

2023 年有望成为 AIGC 的拐点

证券研究报告

2023 年 01 月 30 日

AIGC 行业专题报告

核心结论

GAN/CLIP/Diffusion 是 AIGC (AI-Generated Content) 发展的三大模型。目前 AIGC 生成正在完成从简单的降本增效（以生成金融/体育新闻为代表）向创造额外价值（以提供绘画创作素材为代表）转移，跨模态/多模态内容成为关键的发展节点。AIGC 拥有三大模型：GAN、CLIP 和 Diffusion。

AIGC 应用场景广泛。1) 文本生成：文本交互成为未来发展方向；2) 音频生成：AI 编曲将成为 AI 音频生成中的快速成长赛道；3) 图像生成：目前仍然难以达到稳定可靠的生成高质量图像；4) 视频生成：跨模态生成领域的中高潜力场景；5) 图像、视频、文本间的跨模态生成：文本到视频转换目前还处于技术尝试阶段；6) 策略生成：国内相关游戏厂商普遍存在内部开发或外部测试需求；7) 虚拟人生成：视频是短期的发展重点。

中国 AIGC 产业仍处于发展初期，底层技术相较国外仍有较大差距。国内最先兴起的是 AI 写作和语音合成模块，虚拟人概念刚兴起，但技术层面还存在差距，众多国内厂商从内容布局入手，同时众多厂商目前商业模式不成熟，还停留在免费引流阶段。此外，AIGC 目前主要客户集中在 B 端，厂商话语权较弱，未来会向 2B+2C 方向发展，B 端降本增效是根本诉求，C 端的付费逻辑在于高效获取信息，替代表达以及扩大用户创造力。

2023 年有望成为 AIGC 发展大年。近期微软、谷歌等全球计软件互联网大厂大力投入 AIGC，并开始探索应用层面的突破，AIGC 已经来到 2 年前自动驾驶的应用突破拐点，2023 年有望成为 AIGC 大年。类似 ChatGPT 的“大模型”作为 AIGC 的底层平台，将成为 AIGC 的核心资源。

建议关注：1、自研 AIGC 大模型或者有相关 AI 技术布局的上市公司：云从科技、拓尔思、金山办公、科大讯飞、商汤科技等；2、掌握独特场景，拥有 AIGC 应用场景的上市公司：汉仪科技、万兴科技、新国都、海天瑞声。

风险提示：政策监管出现变化；技术更新不及预期。

行业评级

超配

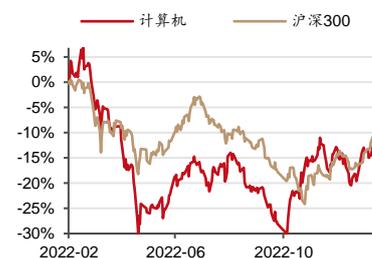
前次评级

超配

评级变动

维持

近一年行业走势



相对表现

1 个月

3 个月

12 个月

计算机

12.46

15.24

-8.40

沪深 300

8.00

18.08

-8.38

分析师



邢开允 S0800519070001



13072123839



xingkaiyun@research.xbmail.com.cn



赵宇阳 S0800522090002



19534078860



zhaoyuyang@research.xbmail.com.cn

相关研究

计算机：安全与发展并重—2023 年计算机行业年度策略 2022-11-30

计算机：职教信息化：用智慧锻造制造强国—职教智慧教育专题 2022-10-22

计算机：高校信息化：疫情影响有限，强国必经之途—高校智慧教育专题 2022-10-17

索引

内容目录

一、GAN/CLIP/Diffusion 是 AIGC 发展的三大模型	4
二、AIGC 应用场景广泛	7
2.1 文本生成：文本交互成为未来发展方向	7
2.2 音频生成：AI 编曲将成为 AI 音频生成中的快速成长赛道	9
2.3 图像生成：目前仍然难以达到稳定可靠的生成高质量图像	11
2.4 视频生成：跨模态生成领域的中高潜力场景	12
2.5 图像、视频、文本间的跨模态生成：文本到视频转换目前还处于技术尝试阶段	13
2.6 策略生成：国内相关游戏厂商普遍存在内部开发或外部测试需求	14
2.7 虚拟人生成：视频是短期的发展重点	15
三、中国 AIGC 产业仍处于发展初期，2023 年有望成为 AIGC 发展大年	16
四、建议关注	17
五、风险提示	18

