描述给定视觉内容（例如图像、视频等）的文本和语音。视觉内容描述技术可以被广泛地应用于赛事转播、智慧交通、影视娱乐等各类应用场景中。虽然现阶段的智能转译技术已经可以初步的描述图像(或视频)

中的人物、物体和环境信息，但如何能够准确地生成有关人物行为和主体关系的描述仍是现有技术亟需突破的问题。相比于智能增强技术，智能转译技术更加关注数字世界中不同模态的数字内容间相互理解、融合和转换的能力，从而丰富智能数字内容孪生技术的应用范围和灵活性。

数字内容孪生技术通过对真实世界中内容的智能增强和转译，将现实世界的物理属性（如物体的大小、纹理、颜色等）和社会属性（如主体行为、主体关系等）高效、可感知地进行数字化，实现现实世界到数字世界的映射，构建了在数字世界中重现现实场景的能力。通过数字内容孪生技术，不同行业的从业者可以更好地在数字世界中进行内容的组织和展示。

2. 理解与控制组成内容编辑能力

在数字内容孪生技术的基础上，智能数字内容编辑的相关技术构建了虚拟数字世界与现实物理世界间的交互通道。一方面，对数字内容的编辑和控制，例如数字人技术，可以直接作用于物理世界，实现实时的反馈和互动，起到对现实世界中主体陪伴或服务等功能；另一方面，数字内容编辑技术是实现数字仿真的基础。例如在自动驾驶仿真场景中，通过智能编辑，可以实现对同一道路上不同车况和天气状况的控制。基于数字内容仿真，算法模型可以在数字世界中学习到相

应的知识和技能，这些知识可以被用来反哺解决现实世界中的问题⁶。

从技术角度看，智能数字内容编辑主要通过数字内容的语义理解和属性控制两类技术来实现对内容的修改和控制。首先，理解数字内容是对其进行编辑和修改的必要前提。例如，在处理音频数据进行人声分离时，算法模型需要先理解输入的原始声音信号，才能进一步分离其中的人声信号和背景音，生成两段独立的音频内容。同理，对于计算机视觉中的图片、视频剪辑和自然语言处理中的摘要生成任务，都需要数字内容的语义理解技术进行相关语义的理解和概括，继而修改输入的原始数据以得到最终的生成结果。

值得注意的是，现实世界中的内容大多是由多种不同的语义信息组成的。例如，一张人脸照片实际上是由人物的身份信息、面部动作、拍摄视角、摄影设备和光照条件等许多语义信息一同决定的。早期的语义理解技术更多的是将某个内容当做一个整体进行理解，在学习到的数字表征中不同类别的语义信息往往是纠缠在一起的。虽然可以应用于解决某些数字内容编辑任务，但却难以对不同的语义进行精确的理解和修改。基于生成模型的可解耦语义学习技术是解决语义纠缠问题的可行解决方案之一，并在近些年间取得了快速的发展。通过理解并学习不同语义成分的变化，可解耦语义学习技术对数据内容具有更深刻的理解，并逐渐开始服务于人工智能试妆、试衣、生成同一个人

⁶ <https://www.nvidia.cn/omniverse/media-entertainment/>

不同年龄照片等新兴应用程序。

在充分理解数字内容语义的基础上，属性控制技术构成了数字内容编辑的另一主要分支。在语义理解的基础上，数字内容的智能属性控制技术将直接根据用户指定的属性，对原有的内容进行精确地修改、编辑和二次生成。常用的属性控制技术已经广泛地应用于智能图像编辑、文本情感改写和智能调音等多项应用中，并潜移默化地服务人们的生活，作为辅助功能提升内容创作者的效率。此外，先进的智能内容编辑技术结合了语义理解技术和属性控制技术，在处理三维动画内容时，在学习可解耦的视角、光照和角色等语义特征的基础上，智能属性控制技术以比传统算法更加高效且稳定的方式完成虚拟现实、游戏、电影中的渲染和操控^[24]；在构造数字人时，属性控制能力可以根据实际需要快速地编辑数字人的外貌、音色、感情、表情等属性，以完成数字人技术在不同场合环境中的应用。

数字内容编辑技术在内容孪生技术的基础上，具备了对现实世界内容进行语义理解和属性操控的能力，从而构建了数字世界对现实世界内容的影响和反馈。在数字世界中的操作和尝试将不受限于场地、成本、资源消耗等客观约束，所得到的经验知识也能够更好地反馈给现实世界，提升生产生活的效率。

3. 模仿与概念学习造就内容创作能力

上述的数字内容的孪生和编辑能力主要面向客观世界中的真实内容，通过对现实内容的智能孪生、理解、控制和编辑，AIGC 算法

可以快速准确地将现实世界的内容映射到虚拟世界中，并通过控制仿真等方法，对现实世界产生正向的反馈和帮助。更进一步，数字内容的智能创作旨在让人工智能算法具备类似甚至超越人的创作能力。

1968年，毕加索曾这样评价计算机技术：“它们是没用的，只能简单的给出答案。”但在54年后的今天，百度已经可以通过人工智能模型进行绘画创作，并被西安美院的教授评价为具有“美院毕业生水平”，在短短24小时内就售出了8700多份，销售额超过17万元⁷。无需基于任何现实世界中存在的内容主体，基于人工智能算法的内容创作能力有望生成海量的原创数字内容。

按照技术的发展进程和实际应用的形态，数字内容的创作能力可划分为基于模仿的创作和基于概念的创作两类。基于模仿的创作需要人工智能模型首先观察人类的作品，通过学习某一类作品的分布特性，人工智能生成模型可以进行模仿式的新创作。以前文中提到的佳士得拍卖的肖像画为例，人工智能算法利用大约15000张创作于14世纪到20世纪的肖像画，从中学习作画的笔法、内容、艺术风格等。最终，人工智能内容生成模型所创作的肖像画通过了视觉图灵测试，让绝大部分人类都难以区分这幅画是艺术家创作的，还是人工智能的作品。不仅仅局限于智能作画，基于模仿的人工智能生成模型在旋律创作、文本写作和诗词创作等具体任务中都取得了不错的表现。对于某一类具体的内容，例如人物画像、押韵诗歌或乐曲旋律，现有的人工

⁷ https://www.sohu.com/a/557118794_362042

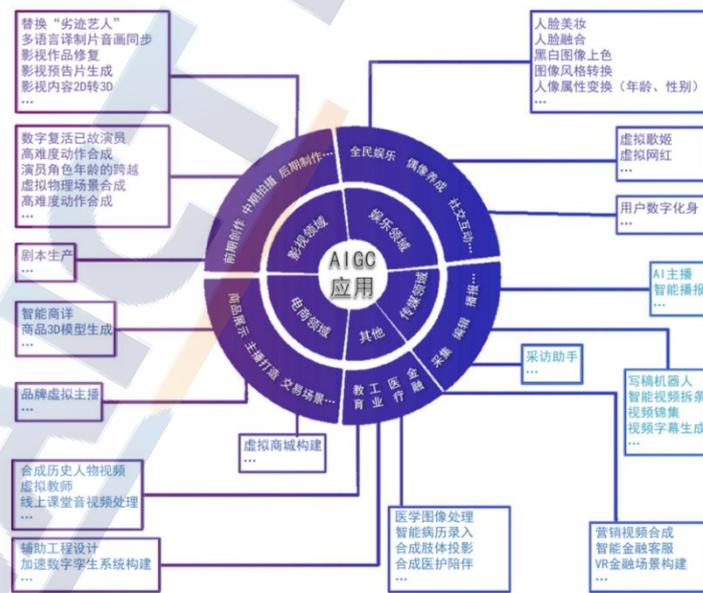
智能技术基本可以创作出让人真假难辨的数字内容。但同时，面对更加复杂的数据内容，例如三维数据、视频数据等，现有的技术所创作的内容相比于真实内容仍有一定差距，需要算法模型的不断完善来缩小这些内容的创作难度。

