



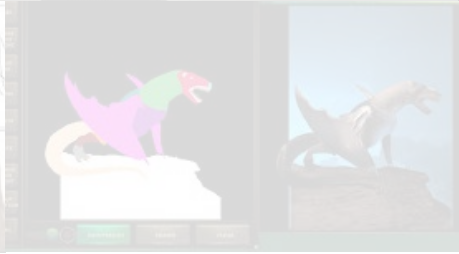
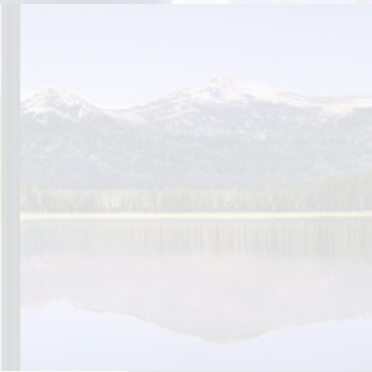
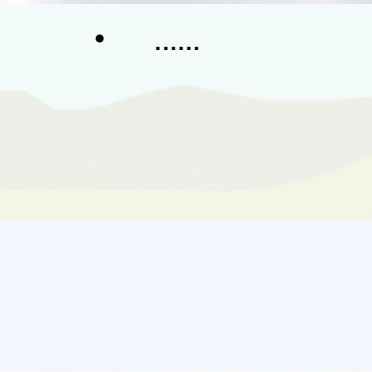
AIGC, AI-GENERATED CONTENT

AIGC/AI生成内容

INDUSTRY INSIGHT REPORT

产业展望报告

- 在技术上，AIGC已经可以完成哪些创作？
- 在价值上，AIGC除了直接生成艺术作品还能做什么？
- 在未来，AIGC将如何改变内容及相关产业？
-



目录



技术及场景篇

- 01 技术定义
- 02 技术及落地场景介绍—— 基于模态
 - 02 文本生成
 - 05 音频及文字—音频生成
 - 06 图像生成
 - 09 视频生成
 - 11 图像、视频、文本间跨模态生成
 - 15 策略生成
- 15 技术及落地场景介绍—综合赛道
 - 15 Game AI
 - 17 虚拟人生成
- 19 相关预测



产业篇

- 20 价值分析
- 21 细分赛道分析
 - 21 内容领域
 - 24 延展领域
- 26 行业发展阶段分析
- 29 产业链分析
- 29 玩家分析

如需进入AIGC相关讨论群、获取相关访谈实录等，请扫码添加量子位智库小助手；如您深耕于AIGC领域，希望和分析师进一步交流或联系业务，请扫码添加分析师（注明身份及来意）



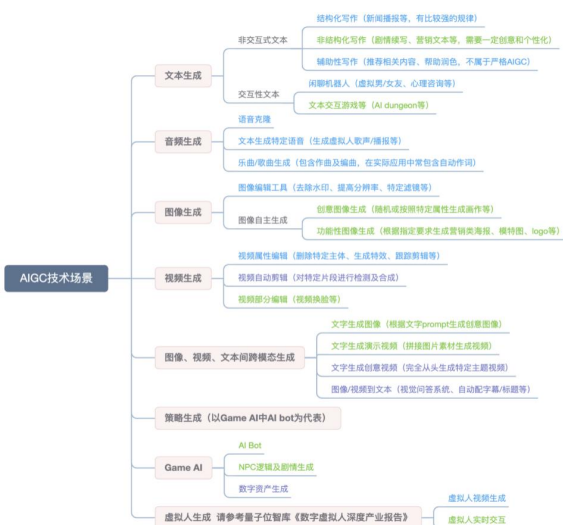
智库小助手



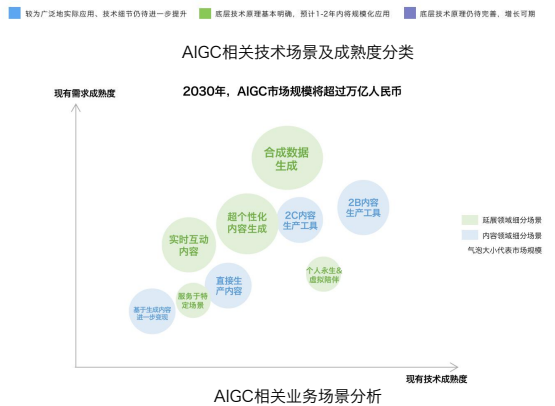
智库负责人

关键结论

1. 区别于Generative AI和Synthetic media, AIGC中的跨模态生成和策略/线索生成应当得到重视, 这两者也是AIGC长期的重要价值增长点。
2. 按照模态区分, AIGC可分为音频生成、文本生成、图像生成、视频生成及图像、视频、文本间的跨模态生成, 细分场景众多, 跨模态生成需要重点关注。Game AI和虚拟人生成是目前变现较为明确的两大综合赛道, 预计对原行业产生整体性影响。
3. 以下技术要素值得关注: 长/开放式文本生成、NeRF模型、Diffusion模型、跨模态大型预训练模型、小样本学习及自监督算法、强化学习及环境学习。以下技术场景值得关注: 闲聊式文本生成、个性化营销文本、富情感及细节TTS、拼凑式视频生成、基于文本的AI绘画、语音复刻。
5. 除降本增效、提供参考外, AIGC的以下两点价值更值得关注: 对不同模态元素进行二次拆解组合, 改变内容生产逻辑及形式; 和其他AI系统或数据库进行联动, 有实现高度个性化/高频优化。
6. 我国AIGC行业仍处于刚刚起步阶段, 距离大规模验证和体系化发展仍有距离, “模块分拆+个性化推荐”的“泛AIGC”形式预计将有所发展。