图表目录

图 1: GAN 网络整体示意	5
图 2: CLIP 模型原理	6
图 3: Imagen 生成结果	7
图 4: 各类生成模型对比图	7
图 5: 澜舟科技文本续写功能	8
图 6: 彩云小梦剧情续写功能	8
图 7: 写作猫文本辅助功能	9
图 8: 小冰岛 AI 聊天界面	9
图 9: 文字转音频功能展示	10
图 10: AI 智能配音功能	10
图 11: 网易-有灵智能创作平台	11
图 12: EditGAN 支持图像进行细节修改	11
图 13: DeepFaceDrawing 草图变完整图像效果	12
图 14: 剪映视频氛围渲染	12
图 15: 视频换脸原理	13
图 16: DALLE2 生成结果	13
图 17: 百度智能视频合成平台功能展示	14
图 18: 腾讯 AI Lab 在游戏领域的全栈布局	14
图 19: 数字分身视频制作界面	15

图 20: 小冰虚拟人产品	16
图 21: AIGC 产业链	16
图 22: ChatGPT 对话展示	17
表 1: AIGC 发展历程	4
表 2: AIGC 技术应用场景	4
表 3: AIGC 的三大模型	5

一、GAN/CLIP/Diffusion是AIGC发展的三大模型

AIGC (AI-Generated Content) 是利用人工智能技术来生成内容。2021年之前，AIGC生成的主要还是文字，而新一代模型可以处理的格式内容包括：文字、语音、代码、图像、视频、机器人动作等等。AIGC被认为是继专业生产内容 (PGC, professional-generated content)、用户生产内容 (UGC, User-generated content) 之后的新型内容创作方式，可以在创意、表现力、迭代、传播、个性化等方面，充分发挥技术优势。

表 1: AIGC 发展历程

阶段	特点	典型事件
早期萌芽阶段 (20世纪50年代至90年代中期)	受限于科技水平，AIGC 仅限于小范围实验	1950年，艾伦·图灵提出著名的“图灵测试”，给出判定机器是否具有“智能”的试验方法； 1957年，第一支由计算机创作的弦乐四重奏《依利亚克组曲 (Illiac Suite)》完成； 1966年，世界第一款可人机对话的机器人“Eliza”问世 80年代中期，IBM 创造语音控制打字机 Tangora。 2007年，世界第一部完全由人工智能创作的小说《1 The Road》问世；
沉淀积累阶段 (20世纪90年代中期至21世纪10年代中期)	AIGC 从实验性向实用性转变，受限于算法瓶颈，无法直接进行内容生产	2012年，微软展示全自动同声传译系统，可将英文演讲者的内容自动翻译成中文语音。 2014年，Ian J. Goodfellow 提出生成式对抗网络 GAN； 2017年，微软“小冰”推出世界首部 100% 由人工智能创作的诗集《阳光失了玻璃窗》；
快速发展阶段 (21世纪10年代中期至今)	深度学习算法不断迭代，人工智能生成内容百花齐放，效果逐渐逼真直至人类难以分辨	2018年英伟达发布 StyleGAN 模型可以自动生成高质量图片 2018年，人工智能生成的图在佳士得拍卖行以 43.25 万美元成交，成为首个出售的人工智能艺术品； 2019年，DepMind 发布 DVD-GAN 模型用以生成连续视频 2021年，OpenAI 推出 DALL-E，主要应用于文本与图像交互生成内容。

资料来源：中国信通院、西部证券研发中心

目前 AIGC 生成正在完成从简单的降本增效 (以生成金融/体育新闻为代表) 向创造额外价值 (以提供绘画创作素材为代表) 转移，跨模态/多模态内容成为关键的发展节点。

表 2: AIGC 技术应用场景

场景	内容
文本生成	结构化写作 (新闻播报等，又比较强的规律)
	非交互式文本 非结构化写作 (剧情续写、营销文本等，需要一定创意和个性化) 辅助性写作 (推荐相关内容、帮助润色，部署于严格 AIGC)
	交互性文本 闲聊机器人 (虚拟男/女友、心理咨询等) 文本交互游戏等 (AI dungeon 等)
音频生成	语音克隆 文本生成特定语言 (生成虚拟人歌声/播报等) 乐曲/歌曲生成 (包含作曲及编曲，在实际应用中常包含自动作词)
	图像生成 图像编辑工具 (去除水印、提高分辨率、特定滤镜等) 图像自动生成 创意图像生成 (随机或按照特定属性生成画作等)