基于概念的创作不再简单的对固定种类的数据进行观察和模仿，而是致力于在海量的数据中学习抽象的概念，进而通过对不同概念的组合进行全新的创作。以文本到图像的生成为例，给定的文本不仅可以描述生成内容中需要包含的主体内容、数量和关系，还可以指定生成图像的风格、年代等属性。在现实世界中，人们可能只能见到“木头制作的椅子”，“狮子在捕猎獾鼠”等内容，但是通过文本描述，基于概念的创作技术可以创作出“牛油果制作的椅子”，“在猎捕狮子的獾鼠”等视觉内容^[25]。在更进一步理解不同主体间动作、行为、和关系基础上，已经有相关的前沿研究开始尝试通过故事或者剧本描述，创作影视短片。总体来说，基于概念的智能创作与上述智能孪生中的转译技术不同，智能转译更关注对已有内容的精确表达和转换，而基于概念的智能创作是在给定模糊概念的基础上，进行自由生成和创作。数字内容基于概念的创作很大程度上依赖于算法模型对多模态数据的理解、对齐、融合和生成，依赖于人类社会中海量的数据以及相关的描述。基于概念的创作摆脱了对简单学习纹理、形状、颜色的模仿，进一步像人类一样开始学习和总结创作中包含的概念元素，实现更通用、更高效、更智能的 AIGC 应用。

伴随着深度神经网络的快速发展，人工智能模型的规模和能力都在不断被刷新，凭借着数据内容的快速增长，算力的爆发以及算法模型的不断迭代，数字内容创作技术突破到了一个新的高度，规模上不断变大，逐步趋近并开始超过人脑的神经元个数，能力上不断增强，展现出强大的多模态理解和生成能力。

三、人工智能生成内容的应用场景

在全球新冠肺炎疫情延宕反复的背景下，各行业对于数字内容的需求呈现井喷态势，数字世界内容消耗与供给的缺口亟待弥合。AIGC以其真实性、多样性、可控性、组合性的特征，有望帮助企业提高内容生产的效率，以及为其提供更加丰富多元、动态且可交互的内容，或将率先在传媒、电商、影视、娱乐等数字化程度高、内容需求丰富的行业取得重大创新发展。



来源：中国信息通信研究院

图 2 AIGC 应用视图

（一）AIGC+传媒：人机协同生产，推动媒体融合

近年来，随着全球信息化水平的加速提升，人工智能与传媒业的融合发展不断升级。AIGC 作为当前新型的内容生产方式，为媒体的内容生产全面赋能。写稿机器人、采访助手、视频字幕生成、语音播报、视频锦集、人工智能合成主播等相关应用不断涌现，并渗透到采集、编辑、传播等各个环节，深刻地改变了媒体的内容生产模式，成为推动媒体融合发展的重要力量。

在采编环节，一是实现采访录音语音转写，提升传媒工作者的工作体验。借助语音识别技术将录音语音转写成文字，有效压缩稿件生产过程中录音整理方面的重复工作，进一步保障了新闻的时效性。2022 年冬奥会期间，科大讯飞的智能录音笔通过跨语种的语音转写助力记者 2 分钟快速出稿。二是实现智能新闻写作，提升新闻资讯的时效。基于算法自动编写新闻，将部分劳动性的采编工作自动化，帮助媒体更快、更准、更智能化地生产内容。比如 2014 年 3 月，美国洛杉矶时报网站的机器人记者 Quakebot，在洛杉矶地震发生后仅 3 分钟，就写出相关消息并进行发布；美联社使用的智能写稿平台 Wordsmith 可以每秒写 2000 篇报道；中国地震台网的写稿机器人在九寨沟地震发生后 7 秒内就完成了相关消息的编发；第一财经“DT 稿王”一分钟可写出 1680 字^[26]。三是实现智能视频剪辑，提升视频内容的价值。通过使用视频字幕生成、视频锦集、视频拆条、视频超分等视频智能化剪辑工具，高效节省人力时间成本，最大化版权内容

价值。2020年全国两会期间，人民日报社利用“智能云剪辑师”快速生成视频，并能够实现自动匹配字幕、人物实时追踪、画面抖动修复、横屏速转竖屏等技术操作，以适应多平台分发要求^[27]。2022年冬奥会期间，央视视频通过使用AI智能内容生产剪辑系统，高效生产与发布冬奥冰雪项目的视频集锦内容，为深度开发体育媒体版权内容价值，创造了更多的可能性。

在传播环节，AIGC应用主要集中于以AI合成主播为核心的新闻播报等领域。AI合成主播开创了新闻领域实时语音及人物动画合成的先河，只需要输入所需要播发的文本内容，计算机就会生成相应的AI合成主播播报的新闻视频，并确保视频中人物音频和表情、唇动保持自然一致，展现与真人主播无异的信息传达效果。纵观AI合成主播在传媒领域的应用，呈现三方面的特点。一是应用范围不断拓展。目前新华社、中央广播电视总台、人民日报社等国家级媒体及湖南卫视等省市媒体都开始积极布局应用AI合成主播，先后推出“新小微”、“小C”等虚拟新闻主持人，并推动其从新闻播报向晚会主持、记者报道、天气预报等更广泛的场景应用，为全国两会、冬奥会、冬残奥会等重大活动传播深度赋能。二是应用场景不断升级。除了常规的新闻播报，AI合成主播开始陆续支持多语种播报和手语播报。2020年全国两会期间，多语种虚拟主播采用中、韩、日、英等多种语言进行新闻报道，实现了一音多语的播报，将中国新闻传递给世界，顺应了信息化时代信息共享的发展潮流^[28]。2022年冬奥会期间，百

度、腾讯等企业推出手语播报数字人，为千万听障用户提供手语解说，进一步推动观赛的无障碍进程。三是应用形态日趋完善。在形象方面，逐步从 2D 向 3D 拓展；在驱动范围上，开始从口型向面部表情、肢体、手指、背景内容素材延伸；在内容构建上，从支持 SaaS 化平台工具构建向智能化生产探索。例如腾讯 3D 手语数智人“聆语”，实现了唇动、面部表情、肢体动作、手指动作等内容的生成，并配套可视化动作编辑平台，支持对手语动作进行精修。

AIGC 对传媒机构、传媒从业者和传媒受众都产生深刻影响。对传媒机构来说，AIGC 通过参与新闻产品的生产过程，大幅提高生产效率，并带来新的视觉化、互动化体验；丰富了新闻报道的形式，加速了媒体的数字化转型，推动传媒向智媒转变。对传媒从业者来说，AIGC 可助力生产更具人文关怀、社会意义和经济价值的新闻作品；将部分劳动性的采编播工作自动化，让其更加专注于需要深入思考和创造力的工作内容，如新闻特稿、深度报道和专题报道等此类更需发挥人类在精准分析事物、妥善处理情感元素等方面优势的细分领域。对传媒受众来说，AIGC 的应用可使其在更短时间内获得以更丰富多元的形态呈现的新闻内容，提高了其获取新闻信息的及时性和便捷性；降低了传媒行业的技术门槛，促使传媒受众具有更多参与内容生产的机会，极大增强其参与感。