我国AIGC行业阶段划分



我国AIGC产业链划分

感谢以下机构及个人参加深度访谈：倒映有声、Deepmusic、高林教授、黄民烈教授、rct AI、腾讯AI Lab、小冰公司



技术及场景篇

1.1 技术定义

AIGC全称为**AI-Generated Content**，指基于生成对抗网络GAN、大型预训练模型等人工智能技术，通过已有数据寻找规律，并通过适当的泛化能力生成相关内容的技术。与之相类似的概念还包括Synthetic media，合成式媒体，主要指基于AI生成的文字、图像、音频等。

Gartner也提出了相似概念Generative AI，也即生成式AI。生成式AI是指该技术从现有数据中生成相似的原始数据。相较于量子位智库认为的AIGC，这一概念的范围较狭窄。

Gartner《2021年预测：人工智能对人类和社会的影响》给出积极预测：

- 至2023年将有20%的内容被生成式AI所创建
- 至2025年，Gartner预计生成式AI产生的数据将占有所有数据的10%，而今天这个比例不到1%
- 根据Gartner披露的“人工智能技术成熟度曲线”，生成式AI仍处于萌芽期，但其广阔的应用场景和巨大需求空间吸引着大量资本和技术的投入，预计将在2-5年内实现规模化应用

一方面，这一概念忽略了**跨模态生成（如基于文本生成图像或基于文本生成视频）**这一愈加重要的AIGC部分。我们会在下一部分对跨模态生成进行重点讲解。另一方面，在结合现有技术能力和落地场景进行分析后，我们认为“生成”和“内容”都应该采取更为广泛的概念。例如，生成中可以包含**基于线索的部分生成、完全自主生成和基于底稿的优化生成**。内容方面，不仅包括常见的图像、文本、音频等**外显性内容**，同样也包括策略、剧情、训练数据等**内在逻辑内容**。

从特定角度来看，AI内容生成意味着AI开始在现实内容中承担新的角色，从“观察、预测”拓展到“直接生成、决策”。

从商业模式来看，我们认为，AIGC本质上是一种**AI赋能技术**，能够通过其高通量、低门槛、高自由度的生成能力广泛服务于各类内容的相关场景及生产者。因此，我们不会将其定义为PGC\UGC之后的**新内容创作模式**，而是认为其在商业模式上会有大量其他交叉。我们会在价值篇对其商业模式进行进一步展开。

1.2 技术及落地场景介绍——基于模态

我们认为，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。

技术视角下，我们认为以下场景将成为未来发展的重点：**文本-图像-视频的跨模态生成、2D到3D生成、多模态理解结合生成**。后文将对以上技术场景的技术原理、现有进展、关键瓶颈等进行展开。

商业视角下我们认为，**未来3年内，虚拟人生成和游戏AI这两种综合性的AIGC场景将趋于商业化成熟**。

在这一部分，我们会基于不同的模态介绍对应的技术和商业场景，并给出各场景下的代表机构及发展现状。**下图中的绿色部分，是我们认为2-3年内具有快速增长潜力的细分赛道**。具体原因我们将在对应位置进行具体展开。



■ 较为广泛地实际应用、技术细节仍待进一步提升 ■ 底层技术原理基本明确，预计1-2年内将规模化应用 ■ 底层技术原理仍待完善，增长可期

📄 文生成

以结构性新闻撰写、内容续写、诗词创作等细分功能为代表，基于NLP技术的文本生成可以算作是AIGC中发展最早的一部分技术，也已经在新闻报道、对话机器人等应用场景中大范围商业落地。

量子位智库在此看好**个性化文本生成**以及**实时文字生成交互**。

• 细分技术介绍

大型预训练模型成为当下主流做法与关键技术节点。

一方面，2020年，1750亿参数的GPT-3在问答、摘要、翻译、续写等语言类任务上均展现出了优秀的通用能力，证明了“大力出奇迹”在语言类模型上的可行性。自此之后，**海量数据、更多参数、多元的数据采集渠道**等成为国内清华大学、智源研究院、达摩院、华为、北京大学、百度等参与者的关注点。