功能性图像生成（根据指定要求生成营销类海报、模特图、logo等）

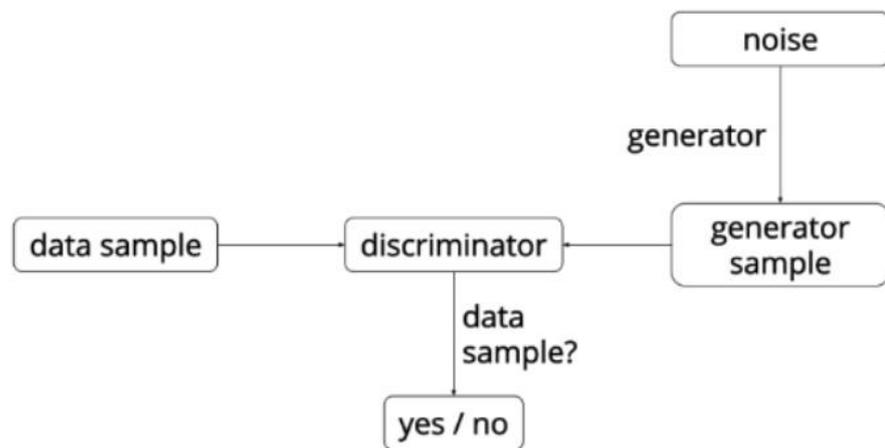
视频生成	视频属性编辑（删除特定主体、生成特效、跟踪剪辑等）
	视频自动剪辑（对特定片段进行检测及合成）
	视频部分编辑（视频换脸等）
图像、视频、文本间跨模态生成	文字生成图像（根据文字 prompt 生成创意图像）
	文字生成演示视频（拼接图片素材生成视频）
	文字生成创意视频（完全从头生成特定主题视频）
	图像/视频到文本（视觉问答系统、自动配字幕/标题等）
策略生成（以 Game AI 中 AI bot 为代表）	
Game AI	AI Bot
	NPC 逻辑剧情生成
	数字资产生成
虚拟人生成	虚拟人视频生成
	虚拟人实时交互

资料来源：量子位智库、西部证券研发中心

AIGC 的三大模型：

“对抗生成网络”GAN (Generative Adversarial Network, 以下简称“GAN”)是很多 AIGC 的基础框架，GAN 包含有两个模型，一个是生成模型 (generative model)，一个是判别模型 (discriminative model)。生成模型的任务是生成看起来自然真实的、和原始数据相似的实例。判别模型的任务是判断给定的实例看起来是自然真实的还是人为伪造的（真实实例来源于数据集，伪造实例来源于生成模型）。

图 1：GAN 网络整体示意



资料来源：AI 遇见机器学习、西部证券研发中心

但是 GAN 有三个不足：1) 对输出结果的控制力较弱，容易产生随机图像；2) 生成的图像分辨率较低；3) 由于 GAN 需要用判别器来判断生产的图像是否与其他图像属于同一类别，这就导致生成的图像是对现有作品的模仿，而非创新。因此依托 GAN 模型难以创作出新图像，也不能通过文字提示生成新图像。

表 3：AIGC 的三大模型

深度学习模型 出现时间	特点
-------------	----

深度学习模型 出现时间		特点
1	GAN	2014 1、生成器用来生成图片，判别器用来判断图片质量，两者互相平衡之后得到结果。 2、对输出结果的控制力较弱，容易产生随机图像、分辨率比较低。
2	CLIP	2021 1、进行自然语言理解和计算机视觉分析。 2、使用已经标注好的“文字-图像”训练数据。一方面对文字进行模型训练。另一方面对图像进行另一个模型的训练，不断调整两个模型内部参数，使得模型分别输出的文字特征值和图像特征值并确认匹配。
3	Diffusion	2022 1、通过增加噪声破坏训练数据来学习，然后找出如何逆转这种噪声过程以恢复原始图像。 2、经过训练，该模型可以应用这些去噪方法，从随机输入中合成新的“干净”数据。

资料来源：腾讯科技、西部证券研发中心

CLIP 模型能够将文字和图像进行关联，比如将文字“狗”和狗的图像进行关联，并且关联的特征非常丰富。因此，CLIP 模型具备两个优势：1) 同时进行自然语言理解和计算机视觉分析，实现图像和文本匹配。2) 为了有足够多标记好的“文本-图像”进行训练，CLIP 模型广泛利用互联网上的图片，这些图片一般都带有各种文本描述，成为 CLIP 天然的训练样本。据腾讯科技公众号显示，CLIP 模型搜集了网络上超过 40 亿个“文本-图像”训练数据，这为后续 AIGC 尤其是输入文本生成图像/视频应用的落地奠定了基础。