（二）AIGC+电商：推进虚实交融，营造沉浸体验

随着数字技术的发展和应用、消费的升级和加快，购物体验沉浸

化成为电商领域发展的方向。AIGC 正加速商品 3D 模型、虚拟主播乃至虚拟货场的构建，通过和 AR、VR 等新技术的结合，实现视听等多感官交互的沉浸式购物体验。

生成商品 3D 模型用于商品展示和虚拟试用,提升线上购物体验。

基于不同角度的商品图像，借助视觉生成算法自动化生成商品的 3D 几何模型和纹理，辅以线上虚拟“看、试、穿、戴”，提供接近实物的差异化网购体验，助力高效提升用户转化。百度、华为等企业都推出商品自动化 3D 建模服务，支持在分钟级的时间内完成商品的 3D 拍摄和生成，精度可达到毫米级。相较于传统 2D 展示，3D 模型可 720°全方位展示商品主体外观，可大幅度降低用户选品和沟通时间，提升用户体验感，快速促成商品成交。同时生成出的 3D 商品模型还可用于在线试穿，高度还原商品或服务试用的体验感，让消费者有更多机会接触到产品或服务的绝对价值。如阿里于 2021 年 4 月上线 3D 版天猫家装城，通过为商家提供 3D 设计工具及商品 3D 模型 AI 生成服务，帮助商家快速构建 3D 购物空间，支持消费者自己动手做家装搭配，为消费者提供沉浸式的“云逛街”体验。数据显示，3D 购物的转化率平均值为 70%，较行业平均水平提升了 9 倍，同比正常引导成交客单价提升超 200%，同时商品退换货率明显降低。此外，不少品牌企业也开始在虚拟试用方向上开展探索和尝试，如优衣库虚拟试衣、阿迪达斯虚拟试鞋、周大福虚拟试珠宝、Gucci 虚拟试戴手表和眼镜、宜家虚拟家具搭配、保时捷虚拟试驾等^[29]。尽管目前还是采用的传统

手动建模方式，但随着 AIGC 技术的不断进步，未来有望涌现更多消费级工具，从而逐步降低 3D 建模的门槛和成本，助力虚拟试穿应用大规模商用。

打造虚拟主播，赋能直播带货。基于视觉、语音、文本生成技术，打造虚拟主播为观众提供 24 小时不间断的货品推荐介绍以及在线服务能力，为商户直播降低门槛。相比真人直播间带货，虚拟主播具备三大优势：**一是**虚拟主播能够填补真人主播的直播间隙，使直播间能不停轮播，既为用户提供更灵活的观看时间和更方便的购物体验，也为合作商家创造更大的生意增量。如欧莱雅、飞利浦、完美日记等品牌的虚拟主播一般会在凌晨 0 点上线，并进行近 9 个小时的直播，与真人主播形成了 24 小时无缝对接的直播服务。**二是**虚拟化的品牌主播更能加速店铺或品牌年轻化进程，拉近与新消费人群的距离，塑造元宇宙时代的店铺形象，未来可通过延展应用到元宇宙中更多元的虚拟场景，实现多圈层传播。如彩妆品牌“卡姿兰”推出自己的品牌虚拟形象，并将其引入直播间作为其天猫旗舰店日常的虚拟主播导购。同时对于已具备虚拟品牌 IP 形象的传统企业，可直接利用已有形象快速转化形成虚拟品牌主播。如在 2020 年 5 月海尔直播大促活动中，大家所熟知的海尔兄弟虚拟 IP 来到直播间，并同主持人和粉丝一起互动，高达千万播放量。**三是**虚拟主播人设更稳定可控。在头部主播有限并且可能“人设崩塌”的情况下，虚拟主播人设、言行等由品牌方掌握，比真人明星的可控性、安全性更强。品牌不必担心虚拟形象

人设崩塌，为品牌带来负面新闻、差评及资金损失。

赋能线上商城和线下秀场加速演变，为消费者提供全新的购物场景。通过从二维图像中重建场景的三维几何结构，实现虚拟货场快速、低成本、大批量的构建，将有效降低商家搭建 3D 购物空间的门槛及成本，为一些原本高度倚重线下门店的行业打开了线上线下融合的想象空间，同时为消费者提供线上线下融合的新消费体验。目前一些品牌已经开始尝试打造虚拟空间。例如奢侈品商 Gucci 在一百周年品牌庆典时，把线下的 Gucci Garden Archetypes 展览搬到了游戏 Roblox 上，推出了为期两周的虚拟展，5 个主题展厅的内容与现实展览相互对应。2021 年 7 月，阿里巴巴首次展示了其虚拟现实计划“Buy+”，并提供 360° 虚拟的购物现场开放购物体验。2021 年 11 月，Nike 和 Roblox 合作，推出虚拟世界 Nikeland，并向所有 Roblox 用户开放。随着基于图像的 3D 重建技术在谷歌地图沉浸式视图功能中的成功应用，虚拟货场的自动化构建未来将得到更好的应用和发展。

（三）AIGC+影视：拓展创作空间，提升作品质量

随着影视行业的快速发展，从前期创作、中期拍摄到后期制作的过程性问题也随之显露，存在高质量剧本相对缺乏、制作成本高昂以及部分作品质量有待提升等发展痛点，亟待进行结构升级。运用 AIGC 技术能激发影视剧本创作思路，扩展影视角色和场景创作空间，极大地提升影视产品的后期制作质量，帮助实现影视作品的文化价值与经济价值最大化。

AIGC 为剧本创作提供新思路。通过对海量剧本数据进行分析归纳，并按照预设风格快速生产剧本，创作者再进行筛选和二次加工，以此激发创作者的灵感，开阔创作思路，缩短创作周期。国外率先开展相关尝试，早在 2016 年 6 月，纽约大学利用人工智能编写的电影剧本《Sunspring》，经拍摄制作后入围伦敦科幻电影（Sci-Fi London）48 小时挑战前十强^[30]。2020 年，美国查普曼大学的学生利用 OpenAI 的大模型 GPT-3 创作剧本并制作短片《律师》。国内部分垂直领域的科技公司开始提供智能剧本生产相关的服务，如海马轻帆推出的“小说转剧本”智能写作功能，服务了包括《你好，李焕英》《流浪地球》等爆款作品在内的剧集剧本 30000 多集、电影/网络电影剧本 8000 多部、网络小说超过 500 万部。

AIGC 扩展角色和场景创作空间。一是通过人工智能合成人脸、声音等相关内容，实现“数字复活”已故演员、替换“劣迹艺人”、多语言译制片音画同步、演员角色年龄的跨越、高难度动作合成等，减少由于演员自身局限对影视作品的影响。如央视纪录片《创新中国》中，央视和科大讯飞利用人工智能算法学习已故配音员李易过往纪录片的声资料，并根据纪录片的文稿合成配音，配合后期的剪辑优化，最终让李易的声音重现。在 2020 年播出的《了不起的儿科医生》中，主角人物的学历事件影响了影视作品的宣传与发行，该作品便采用了智能影视换脸技术将主角人物进行替换，从而减少影视作品创作过程中的损失。2021 年，英国公司 Flawless 针对多语言译制片中角色唇形