目前，**大型文本预训练模型作为底层工具，商业变现能力逐渐清晰**。以GPT-3为例，其文本生成能力已被直接应用于Writesonic、Conversion.ai、Snazzy AI、Copysmith、Copy.ai、Headlime等文本写作/编辑工具中。同时也被作为部分文本内容的提供方，服务于AI dungeon等文本具有重要意义的延展应用领域。

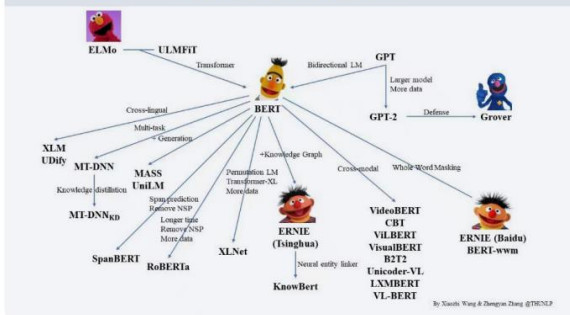
另一方面，以Transformer架构为重要代表，相关的底层架构仍在不断精进。研究者们正通过增加K-adapter、优化Transformer架构、合理引入知识图谱及知识库、增加特定任务对应Embedding等方式，增加文本对于上下文的理解与承接能力、对常识性知识的嵌入能力、中长篇幅生成能力、生成内容的内在逻辑性等。

Transformer架构

引入了注意力机制的Transformer架构能够基于每两个单词间的关系进行建模，有效理解单词在上下文中的意思，支持并行训练，使语言模型的训练效果达到了新高度。

在Transformer架构中，分为从左向右的自回归系列（GPT-3，和生成任务逻辑相同，特别适用于生成性任务），双向Transformer+Mask的自编码系列（BERT系列，更适用于自然语言理解），Encoder-Decoder架构（T5，分别使用双/单向Attention，适用于条件文本生成）

预训练模型演变历史



以下为各技术场景下对应的基本原理。

细分场景	技术原理
内容续写 如完形填空和文章续写	通过随机Mask（即遮挡）数据库文本中的词语或语段，让神经网络自主学习复原被遮挡部分，从而拥有“猜测”缺失内容的能力，产出预训练模型。再通过大规模预训练模型理解上文或给定条件，从概率层面推测最符合要求的输出结果。 其本质是借助超大规模的训练参数猜测上下文的过程。
摘要/标题生成 以TLDR为重要代表	首先通过词嵌入(Word Embedding)将字、词、句进行区分，然后基于特征评分、序列标注、分类模型等提取内容特征计算相关文本单元权重；其次选择相应的文本单元子集组成摘要候选集，完成内容选择；最后是针对字数要求等限定条件，对候选集的内容进行整理形成最终摘要，完成内容组织。 其细分路径又包含生成式文本摘要（AATS），即形成抽象认知并创造新词灵活概括，和抽取式文本摘要（EATS），即直接抽取原始素材并拼接成简单概要。
文本风格迁移 实现情绪、时态、性别、政治倾向等的分离及迁移	主流思路是分离文本属性及文本内容。 隐式方法即使用某类无监督学习学习或强化学习模式将文本属性及内容自动分离，常见的有生成对抗方式，即通过GAN实现目标属性和文本属性完全由不同的编码控制的状态。 显式方法首先寻找并删除代表文风的短语，其次检索与目标文风最匹配的相似短语，最后生成目标语句并保证语句通顺、原意不变。要实现多种风格的转化，典型方法有在通用语料库上预训练基于Transformer的语言模型来初始化编码器-解码器，然后以多种风格语言模型作为鉴别器来增强其对多个目标风格维度的转换能力。
整段文本生成 对话式&结构性文本生成	对话式文本生成适用于智能客服等任务型和闲聊型机器人等非任务型人机交互场景，可分类为管道模式及端对端模式。 管道模式即将对话拆分成四个主要模块（自然语言理解、历史状态追踪、对话策略选择、自然语言生成）并分别进行模型训练。 端对端模式将对话过程转化为历史对话信息到系统回复的映射问题，利用一个Seq2Seq框架构建整个对话系统，并利用强化学习摆脱大量标注数据的限制，通过离散隐向量学习句子之间的依赖关系。 结构性的文本生成，首先通过注意力机制、多层感知器等系统进行语句内容预选，对数值、时间等类型数据进行推理，增强数据间的结构信息；其次通过Transformer等模式结合上下文进行推导，控制句法及文本连贯性，将语义与句法统一分析，最后采用Seq2Seq等模式，以BiLSTM为基础构建文本生成器，生成最终文本。 目前而言，文本生成普遍具有上下文间逻辑问题、关键信息位置混淆、内容无中生有等问题

从现有的落地场景来看，我们将其划分为**应用型文本和创作型文本生成**，前者的进展明显优于后者。此外，从应用推广的角度来说，**辅助文本创作是目前落地最为广泛的场景。**

• 应用型文本生成

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。2015年发展至今，商业化应用已较为广泛，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。据分析师评价，由AI完成的新闻初稿已经接近人类记者在30分钟内完成的报道水准。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。