在预测阶段，也是通过一系列生成的文本对和目标图像，计算 cosine similarities 从而获取预测值。

图 2：CLIP 模型原理

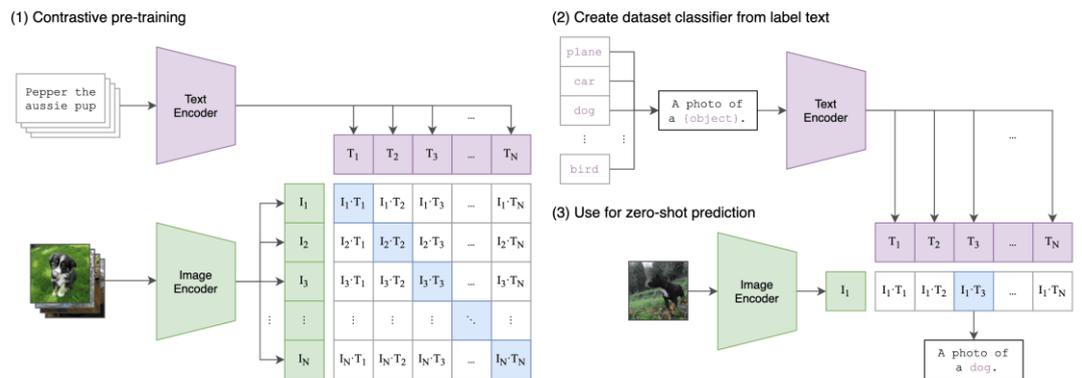


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

资料来源：HIGH-FLYER、西部证券研发中心

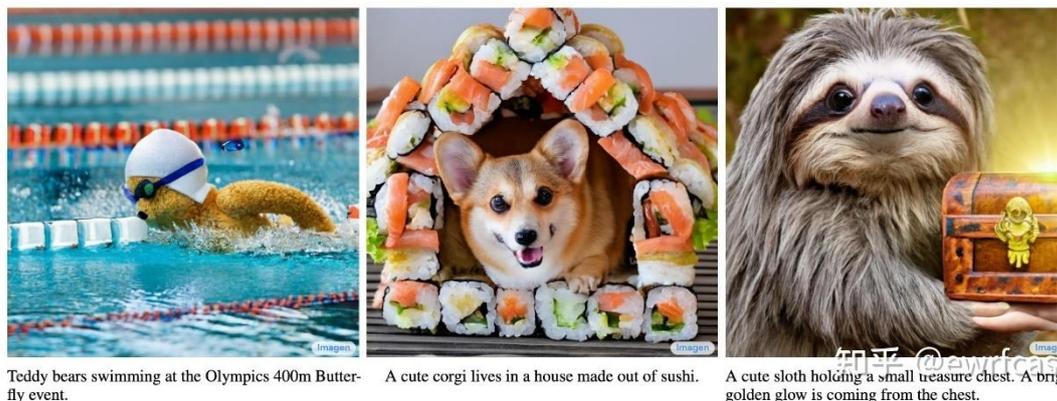
随后出现的 Diffusion 扩散化模型，则真正让文本生成图像的 AIGC 应用为大众所熟知，也是 2022 年下半年 Stable Diffusion 应用的重要推手。

Diffusion 模型有两个特点：1) 给图像增加高斯噪声，通过破坏训练数据来学习，然后找出如何逆转这种噪声过程以恢复原始图像。经过训练，该模型可以从随机输入中合成新的数据。2) Stable Diffusion 把模型的计算空间从像素空间经过数学变换，降维到一个可能性空间(Latent Space)的低维空间里，这一转化大幅降低了计算量和计算时间，使得模型训练效率大大提高。这算法模式的创新直接推动了 AIGC 技术的突破性进展。

一句话概括 diffusion model，即存在一系列高斯噪声 (T 轮)，将输入图片 X₀ 变为纯高斯噪声 X_t。而我们的模型则负责将 X_t 复原回图片 X₀。这样一来其实 diffusion model 和 GAN 很像，都是给定噪声 X_t 生成图片 X₀，但是要强调的是，这里噪声 X_t 与图片 X₀ 是同维度