不同步的问题推出了可视化工具 TrueSync，能通过 AI 深度视频合成技术精准调整演员的面部特征，让演员的口型和不同语种的配音或字幕相匹配。二是通过人工智能合成虚拟物理场景，将无法实拍或成本过高的场景生成出来，大大拓宽了影视作品想象力的边界，给观众带来更优质的视觉效果和听觉体验。如 2017 年热播的《热血长安》，剧中的大量场景便是通过人工智能技术虚拟生成。工作人员在前期进行大量的场景资料采集，经由特效人员进行数字建模，制作出仿真的拍摄场景，演员则在绿幕影棚进行表演，结合实时抠像技术，将演员动作与虚拟场景进行融合，最终生成视频^[31]。

AIGC 赋能影视剪辑，升级后期制作。一是实现对影视图像进行修复、还原，提升影像资料的清晰度，保障影视作品的画面质量。例如中影数字制作基地和中国科技大学共同研发的基于 AI 的图像处理系统“中影·神思”，成功修复《厉害了，我的国》《马路天使》等多部影视剧。利用 AI 神思系统，修复一部电影的时间可以缩短四分之三，成本可以减少一半。同时，爱奇艺、优酷、西瓜视频等流媒体平台都开始将 AI 修复经典影视作品作为新的增长领域开拓。二是实现影视预告片生成。IBM 旗下的人工智能系统 Watson 在学习了上百部惊悚预告片的视听手法后，从 90 分钟的《Morgan》影片中挑选出符合惊悚预告片特点的电影镜头，并制作出一段 6 分钟的预告片。尽管这部预告片需要在制作人员的重新修改下才能最终完成，但却将预告片的制作周期从一个月左右缩减到 24 小时。三是实现将影视内容从

2D 向 3D 自动转制。聚力维度推出的人工智能 3D 内容自动制作平台“峥嵘”支持对影视作品进行维度转换，将院线级 3D 转制效率提升 1000 多倍。

（四）AIGC+娱乐：扩展辐射边界，获得发展动能

在数字经济时代，娱乐不仅拉近了产品服务与消费者之间的距离，而且间接满足了现代人对归属感的渴望，重要性与日俱增。借助于 AIGC 技术，通过趣味性图像或音视频生成、打造虚拟偶像、开发 C 端用户数字化身等方式，娱乐行业可以迅速扩展自身的辐射边界，以更加容易被消费者所接纳的方式，获得新的发展动能。

实现趣味性图像或音视频生成，激发用户参与热情。在图像视频生成方面，以 AI 换脸为代表的 AIGC 应用极大满足用户猎奇的需求，成为破圈利器。例如 FaceAPP、ZAO、Avatarify 等图像视频合成应用一经推出，就立刻病毒式在网络上引发热潮，登上 App Store 免费下载榜首位；人民日报新媒体中心在国庆 70 周年推出互动生成 56 个民族照片人像的应用刷屏朋友圈，合成照片总数超 7.38 亿张；2020 年 3 月，腾讯推出化身游戏中的“和平精英”与火箭少女 101 同框合影的活动，这些互动的内容极大地激发出了用户的情感，带来了社交传播的迅速破圈。在语音合成方面，变声增加互动娱乐性。如 QQ 等多款社交软件、和平精英等多款游戏均已集成变声功能，支持用户体验大叔、萝莉等多种不同声线，让沟通成为一种乐此不疲的游戏。

打造虚拟偶像，释放 IP 价值。一是实现与用户共创合成歌曲，

不断加深粉丝黏性。以初音未来和洛天依为代表的“虚拟歌姬”，都是基于 VOCALOID 语音合成引擎软件为基础创造出来的虚拟人物，由真人提供声源，再由软件合成人声，都是能够让粉丝深度参与共创的虚拟歌手。以洛天依为例，任何人通过声库创作词曲，都能达到“洛天依演唱一首歌”的效果。从 2012 年 7 月 12 日洛天依出道至今十年的时间内，音乐人以及粉丝已为洛天依创作了超过一万首作品，通过为用户提供更多想象和创作空间的同时，与粉丝建立了更深刻联系。

二是通过 AI 合成音视频动画，支撑虚拟偶像在更多元的场景进行内容变现。随着音视频合成、全息投影、AR、VR 等技术的成熟，虚拟偶像变现场景逐步多元化，目前可通过演唱会、音乐专辑、广告代言、直播、周边衍生产品等方式进行变现。同时随着虚拟偶像商业价值被不断发掘，品牌方与虚拟 IP 的联动意愿随之提升。如由魔珐科技与次世文化共同打造的网红翎 Ling 于 2020 年 5 月出道至现在已先后与 VOGUE、特斯拉、GUCCI 等品牌展开合作。

开发 C 端用户数字化身，布局消费元宇宙。自 2017 年苹果手机发布 Animoji 以来，“数字化身”技术迭代经历了由单一卡通动物头像，向 AI 自动生成拟真人卡通形象的发展，用户拥有更多创作的自主权和更生动的形象库。各大科技巨头均在积极探索“数字化身”相关应用，加速布局“虚拟数字世界”与现实世界大融合的“未来”。例如百度在 2020 年世界互联网大会上展现了基于 3D 虚拟形象生成和虚拟形象驱动等 AI 技术设计动态虚拟人物的能力。在现场只需拍

摄一张照片，就能在几秒内快速生成一个可以模仿“我”的表情、动作的虚拟形象。在 2021 年的云栖大会开发者展区，阿里云展示了最新技术——卡通智绘项目，吸引了近 2000 名体验者，成为了大会爆款。阿里云卡通智绘采用了隐变量映射的技术方案，对输入人脸图片，发掘其显著特征如眼睛大小、鼻型等，可以自动化生成具有个人特色的虚拟形象，同时还可跟踪用户的面部表情生成实时动画，让普通人也能有机会创造属于自己的卡通形象。在可预见的未来，作为用户在虚拟世界中个人身份和交互载体的“数字化身”，将进一步与人们的生产生活相融合，并将带动虚拟商品经济的发展。

（五）AIGC+其他：推进数实融合，加快产业升级

除以上行业之外，教育、金融、医疗、工业等各行各业的 AIGC 应用也都在快速发展。**教育领域**，AIGC 赋予教育材料新活力。相对于阅读和讲座等传统方式，AIGC 为教育工作者提供了新的工具，使原本抽象、平面的课本具体化、立体化，以更加生动、更加令人信服的方式向学生传递知识。例如制作历史人物直接与学生对话的视频，给一场毫无吸引力的演讲注入新的活力；合成逼真的虚拟教师，让数字教学更具互动性和趣味性等。**金融领域**，AIGC 助力实现降本增效。一方面可通过 AIGC 实现金融资讯、产品介绍视频内容的自动化生产，提升金融机构内容运营的效率；另一方面，可通过 AIGC 塑造视听双通道的虚拟数字人客服，让金融服务更有温度。**医疗领域**，AIGC 赋能诊疗全过程。在辅助诊断方面，AIGC 可用于改善医学图像质量、

录入电子病历等，完成对医生的智力、精力的解放，让医生资源专注到核心业务中，从而实现医生群体业务能力的提升。在康复治疗方面，AIGC 可以为失声者合成语言音频，为残疾者合成肢体投影，为心理疾病患者合成无攻击感的医护陪伴等，通过用人性化的方式来抚慰患者，从而舒缓其情绪，加速其康复。工业领域，AIGC 提升产业效率和价值。一是融入计算机辅助设计 CAD（Computer-aided Design），极大缩短工程设计周期。AIGC 通过将工程设计中重复的、耗时的和低层次的任务自动化，可使原来需要耗费数千小时的工程设计缩短到分钟级。同时支持生成衍生设计，为工程师或设计师提供灵感。此外，还支持在设计中引入变化，实现动态模拟。如宝马公司在其 BMW VISION NEXT 100 概念车中通过 AIGC 辅助设计开发了汽车动态功能性外表皮和内饰。二是加速数字孪生系统的构建。通过将基于物理环境形成的数字几何图形，快速转化为实时参数化的 3D 建模数据，高效创建现实世界中工厂、工业设备和生产线等的数字孪生系统。总体来看，AIGC 正在发展成与其他各类产业深度融合的横向结合体，其相关应用正加速渗透到经济社会的方方面面。