在结构化写作场景下，代表性垂直公司包括Automated Insights（美联社Wordsmith）、Narrative Science、textengine.io、AX Semantics、Yseop、Arria、Retresco、Viable、澜舟科技等。同时也是小冰公司、腾讯、百度等综合性覆盖AIGC领域公司的重点布局领域。

• 创作型文本生成

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。**据聆心智能创始人黄民烈教授介绍，目前文字生成主要捕捉的是浅层次，词汇上统计贡献的问题。但长文本生成还需要满足语义层次准确、在篇章上连贯通顺的要求，长文本写作对于议论文写作、公文写作等等具有重要意义。**未来四到五年，可能会出现比较好的千字内容。**

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。**我们认为，短期内创作型文本**更适合在特定的赛道下，基于集中的训练数据及具体的专家规则进行场景落地。**

在创作型文本领域，代表性的国内外公司包括Anyword、Phrasee、Persado、Pencil、Copy.ai、Friday.ai、Retresco、Writesonic、Conversion.ai、Snazzy AI、Rasa.io、LongShot.AI、彩云小梦等。

• 文本辅助生成

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。尽管目前能够提升生产力，但我们认为相对于实现技术意义上的AI生成，能否结合知识图谱等提供素材联想和语句参考等更具有实用意义。

这部分的国内代表产品包括写作猫、Gilso写作机器人、Get写作、写作狐、沃沃AI人工智能写作。

• 重点关注场景

长期来看，我们认为**闲聊型文本交互**将会成为重要潜在场景，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。随着线上社交逐渐成为了一种常态，社交重点向转移AI具有其合理性，**我们预估未来1-2年内就会出现明显增长。**目前较为典型的包括**小冰公司推出的小冰岛**，集中在精神心理领域的**聆心智能**、开发了AI dungeon的**Latitude.io**等。

代表公司：聆心智能

由国内NLP、对话系统领域专家黄民烈教授创办，专注于针对精神心理的AI驱动的高质量数字疗法。在技术上，聆心智能让AI围绕认知、情绪和行为三个维度对进行评估和干预。以多模态对话系统为核心交互框架，通过丰富的策略设计，让AI和用户产生足够的情感链接，生成用户个人模型及千人千面的治疗方案，解决用户的情绪困扰，完成心理治疗。目前，聆心智能已与北医六院、湘雅二院、中日友好医院、清华大学玉泉医院等达成合作。

此外，量子位智库重点推荐AI写作在个性化内容营销中的应用场景。该场景的应用与经济价值目前已经在国外Oneword等公司中得到明确验证，而国内目前尚未出现相关布局。有关这一场景，我们将在产业篇的延展细分赛道进行详细展开。

🎵 音频及文字—音频生成

整体而言，此类技术可应用于流行歌曲、乐曲、有声书的内容创作，以及视频、游戏、影视等领域的配乐创作，大大降低音乐版权的采购成本。我们目前最为看好的场景是自动生成实时配乐、语音克隆以及心理安抚等功能性音乐的自动生成。

• TTS (Text-to-speech) 场景

TTS在AIGC领域下已相当成熟，广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。

目前技术上的关键，在于如何通过富文本信息（如文本的深层情感、深层语义了解等）更好的表现其中的抑扬顿挫，以及基于用户较少的个性化数据得到整体的复制能力（如小样本迁移学习）。基于深度学习的端到端语音合成模式也正在逐步替代传统的拼接及参数法，代表模型包括WaveNet、Deep Voice及Tacotron等。

目前的垂直代表公司包括倒映有声、科大讯飞、思必驰（DUI）、Readspeaker、DeepZen和Sonantic。

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。

在TTS领域，语音克隆值得特别关注。语音克隆是本质上属于指定了目标语音（如特定发言人）的TTS。技术流程如下：

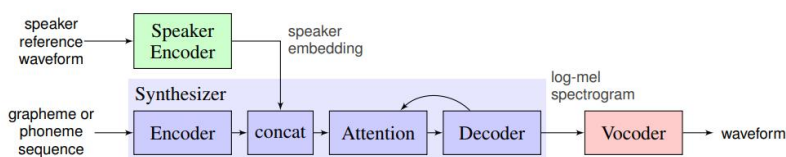


Figure 1: Model overview. Each of the three components are trained independently.