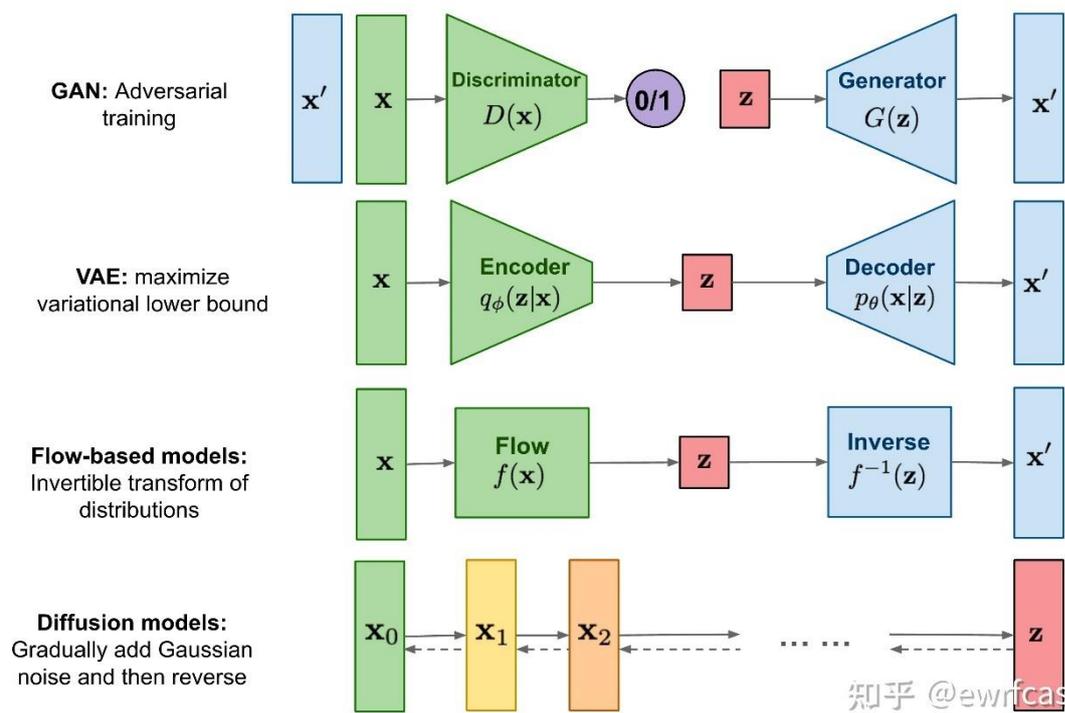
的。

图 3: Imagen 生成结果



资料来源：计算机视觉论文速递、西部证券研发中心

图 4: 各类生成模型对比图



资料来源：计算机视觉论文速递、西部证券研发中心

二、AIGC应用场景广泛

2.1 文本生成：文本交互成为未来发展方向

应用型文本：大多为结构化写作，以客服类的聊天问答、新闻撰写等为核心场景。最为典型的是基于结构化数据或规范格式，在特定情景类型下的文本生成，如体育新闻、金融新闻、公司财报、重大灾害等简讯写作。Narrative Science 创始人甚至曾预测，到 2030 年，90% 以上的新闻将由机器人完成。

图5：澜舟科技文本续写功能



资料来源：澜舟科技官网、西部证券研发中心

创作型文本：主要适用于剧情续写、营销文本等细分场景等，具有更高的文本开放度和自由度，需要一定的创意和个性化，对生成能力的技术要求更高。我们使用了市面上的小说续写、文章生成等 AIGC 工具。发现长篇幅文字的内部逻辑仍然存在较明显的问题、且生成稳定性不足，尚不适合直接进行实际使用。

除去本身的技术能力之外，由于人类对文字内容的消费并不是单纯理性和基于事实的，创作型文本还需要特别关注情感和语言表达艺术。

图6：彩云小梦剧情续写功能



资料来源：彩云小梦官网、西部证券研发中心

文本辅助：除去端到端进行文本创作外，辅助文本写作其实是目前国内供给及落地最为广泛的场景。主要为基于素材爬取的协助作用，例如定向采集信息素材、文本素材预处理、自动聚类去重，并根据创作者的需求提供相关素材。

图 7：写作猫文本辅助功能



资料来源：写作猫官网、西部证券研发中心

文本交互：例如虚拟伴侣、游戏中的 NPC 个性化交互等。2022 年夏季上线的社交 AIGC 叙事平台 Hidden Door 以及基于 GPT-3 开发的文本探索类游戏 AI dungeon 均已获得了不错的消费者反馈。