四、人工智能生成内容发展面临的问题

随着人工智能技术发展步入快车道，AIGC 因为其快速的反应能力、生动的知识输出、丰富的应用场景，在社会生产和生活的方方面面发挥着重要的作用。但与此同时，AIGC 的关键技术、企业核心能力和相关法律法规尚未完善，围绕公平、责任、安全的争议日益增多，

引发了一系列亟待解决的问题。

关键技术不够完全成熟，大规模推广落地尚存痛点、难点。目前，AIGC 技术不断升级，进一步释放内容生产力，但其在人工智能关键技术方面尚存在局限，掣肘产业发展进程。**一是人工智能算法存在固有缺陷。**人工智能算法在透明度、鲁棒性、偏见与歧视方面存在尚未克服的技术局限，导致算法应用问题重重。**在透明度方面**，由于算法模型的黑箱运作机制，其运行规律和因果逻辑并不会显而易见的摆在研发者面前。这一特性使人工智能算法的生成机理不易被人类理解和解释，一旦算法出现错误，透明度不足无疑将阻碍外部观察者的纠偏除误。**在鲁棒性方面**，算法运行容易受到数据、模型、训练方法等因素干扰，出现非鲁棒特征。例如，当训练数据量不足的情况下，在特定数据集上测试性能良好的算法很可能被少量随机噪声的轻微扰动影响，从而导致模型给出错误的结论；在算法投入应用之后，随着在线数据内容的更新，算法很可能会产生系统性能上的偏差，进而引发系统的失灵。**在偏见与歧视方面**，算法以数据为原料，如果初始使用的是有偏见的的数据，这些偏见可能会随着时间流逝一直存在，无形中影响着算法运行结果，最终导致 AI 算法生成的内容存在偏见或歧视，引发用户对于算法的公平性争议。**二是 AIGC 内容编辑与创作技术不够完善。**人工智能技术加持的内容编辑与创作技术仍然受短板制约，导致产业发展存在技术门槛。**文本生成方面**，企业在自然语言理解技术上存在瓶颈，往往只简单地套用模板生成机械化的填充，导致文本

结构雷同、千篇一律，而且难以真正产出感性的、拟人的表达，背离用户对于文本合成产品的易读化、优质化期待。**语音合成方面**，语音表达不够流畅、声音机械感较强等问题突出。语音的情感嵌入需要大规模的数据量支持训练，并且对于建模的要求非常高，由此导致使用复杂度提升，也使得相应的成本难以控制，制约企业释放技术价值。**视觉生成方面**，存在智能图像的处理效果不够理想，实时动作捕捉精准度不足等问题。在应用中，由于视觉大模型同时完成多种视觉感知任务的能力不足，机器视觉的精准度、还原度、仿真度不能周全，需要后期人工标注，因而技术门槛高、制作效率低的问题没有得到很好解决。

企业核心能力参差不齐，威胁网络内容生态健康安全发展。随着数字技术的开源开放，AIGC 技术研发门槛、制作成本等不断降低，致使市场上的平台企业泥沙俱下，企业核心能力不足对良好网络生态构建造成严重障碍。**一是内容审核能力有待提升。**近年来，各 AIGC 企业通过建立内容审核机制的方式落实互联网内容治理主体责任，“机审+人审”已成为其基本审核方式。在机审方面，审核准确率受审核类型、内容违规变种繁杂、网络黑灰产对抗手段加剧等影响而导致误报率偏高，需要人工叠加审核。在人审方面，使用人审外包服务已经成为市场主流，但不同的人审团队在人员管理、业务流程管理、审核能力等方面表现各异，行业内也未形成统一的标准。总体而言，缺乏合格的审核人员可能会导致包含虚假、不良信息的违法违规内容

流出，严重影响产业甚至整个网络生态环境。二是**企业技术管理能力建设不足**。由于 AIGC 技术愈发复杂，且在企业中的运用往往具有高动态性等特点，要求企业作为技术设计者和服务提供者具备相应的技术管理能力。然而，企业具有商业属性，这就决定了在资源有限的情况下其往往倾向于首先满足自身利益，而对技术安全和制度保障投入不足。在这方面，各企业的差距十分明显。投资积累“家底”厚、发展时间长的企业，就更有可能技术防护和管理水平较好，反之不然。诸多初入市场的小型企业和技术管理能力不达标的情况将 AIGC 投入应用，为抄袭侵权、内容造假、恶意营销等灰黑产业链提供温床。

三是**企业风险治理能力尚未完善**。《关于加强互联网信息服务算法综合治理的指导意见》明确提出强化企业主体责任。企业应构建完善的人工智能管理能力，切实防范人工智能发展过程中的各项风险。但是，当前 AIGC 技术仍处于发展初期，其风险具有未知性和复杂性等特点，很多企业对于对风险的预测、防范和应急处置能力均尚未完善，风险治理理念也未落实到工程技术实践中。这一问题导致企业很可能错失把风险拦截在萌芽状态的机会，在复杂的网络安全博弈中处于被动，一旦遭受内部威胁或外部攻击，极易引发网络信息内容生态安全风险。

相关规范指引尚需完善，发展与治理之间存在匹配问题。近年来，人工智能产业规范指引不断推出，治理体系初显格局，但随着科技进步加快，制度建设亦步亦趋也未必严丝合缝，这又引发了技术创新发展与政策支持、法律规制的匹配问题。一是**产业发展需落实支持政策**。

2021年3月，我国十四五规划纲要出台，提出“打造数字经济新优势”的建设方针并强调了人工智能等新兴数字产业在提高国家竞争力上的重要价值。在规划纲要的指引下，面对人工智能生成内容关联产业——尤其是数字文化产业的迅速发展，中央政府相继出台了多项政策推动发展数字文化产业新型业态。2022年5月，最新出台的《关于推进实施国家文化数字化战略的意见》，要求研究制定扶持文化数字化建设的产业政策，强调各地要因地制宜制定具体实施方案，相关部门要细化政策措施。未来，各地、各部门政策的支持力度、推进落实和动态调整情况将决定着技术与社会的相互建构程度，将对AIGC技术在社会情境中的发展起到重要作用。二是AIGC可版权性有待厘清。当前，我国《著作权法》中规定，著作权的指向对象为“作品”。仅从法律文本来看，我国现行知识产权法律体系均规定法律主体为享有权利、负有义务和承担责任的人，因此非人生产的智能化内容难以通过“作品—创作—作者”的逻辑获得著作权的保护⁸，这一观点获得了2019年北京互联网法院的判决支持。而在2020年腾讯公司诉网贷之家网站转载机器人自动撰写的文章作品一案中，深圳南山区法院认为在满足独创性要求的情况下，人工智能撰写的文章属于著作权保护的作品。法律概念的模糊引发司法裁判的翻转，导致AIGC作品存在着著作权归属不清的现实困境。这一问题不仅可能导致使用AIGC技术

⁸中华人民共和国国家互联网信息办公室：“人工智能生成内容的著作权”，http://www.cac.gov.cn/2019-08/21/c_1124902661.htm.