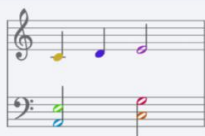
该技术目前被应用于虚拟歌手演唱、自动配音等，在声音IP化的基础上，对于动画、电影、以及虚拟人行业有重要意义。代表公司包括标贝科技、Modulate、overdub、replika、Replica Studios、Lovo、Voice mod、Resemble Ai、Respeecher、DeepZen、Sonantic、VoicelD、Descript。

• 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。

其中，AI作曲可以简单理解为“以语言模型（目前以Transformer为代表，如谷歌Magenta、OpenAI Jukebox、AIVA等）为中介，对音乐数据进行双向转化（通过MIDI等转化路径）”。此方面代表性的模型包括MelodyRNN、Music Transformer。据Deepmusic介绍，为提升整体效率，在这一过程中，由于相关数据巨大，往往需要对段落、调性等高维度的乐理知识进行专业提取，而节奏、音高、音长等低维度乐理信息由AI自动完成提取。

以音乐创作模型MusicAutobot为例, Transformer跨模态应用以文本为转化中介



Tokenized:

```
xxbos xxpad n60 d4 n52 d8 n45 d8
xxsep d4 n62 d4
xxsep d4 n64 d8 n55 d8 n48 d8
```

- 将MIDI音乐数据转换为语言模型可识别的编码文本
- 利用Transformer-XL架构, 训练能够自动生成token编码数据的模型
- 基于和第一步相同的规则, 将生成的token编码结果转化为MIDI音乐数据
- 完成编曲

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。

AI编曲则指对AI基于主旋律和创作者个人的偏好, 生成不同乐器的对应和弦(如鼓点、贝斯、钢琴等), 完成整体编配。在这部分中, 各乐器模型将通过无监督模型, 在特定乐曲/情绪风格内学习主旋律和特定要素间的映射关系, 从而基于主旋律生成自身所需和弦。对于人工而言, 要达到乐曲编配的职业标准, 需要7-10年的学习实践。

人声录制则广泛见于虚拟偶像的表演现场(前面所说的语音克隆), 通过端到端的声学模型和神经声码器完成, 可以简单理解为将输入文本替换为输入MIDI数据的声音克隆技术。混音指将主旋律、人声和各乐器和弦的音轨进行渲染及混合, 最终得到完整乐曲。该环节涉及的AI生成能力较少。

该场景下的代表企业包括Deepmusic、网易-有灵智能创作平台、Amper Music、AIVA、Landr、IBM Watson Music、Magenta、Loudly、Brain.FM、Splash、Flow machines。其中, 自动编曲功能已在国内主流音乐平台上线, 并成为相关大厂的重点关注领域。以QQ音乐为例, 就已成为Amper music的API合作伙伴。

对这一部分工作而言, 最大的挑战在于音乐数据的标注。在标注阶段, 不仅需要按时期、流派、作曲家等特征, 对训练集中乐曲的旋律、曲式结构、和声等特征进行描述, 还要将其有效编码为程序语言。此外, 还需要专业人员基于乐理进行相关调整润色。以Deepmusic为例, 音乐标注团队一直专注在存量歌曲的音乐信息标注工作上, 目前已经形成了全球最精确的华语歌曲音乐信息库, 为音乐信息检索(MIR)技术研究提供数据支持。

代表公司: Deepmusic

公司致力于运用AI技术从作词、作曲、编曲、演唱、混音等方面全方位降低音乐创作及制作门槛, 为音乐行业提供新的产品体验, 提升效率。产品包括针对视频生成配乐的配乐猫、支持非音乐专业人员创作的口袋音乐、可AI生成歌词的LYRICA、AI作曲软件LAZYCOMPOSER。目前已与国内多家音乐平台厂商达成合作。

场景推荐

我们认为, 以乐曲二创、辅助创作等场景为代表, AI编曲将在短期内成为AI音频生成中的快速成长赛道。特别是由于可以指定曲目风格、情绪、乐器等, AIGC音乐生成对于影视剧、游戏等多样化、乃至实时的背景音乐生成有重要意义。



图像生成

· 关键技术介绍

生成对抗网络GAN — 图像生成传统思路

GAN指Generative Adversarial Nets, 生成式对抗网络, 由生成器和判别器两部分组成, 生成器将抓取数据、产生新的生成数据, 并将其混入原始数据中送交判别器区分。这一过程将反复进行, 直到判别器无法以超过50%的准确度分辨出真实样本。

相较于2014年初现期，现有的GAN在神经网络架构、损失函数设计、模型训练稳定性、模型崩溃问题上取得了相应突破，提升了最终图像的特定细节、内在逻辑、生成速度等。但要在实际应用中大规模稳定应用，GAN仍需解决以下问题：

GAN领域仍待解决的问题及创新思路

• 解决训练不稳定问题

GAN 模型的相互博弈过程容易造成训练不稳定，使得训练难以收敛。近期突破思路有Relativistic GAN。有别于传统 GAN 测量生成样本是否为真的概率这一做法，Relativistic GAN 将关注点放在测量生成样本比真实样本“更真”的概率，这一尝试使得 GAN 获得了更好的收敛性。

• 缓解GAN的模式坍塌 model collapse，即生成样本大量重复相似

模式坍塌被认为是应用 GAN 进行图像生成时最难解决的问题之一，它会造成训练结果冗余、生成图像质量差、样本单一等问题。近期突破性思路有包含两个判别网络的D2GAN。

• GAN的结构改善

将 GAN 与机器学习中最理论成果进行交叉训练，引入迁移学习、强化学习等，使 GAN 与计算机图形学等技术更好地融合，推动明显结构的改善，有望成为下一步的研究热点。