案例：小冰发布小冰岛 APP，每个用户均可创造自己的岛屿，并连带拥有一个功能类似于微信和 LINE 等社交产品的完整社交交互界面。用户不仅能在岛屿中体验丰富的视觉和自然音场，也不仅能与人工智能个体进行对话，而是再造了完整的一对一对话、群聊、朋友圈和技能生态体验。

图 8：小冰岛 AI 聊天界面



资料来源：小冰岛 APP、西部证券研发中心

2.2 音频生成：AI编曲将成为AI音频生成中的快速成长赛道

TTS (Text-to-speech) 场景：泛应用于客服及硬件机器人、有声读物制作、语音播报等任务。例如倒映有声与音频客户端“云听”APP 合作打造 AI 新闻主播，提供音频内容服务的一站式解决方案，以及喜马拉雅运用 TTS 技术重现单田芳声音版本的《毛氏三兄弟》和历史类作品。这种场景为文字内容的有声化提供了规模化能力。

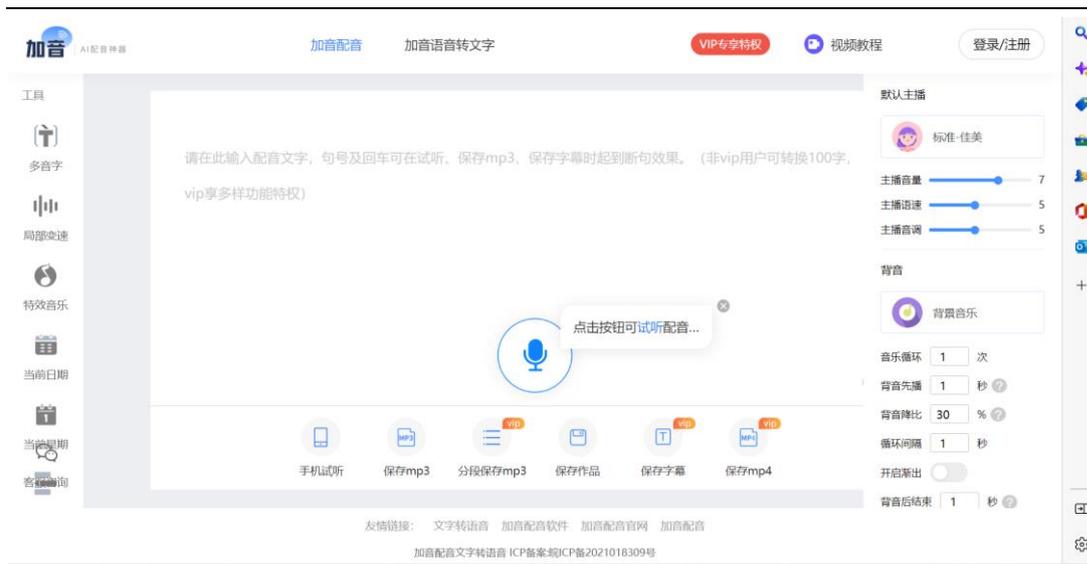
图9：文字转音频功能展示



资料来源：倒映有声官网、西部证券研发中心

随着内容媒体的变迁，短视频内容配音已成为重要场景。部分软件能够基于文档自动生成解说配音，上线有150+款包括不同方言和音色的AI智能配音主播。代表公司有九锤配音、加音、XAudioPro、剪映等。

图10：AI智能配音功能



资料来源：加音官网、西部证券研发中心

乐曲/歌曲生成：AIGC在词曲创作中的功能可被逐步拆解为作词（NLP中的文本创作/续写）、作曲、编曲、人声录制和整体混音。目前而言，AIGC已经支持基于开头旋律、图片、文字描述、音乐类型、情绪类型等生成特定乐曲。

通过这一功能，创作者即可得到AI创作的纯音乐或乐曲中的主旋律。2021年末，贝多芬管弦乐团在波恩首演人工智能谱写完成的贝多芬未完成之作《第十交响曲》，即为AI基于对贝多芬过往作品的大量学习，进行自动续写。

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

【西部证券】AIGC行业专题报告：2023年有望成为AIGC的拐点.pdf

请登录 <https://shgis.com/post/1361.html> 下载完整文档。

手机端请扫码查看：