创作的作品无法获得著作权保护，阻碍人工智能技术发挥其创作价值，还有可能因人工智能的海量摹写行为稀释既有作品权利人的独创性，威胁他人的合法权益。三是新技术增加监管难度。近年来，随着人工智能技术不断成熟，机器深度学习后生成的内容愈发逼真，能够达到“以假乱真”的效果。相应地，应用门槛也在不断降低，人人都能轻松实现“换脸”、“变声”，甚至成为“网络水军”中的一员。由于契合民众“眼见为实”的认知共性，技术滥用后很可能使造假内容以高度可信的方式通过互联网即时触达用户，导致公众在观念博弈中判断失灵，难以甄别水军和虚假信息。而这又牵涉到一个现实的难题，那就是由于互联网提供的虚拟身份外衣和相关技术的发展，造假内容生产者具有分散性、流动性、大规模性和隐蔽性的特点，导致追踪难度和复杂性与日俱增，再加上规范指引的模糊和滞后，对于那些擦边球性质的造假行为存在难以界定的现实困境，这无疑对内容监管行动造成了严重阻碍。

五、发展建议与展望

（一）发展建议

1. 深化正向引导、稳步推进监管治理

发展促进角度，建议相关行业主管部门可以通过政策引导充分释放 AIGC 的正面应用价值，结合自身资源禀赋和发展需求，藉由规划指引、财政补贴、试点示范、揭榜挂帅等方式，加强数据归集、算力

统筹、算法开源等平台 and 基础能力建设，支持 AIGC 赋能数字经济。一方面，引导数字内容产业发展，支持 AIGC 技术在影视传媒、社交娱乐、电商零售场景构建等领域创造价值增量，推动数字内容产业跨界衍生，刺激信息消费；培育与 AIGC 产业发展相匹配的内容审核技术产品和服务体系。另一方面，坚持“以虚促实、以虚强实”基本导向，结合产业数字化转型工作，推动 AIGC 技术参与生产和生活方式重构。例如，以 AIGC 打造逼真在线服务体验，在社会服务、商贸服务、医疗教育服务等领域催生新服务模式、新商业模式，鼓励弱势或特殊群体援助、教育资源公平等社会民生领域的应用场景构建；探索以 AIGC 技术为基础，对现实世界进行人机交互更为自然、环境细节更为逼真的虚拟空间模拟仿真，推动 AIGC 赋能工业制造、城市规划、科学研究、生产办公等领域的效力提升和价值再造，从虚拟维度优化现实世界运转。

风险治理角度，建议统筹推进“技术”监管与“内容”治理，逐步完善保障 AIGC 良性发展的法律法规体系。近年来，我国相继藉由《网络信息内容生态治理规定》《民法典人格权编》《网络数据安全管理条例（征求意见稿）》《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定（征求意见稿）》等政策法规，对 AIGC 相关技术应用划定了法律底线，但责任认定与归结、知识产权保护等问题总体处于探索阶段。建议政府部门和研究机构前瞻布局，一是探索明确 AIGC 技术服务、内容传播与技术应用各相关方面的法律和社

会责任。探索结合使用意图和使用效果，从法律法规层面进一步明确违法有害的 AIGC 应用范围；针对基于深度伪造技术的虚假内容制作和散布等问题，探索对制作者披露和标注生成内容、对传播媒介提高鉴别能力提出明确要求、细化主体责任；针对内容爆发、信息过载带来的沉溺迷失等问题，探索出台 AIGC 分级分类指南。二是开展针对 AIGC 作品权属认定等方面的法律法规研究，形成模仿、抄袭或合理使用的认定标准，探索侵权主体及侵权责任的认定路径，对 AIGC 作品及其智力与经济投入、技术工具及作品所依赖的数据基础等提供平衡的保护，确保相关数字内容的合规有序流通。同时，谨遵包容审慎理念，鼓励立法研究的多方参与、监管手段的分级分类、行业治理的公私合作，针对 AIGC 技术及应用快速迭代的特点，探索形成法律法规动态评估、修改和废止机制，确保法律法规和监管治理的科学化、精细化、敏捷化。

2. 呼吁自律自治、共建产业发展生态

发展促进角度，建议政产学研用各主体基于开源共享平台促成协同合作、加速应用创新。围绕 AIGC 产业发展与治理需求，推动行业层面在算力能力、算法技术、AI 工程化等方面的联合攻关，特别是聚力突破算法透明度、鲁棒性、偏见与歧视等技术局限，打破行业发展瓶颈；鼓励超大神经网络模型的联合研发与开源，支持产学研各界通过合作构建训练与标准测试数据集、搭建面向重点行业应用的开发者协同平台等方式，加速共性技术和基础产品研发、降低 AIGC 应用门

槛，促进产业繁荣发展；探索构建行业级 AIGC 参考实施框架，以生态协议、行业标准的互联互通支撑内容生态的共建共享；探索通过推广制度模板、创新多边议程、利用市场工具等手段，积极参与 AIGC 应用与治理等领域的国际规则制定和全球发展合作，藉由规则输出和辐射影响力的扩大，支持相关产业和重点企业增强海外竞争力、争取更大国际市场。

风险治理角度，建议行业层面强化“伦理先行”意识、加强自律自治，共同打造 AIGC 良性发展生态。鼓励相关龙头企业、行业联盟、研究机构和标准组织等发挥资源统筹优势，加强内容识别、事实核查、问题感知、违法打击、舆情治理、版权保护等治理技术能力研究，围绕 AIGC 的合法合规应用和健康有序发展联合制定公约、标准、指南、准则等行业制度规范，建立争议解决、行业黑名单、危机应对联动等行业自治机制，并积极开展制度宣传、标准推广、测试评估、标杆塑造等活动；特别是针对“算法偏见”、“算法黑箱”带来的数字内容不公正、责任主体难界定等问题，牵头支持可解释、可信赖的 AIGC 技术框架、标准体系、评测机制等的研究和推广。

3. 加速工程实践、积极履行社会责任

发展促进角度，建议人工智能企业把握数字内容工业化趋势，加速 AIGC 技术工程化。人工智能企业可以在提高算法、算力等基础能力的基础上，一方面，通过完善数字内容模板库、知识图谱储备，加强渲染引擎、仿真引擎等标准化创作工具研发，完善 AIGC 工具体系，

提高孪生、编辑、创作关键技术工程化水平，加速内容生产、打造海量应用场景；另一方面，以 AIGC 工具体系为基础，通过优化模型开发和技术管理流程等方式提高定制化解决方案交付效率，从而提升产品和服务质量、快速响应多元化内容生成需求。同时，可以积极向内容创作与运营、内容传播、内容分发、内容平台、内容审核等数字内容产业上下游环节延伸，加快构建产业发展先发优势；围绕服务国家战略大局，不断迭代技术和产品，主动赋能实体经济。

风险治理角度，建议相关企业践行科技向善理念、积极履行社会责任，把伦理、安全等底线问题和企业发展摆在同等重要的地位。建议相关企业可以针对生成内容主动进行合理标识，特别是对产业链和用户等多方主体起到纽带作用的大型互联网企业和平台型企业，应加强与政府部门和行业组织的治理合作，通过加强鉴别和审核技术能力部署、支持平台用户自主标记不良内容、建立应急辟谣机制、发布平台审核透明度报告等方式，主动承担风险防范、内容治理责任。

4. 构建全民共享和多元共治发展格局

发展促进角度，建议行业组织、媒体机构、职业群体等各类社会力量及时做好正面宣传，可以通过 AIGC 技术在大型公共活动中的应用等方式，帮助广大群众及时了解 AIGC 应用水平、行业动态及发展趋势，推动 AIGC 技术较快较好融入各类场景。如今年的五四晚会，央视携手腾讯音乐虚拟世界 TMELAND，将超百万青年人带入数实融