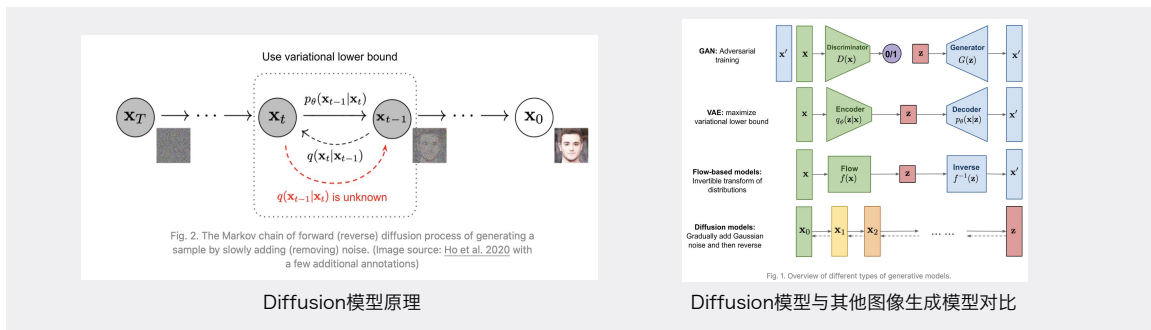
• GAN的压缩

目前图像生成技术想要落地，势必要根据需求调整模型的大小，结合基于“知识蒸馏”的模型压缩方法，进行匹配需求的优化和压缩，从而内嵌入小型软件中，拓宽应用领域同样为当前的研究热点。

扩散模型Diffusion Model — 新一代图像生成主流模型

2022年，Diffusion model成为图像生成领域的重要发现，甚至有超越GAN的势头。尽管大多通过Imagen、DALL · E 2等项目知名，但Diffusion本身服务于图像生成。我们将在此对Diffusion model进行介绍。

Diffusion是指扩散模型。最初受到热力学概念的启发，扩散模型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。一旦经过训练，扩散模型就可以应用这些去噪方法从随机输入中合成新颖的“干净”数据。



相较于其他的图像生成模型（比如GAN、VAE和基于流的模型），在所需数据更少的背景下，Diffusion Model的图像生成效果有明显提升。我们将在跨模态部分对基于Diffusion model的知名模型，如Guide Diffusion、Disco Diffusion、DALL · E2等进行具体介绍。

神经辐射场模型NeRF — 符合3D内容消费趋势的新一代模型

NeRF通过将场景表示为隐式的神经辐射场，渲染时通过神经网络查询位置上的场景信息生成新视角图像。简单来说，NeRF利用深度学习完成了计算机图形学中的3D渲染任务。

此前，2D到3D生成的领域也包含GAN方面的尝试，比如英伟达20-21年推出的GANverse3D能够生成可自定义并生成动画的3D图形。但由于GAN在对抗训练中会受限样本特征，该模型当时仅适用于汽车、马匹和鸟类。而NeRF在很大程度上克服了这一问题。

NeRF模型基本原理

- 将场景的体积表示优化为向量函数，该函数由位置和视图方向组成的连续5D坐标定义。具体而言，是沿相机射线采样5D坐标，来合成图像
- 将场景表示参数化为一个完全连接深度网络（MLP），该网络将通过5D坐标信息，输出对应的颜色和体积密度值
- 通过体积渲染技术将这些值合成为RGB图像
- 基于可微分的渲染函数，通过最小化合成图像和真实图像之间的残差，优化场景表示

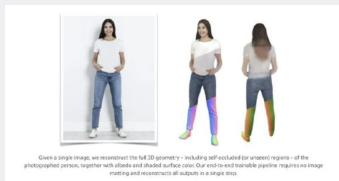
随着元宇宙等概念、深度相机等相关基础设施、VR眼镜等消费渠道的发展，3D内容生成将成为未来内容领域的必经之路，可以预见，通过降低对创作者3D建模能力和渲染能力的要求，**NeRF对于虚拟人创建、3D训练环境构建、增强现实、线上游戏及电影特效等都具有重要意义。**

自NeRF在ECCV 2020提出后，NeRF模型也持续在生成范围、生成效果、乃至于所需基础数据上进行改进。

例如包含光影变化效果（谷歌团队联合MIT、UC伯克利发布的NeRV）、动态NeRF（HyperNeRF和ST-NeRF），接近实时生成（英伟达的Instant-NGP和Instant NeRF），全场景NeRF（谷歌，mip-NeRF 360，对合成3D图像的视角和深度有了大幅提升。可以渲染完整的3D物体及场景，并支持在整个空间内从任何角度观看）、单张生成模型（2022年谷歌PHORHUM，甚至能够重现不可视区域；苏黎世联邦理工Pix2NeRF）、可进一步生成3D几何数据（英伟达及斯坦福，EG3D，可生成高分辨率人脸的几何结构）

量子位智库认为，其中的关键节点是能够基于单张图像生成3d模型，这预示着3D建模效率的明显提升。

基于单张图像生成3D模型



PHONUM



Pix2NeRF

尽管NeRF在3D内容生成领域受到了广泛关注，但具有创新架构设计的3D GAN依旧在扮演重要角色。例如，在以单张图像生成3D模型上，作为纯隐式神经网络的NeRF确定位置的速度较慢，计算量较大，会导致画质低、变形等问题。而英伟达与斯坦福大学推出的结合了显隐式神经网络特征的EG3D模型，在速度和内存消耗上都有明显改进。

• 技术及商业场景介绍

基于对不同技术原理的梳理，我们将图像生成领域的技术场景划分为**图像属性编辑、图像局部生成及更改、以及端到端的图像生成**。其中，前两者的落地场景为**图像编辑工具**，而端到端的图像生成则对应**创意图像及功能性图像生成两大落地场景**。

图像属性及部分编辑

属性编辑部分，可以直观的将其理解为**经AI降低门槛的PhotoShop**。目前而言，图片去水印、自动调整光影、设置滤镜（如Prisma、Versa、Vinci和Deepart）、修改颜色纹理（如DeepAI）、复刻/修改图像风格（DALL·E 2已经可以仅凭借单张图像进行风格复刻、NightCafe等）、提升分辨率等已经常见。

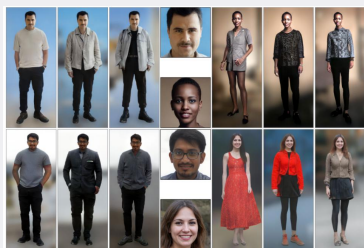
关于此类场景，初创公司数量庞大，Adobe等相关业务公司（Adobe Sensai）也在进入这一领域。我们认为，对于

此类业务而言，引流渠道和与技术流程挂钩的产品使用体验非常重要。现有的代表公司包括美图秀秀（美图AI开放平台）、Radius5、Photokit、Imglarger、Hotpot、Remove.bg、Skylum (Mask AI)、Photodiva。

图像部分编辑部分，指部分更改图像部分构成（如英伟达CycleGAN支持将图内的斑马和马进行更改）、修改面部特征（Metaphysics，可调节自身照片的情绪、年龄、微笑等；以Deepfake为代表的图像换脸）。

由于技术限制，图像的各部分之间需要通过对齐来避免扭曲、伪影等问题，目前GAN还难以直接生成高质量的完整图像。2019年，曾宣传能够直接生成完整模特图的日本公司DataGrid目前已无动向。但同时，也出现了由局部生成拼接为完整图像的生成思路。典型代表为选入CVPR2022的InsetGAN，该模型由Adobe推出。

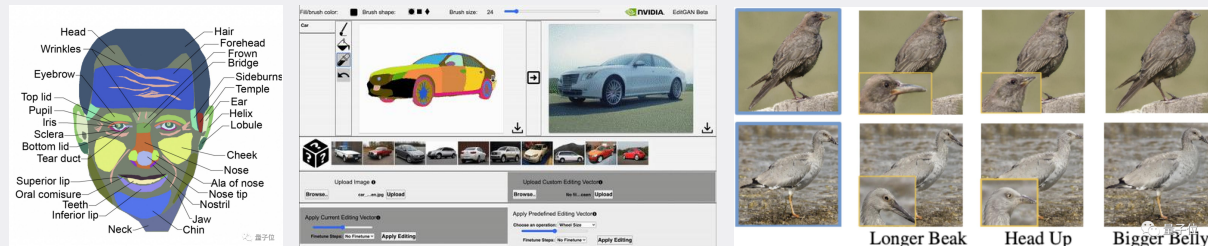
通过组合多个局部预训练模型生成完整图像的InsetGAN



该模型由两类GAN组成，全身GAN（Full-Body GAN），基于中等质量的数据进行训练并生成一个人体和部分GAN，其中包含了多个针对脸部、手、脚等特定部位进行训练的GAN。

该模型会通过检测部分GAN生成的特定区域在底层画布，也就是全身GAN生成的区域中的位置，经过裁剪后再将特定区域嵌入。

同时，细粒度、分区域的图像编辑能力也较为关键，代表为英伟达的EditGAN。该模型将需要编辑的原图像x嵌入到EditGAN的潜空间，借助语义分割图的相同潜码，将原图x分割成高度精细的语义块（segmentation mask），并得到分割图y。接着，使用简单的交互式数字绘画或标签工具进行手动修改。模型最终会共享潜码的优化，以保持新分割图与真实图像的RGB外观一致，如图所示：



EditGAN，支持对图像进行细节修改

图像端到端生成

此处则主要指基于草图生成完整图像（VansPortrait、谷歌Chimera painter可画出怪物、英伟达GauGAN可画出风景、基于草图生成人脸的DeepFaceDrawing）、有机组合多张图像生成新图像（Artbreeder）、根据指定属性生成目标图像（如Rosebud.ai支持生成虚拟的模特面部）等。