合的虚拟音乐世界，通过打造 AIGC 融合应用范例向公众呈现了前沿技术成果和场景探索可能性。

风险治理角度，建议通过制度安排、服务规范、媒体引导和产品设计等方式，全社会共同参与、全民提高“数字素养”，一方面，深化公众对于 AIGC 技术滥用带来的风险和危害的认知，逐步提高个人信息保护和对不良 AIGC 的防范意识，并探索通过开放免费技术工具、设立咨询专线等方式鼓励公众对于互联网传播信息的多方验证，提高广大群众对于虚假信息的鉴别能力、减少潜在损失；另一方面，培养公众作为 AIGC 技术使用者和内容受众的社会责任意识，鼓励广大群众主动参与内容制作与传播等环节的舆论监督，共同营造良性发展氛围。

（二）未来展望

1. 核心技术持续演进

（1）从真实可控向多样组合发展

从技术层面上看，目前 AIGC 的相关算法已经具备了真实复刻和创造某类既定内容的能力，同时相关模型对简单场景的内容生成也取得了较好成果，但面对多样性变化和复杂场景内容生成的挑战，现有 AIGC 的算法能力仍需进一步提升。举例来说，目前 AIGC 对于图像生成和编辑已经取得了令人惊叹的成果，例如对高清人脸图像或数字人头像的生成，相关算法已经能做到以假乱真。相较而言，动画视频的动态复杂性和可能性的复杂程度以几何倍数增长，高质量的视频创

作仍有巨大提升空间。同时，仅仅依靠单一生成个体的内容生成对构建理想的数字世界甚至元宇宙是远远不够的。AIGC 技术的下一个发展方向将是**通过不同生成个体间的交互进行内容创作**。通过对整体的、多模态的复杂场景创作，AIGC 将有望实现更多数智内容，进而反哺核心及关联领域，促进共同发展。

（2）从本地化集中式向大规模分布式发展

作为数字经济和实体经济深度融合的新模式，AIGC 通过应用大量新型人工智能技术来创造和制作丰富的创新型、高质量、可交互的数字内容，这给当前人工智能技术的研究带来了全新的挑战，**针对大规模分布式 AIGC 交互算法的研究是时下紧迫且热门的主题，也是 AIGC 技术未来的发展方向之一**。究其原因，一是 AIGC 离不开大规模分布式深度学习技术和多智能体强化学习技术。大规模分布式 AIGC 的发展，将有助于高效利用 GPU 算力，将计算流程拆解到一系列算力平台和边缘计算设备上，通过多设备分布式计算的方式，加速内容制作流程，提高生成效率和质量。目前以谷歌和微软为代表的人工智能公司，已经开始布局下一代超大规模人工智能模型的分布式计算平台，如 Pathways，DeepSpeed 等，以此来针对性解决大规模模型训练面临的算力紧缺、资源利用率低、无法高效进行模型制作等难题。二是大规模多智能体网络在分布式计算的框架下通过个体间的合作和竞争可以完成单体无法完成的任务。AIGC 作为构建数字世界乃至元宇宙的重要生产工具，需要模拟现实世界复杂的大规模多智能体网

络系统，如动物群体、社交网络、城市复杂体等。通过对大规模分布式多智能体算法的研究，探索多智能体的拓展性、安全性、稳定性、迁移性等，将是未来重点探索的方向之一。

2. 关键能力显著增强

随着 AIGC 核心技术的不断发展，其内容孪生、内容编辑、内容创作三大基础能力将显著增强。一是随着渲染技术、仿真加速、XR 套件及开发组件等技术的提升，基于内容孪生的数字孪生能力可以将现实世界更逼真地复制到虚拟世界，复刻出人物更丰满、物体更逼真、细节更丰富的虚拟镜像，并依托新一代传感器以及云边端协同作业，进行实时动态更新。二是依托内容编辑的数字伴生能力将进一步打通现实世界与虚拟世界的双向通道。通过现实世界与虚拟世界的双向联通，将现实世界的物理问题抽象化与数字化，进而转化为虚拟世界的计算问题，并将计算的最优解以物理形式输出至现实世界。未来，依托虚拟优化、智能控制、可信认证等关键技术提升，数字伴生将进一步扩大现实世界发现问题、解决问题的能力、同时降低成本，提高产出。三是基于内容创作的数字原生能力的潜能将得到真正释放。随着未来 AIGC 数字原生能力的大幅提升，基于更先进算法的人工智能技术将使 AIGC 摆脱对专业生成内容（PGC）和用户生成内容（UGC）的依赖，完全独立自主进行内容创作，充分释放其创作潜力。内容将以高质量、多样化、高自由度持续输出，填补目前专业生成内容（PGC）和用户生成内容（UGC）的产能及监管缺口，特别是随着人工智能技

术的持续发展和不断迭代，AIGC 将从目前以辅助内容生成为主向以自主内容生成为主跃进，极大满足未来消费者对内容数量及质量的双重刚性需求。

3. 产品类型逐渐丰富

近年来，随着元宇宙概念的兴起以及科技进步而快速发展的数字人是未来 AIGC 应用的一个重要细分赛道。数字人作为现实与虚拟世界的交互媒介，通过其特有的人设、性格、思维、职业等并辅以专属的创作内容，可以打破传统的物理及时空边界，通过 VR、AR 等技术和新一代智能终端设备，带给用户丰富的沉浸式体验。而可以自主生成内容的数字人将是构建人机交融、数实融合的未来世界的最佳载体，同时也是未来人类构建“自生成、自更新”的元宇宙新世界的必经之路。随着人工智能相关技术的持续发展，数字人开发的自由度将大幅提高，不同个人、不同企业的数字人将更具辨识度和独立特征，开发成本将大幅降低以促进数字人的普世化发展。同时，具有自主内容生成能力的“智能化”乃至“智慧化”数字人意味着可以进行无限的内容创作，元宇宙将会提供人类可以自由探索的广袤空间，人类将不再受真实时空的束缚，基于 AIGC 的元宇宙世界不再依赖于现实世界的投射与编辑，而将脱离于现实世界自我生成、自我发展、自我更新。人工智能的不断发展及推进，也将使 AIGC 模态不再仅仅局限于文本、音频、视觉三种基本形态。嗅觉、触觉、味觉、情感等多重信息感知和认知能力将以数字化的形式传输并指导人工智能进行内容创作，未

来的某一天，人工智能能创造出酸甜苦辣外的另一种味道也未可知。

4. 场景应用趋于多元

目前，AIGC 已经在多个领域得到广泛应用，如金融、传媒、文娱、电商等，未来应用场景会进一步多元化。例如在“AIGC+数据科学”领域，可以自动生成具有安全性、标签化、经过预处理的标准数据以满足越来越“饥饿”的人工智能模型。目前由人工智能生成的数据占有所有数据的 1% 不到，根据 Gartner 预测，到 2025 年，人工智能生成数据占比将达到 10%^[32]；在“AIGC+游戏”领域，通过训练，人工智能可以生成针对不同玩家阶层的游戏攻略和教学手册，并且无重复自动创建不同难度、高交互性、高可玩性的剧情和关卡；在“AIGC+医药”领域，相关模型可以克服医疗数据的稀缺性，自动寻找带有特定性质的分子结构，以此大幅缩短新药的研发和临床试验的科研成本，据 Gartner 预测，到 2025 年，超过 30% 的药物和材料将通过生成式人工智能（AIGC 的工具之一）被发现^[33]；在“AIGC+安全”领域，在公众场合或者活动中，人工智能将会自动生成用户虚拟形象以保护用户的数据安全和个人隐私；在“AIGC+艺术”领域，除目前火爆的 NFT（Non-Fungible Token，非同质化代币）以外，AIGC 还可涉及绘画、作曲、演唱、编剧、设计等，不同于目前在这些细分领域以辅助内容生成为主的现状，未来这些领域的自主生成内容将会达到人类水平，经人工智能创作后无需人为优化即可投放市场。除应用场景的横向扩展外，场景与场景的融合交互也是未来的发展趋势之一。例如通过“文

旅+游戏”，以高度沉浸式体验深度挖掘传统文旅产业新特点，以新颖的、无重复的、高吸引力互动游戏形式吸引年轻消费群体深度探索传统文旅的深刻内涵，激发传统文旅产业新活力；通过“教育+政务”，AIGC 可以针对不同年龄、不同学历、不同职业、不同地区的人群依照政策导向生成类型各异的教育及科普内容，极大平衡教育资源、更好普及全民教育、创造全民科学氛围、提升全民科学素养；通过“商业+艺术”，AIGC 可以打造更具人文关怀和时代意义的数字藏品，数字藏品将从目前缺乏实质内涵的虚拟货品蜕变为拥有特定纪念意义的虚拟化身，以此深度挖掘艺术的商业潜力，从而反哺并推进未来艺术高质量发展。

5. 生态建设日益完善

随着 AIGC 的不断成熟，以标准规范、技术研发、内容创作、行业应用、资产服务为核心的生态体系架构将日趋完善，无论是以 AIGC 赋能产业升级还是以 AIGC 自主释放价值都将在此框架下健康有序发展。标准规范为 AIGC 生态构建了从技术、内容、应用、服务、监管的全过程一体化标准体系，促进 AIGC 在合理合规合法的框架下良性发展。同时，在核心技术持续演进和关键能力显著增强的背景下，性能更强大、逻辑更智能的人工智能算法将被应用于 AIGC，技术研发的不断创新将强有力地推动内容创作，提高生成内容质量，使内容更接近人类智力水平和审美标准，同时应用于各类行业各种场景。AIGC 的繁荣发展将促进资产服务快速跟进，通过对生成内容的合规

评估、资产管理、产权保护、交易服务等构成 AIGC 的完整生态链，并进行价值重塑，充分释放其商业潜力。随着 5G、云计算、VR、AR 等前沿技术的快速发展和新一代智能终端设备的研发创新，完整的 AIGC 生态链是未来释放数据要素红利、推动传统产业升级、促进数字经济发展、构建数实融合一体、创造元宇宙世界最重要的推动力之一。

参考文献

- [1] Krizhevsky A., et al. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in Neural Information Processing Systems, 2012.
- [2] Kingma D., et al. Auto-encoding variational bayes [C]. In International Conference on Learning Representations, 2014.
- [3] Goodfellow I., et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014.
- [4] Ganin, et al. Synthesizing programs for images using reinforced adversarial learning [C]. In International Conference on Machine Learning, pages 1666-1675, 2018.
- [5] Rezende D., et al. Variational Inference with Normalizing Flows [C]. In International Conference on Machine Learning, pages 1530-1538, 2015.
- [6] Ho, J., et al. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, pages 6840-6851, 2020.
- [7] Karras, T., et al. A style-based generator architecture for generative adversarial networks[C]. In Computer Vision and Pattern Recognition, pages 4401-4410, 2019.
- [8] Mildenhall, B., et al. Nerf: Representing scenes as neural radiance fields for view synthesis[C]. In European conference on computer vision,

- pages 405-421, 2020.
- [9] Repecka, D., et al. Expanding functional protein sequence spaces using generative adversarial networks[J]. *Nature Machine Intelligence*, pages 324-333, 2021.
- [10] Lu H., et al. Machine learning-aided engineering of hydrolases for PET depolymerization[J]. *Nature*, pages 662-667, 2022.
- [11] Zhang J., et al. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things[J]. *IEEE Internet of Things Journal*, 2020, 8(10): 7789-7817.
- [12] Dosovitskiy A., et al. An image is worth 16x16 words: transformers for image recognition at scale[C]. In *International Conference on Learning Representations*. 2021.
- [13] Xu Y., et al. Vitae: vision transformer advanced by exploring intrinsic inductive bias[J]. *Advances in Neural Information Processing Systems*, 2021.
- [14] Zhang Q., et al. Vitaev2: vision transformer advanced by exploring inductive bias for image recognition and beyond[J]. *arXiv preprint arXiv:2202.10108*, 2022.
- [15] Xu Y., et al. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation[J]. *arXiv preprint arXiv:2204.12484*, 2022.
- [16] Liu Z., et al. Swin transformer v2: scaling up capacity and

- resolution[J]. arXiv preprint arXiv:2111.09883, 2021.
- [17] Wang D., et al. An Empirical Study of Remote Sensing Pretraining[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022.
- [18] Devlin, J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics. 2019.
- [19] Radford, A., et al. Improving language understanding by generative pre-training [J]. OpenAI Tech Report. 2018.
- [20] Brown, T., et al. Language models are few-shot learners. In Annual Conference on Neural Information Processing Systems [J]. 2020.
- [21] Jewitt C.. K. O'Halloran: multimodal discourse analysis: systemic functional perspectives. continuum, 2004[J]. Applied Linguistics, Pages 335–337, 2006.
- [22] Martin-Brualla R., et al. Nerf in the wild: neural radiance fields for unconstrained photo collections[C]. In Computer Vision and Pattern Recognition, pages 7210-7219, 2021.
- [23] Vinyals, Oriol, et al. Show and tell: A neural image caption generator[C]. In Computer Vision and Pattern Recognition, pages 3156-3164, 2015.
- [24] Wang C., et al. Tag disentangled generative adversarial networks for

- object image re-rendering[C]. In International Joint Conference on Artificial Intelligence, pages 2901-2907, 2017.
- [25] Gafni O., et al. Make-a-scene: Scene-based text-to-image generation with human priors[J]. arXiv preprint arXiv:2203.13131, 2022
- [26] 丁琼.人工智能背景下机器人新闻写作的应用与反思[J].声屏世界,2021(02):30-31.
- [27] 柴艳霞.人工智能背景下影视传媒教育的转型与发展[J].艺术教育,2021(10):118-121.
- [28] 余尚娇.浅析 AI 合成主播在电视媒介中的应用与发展[J].采写编,2021(11):15-16.
- [29] 龚思颖.论沉浸式媒介在数字化零售空间中的应用[J].商业经济研究,2020(22):5-9.
- [30] 巴胜超,姜佳娟.人工智能技术与影视创作融合发展的的问题与对策[J].昆明理工大学学报(社会科学版), 2021,21(02):99-105.DOI:10.16112/j.cnki.53-1160/c.2021.02.014.
- [31] 柴艳霞.人工智能为影视行业带来的机遇与挑战[J].现代电影技术,2020(10):51-55.
- [32] STAMFORD, Conn. “Gartner Identifies the Top Strategic Technology Trends for 2022”, Gartner, October 18, 2021.
- [33] Laurence Goasduff. “The 4 Trends That Prevail on the Gartner Hype Cycle for AI, 2021”, Gartner, September 22, 2021.

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62302914，021-64171028

传真：010-62304980

网址：www.caict.ac.cn