该部分包含两类场景，分别为**创意图像生成与功能性图像生成**。前者大多以NFT等形式体现，后者则大多以营销类海报/界面、logo、模特图、用户头像为主。垂直代表公司/产品包括Deepdream Generator、Rosebud.ai、AI Gahaku、artbreeder、nightcafe、starryai、wombo、deepart、obvious、阿里鹿班、ZMO.ai、Datagrid、诗云科技、道子智能绘画系统等。

由于图像的生成复杂度远高于文字，在整体生成上，目前仍然难以达到稳定可靠的生成高质量图像。但据高林教授评价，人脸生成的应用将预计有更快的发展。从VAQ、VAE等技术选型开始，人脸生成的技术研究已经有了较好的效

果，同时人脸数据集也较为充足。同时，单张的人脸生成价值相对有限。要进一步发挥其价值，可以考虑将其与NeRF、也即3D内容生成相结合，支持从不同的视角和动作还原特定对象面部，能够在发布会、面见客户等场景中具有重要作用。而对于近年视频换脸效果不佳的情况，高教授认为这与底层设计优化，例如除纹理相似度之外，在解编码中考虑更多的时间、动作、甚至情感等因素，并叠加考虑数据、渲染能力等因素。

视频生成

量子位智库认为，**视频生成将成为近期跨模态生成领域的中高潜力场景。其背后逻辑是不同技术带来的主流内容形式的变化。**本部分主要包括视频属性编辑、视频自动剪辑、视频部分编辑。

• 视频属性编辑

例如视频画质修复、删除画面中特定主体、自动跟踪主题剪辑、生成视频特效、自动添加特定内容、视频自动美颜等。代表公司包括Runway ML、Wisecut、Adobe Sensei、Kaleido、帝视科技、CCTV AIGC、影谱科技、Versa（不咕剪辑）、美图影像研究院等。

• 视频自动剪辑

基于视频中的画面、声音等多模态信息的特征融合进行学习，按照氛围、情绪等高级语义限定，对满足条件片段进行检测并合成。目前还主要在技术尝试阶段。典型案例包括Adobe与斯坦福共同研发的AI视频剪辑系统、IBM Watson自动剪辑电影预告片、以及Flow Machine。我国的影谱科技推出了相关产品，能够基于视频中的画面、声音等多模态信息的特征融合进行学习，按照氛围、情绪等高级语义限定，对满足条件片段进行检测并合成。

• 视频部分生成（以Deepfake为典型代表）

技术原理

视频到视频生成技术的本质是基于目标图像或视频对源视频进行编辑及调试，通过基于语音等要素逐帧复刻，能够完成人脸替换、人脸再现（人物表情或面部特征的改变）、人脸合成（构建全新人物）甚至全身合成、虚拟环境合成等功能。

其原理本质与图像生成类似，强调将视频切割成帧，再对每一帧的图像进行处理。视频生成的流程通常可以分为三个步骤，即数据提取、数据训练及转换。以人脸合成为例，首先需要对源人物及目标人物的多角度特征数据提取，然后基于数据对模型进行训练并进行图像的合成，最后基于合成的图像将原始视频进行转换，即插入生成的内容并进行调试，确保每一帧之间的流程度及真实度。目前的技术正在提升修改精准度与修改实时性两方面。

落地分析

在我们看来，该场景的底层商业逻辑与虚拟偶像类似。本质上是以真人的肖像权作为演员，实际表演者承担“中之人”的角色。其主要落地场景包含两方面：

一方面，可以选择服务于明星，在多语言广告、碎片化内容生成等领域使用，快速提升明星的IP价值。例如Synthesia为Snoop Dogg制作的广告，通过使用deepfake改变其嘴部动作，就能够将原始广告匹配到另一品牌。

另一方面，则可以服务于特定商务场景，例如培训材料分发（如WPP的全球培训视频）、素人直播及短视频拍摄等。由于技术要求，需要对最终脸部所有者进行大量数据采集，需要相关从业公司获取大量面部数据授权，针对市场需求进行相关运营，完善后续的配套监管和溯源措施。量子位智库接触了“中国马斯克”的创作团队，该团队目前已经将换脸能力在专业MCN中进行商业化。这可能会引入一种新的商业模式，基于生物资产（声音复刻/面孔授权）的创作者经济。与此同时，底层的IP授权和配套监管值得关注。

除了deepfake之外，我们还观察到了在视频中的虚拟内容植入，也即利用计算机图形学和目标检测在视频中生成物理世界并不存在的品牌虚拟元素，如logo、产品、吉祥物等。以国外公司Marriad为代表，该公司目前已经为腾讯视频服务，后者准备在插入虚拟资产的基础上，个性化展示广告。这将极大的简化商业化内容的生成过程。



图像、视频、文本间的跨模态生成

模态是指不同的信息来源或者方式。目前的模态，大多是按照信息媒介所分类的音频、文字、视觉等。而事实上，在能够找到合适的载体之后，很多信息，诸如人的触觉、听觉、情绪、生理指标、甚至于不同传感器所对应的点云、红外线、电磁波等都能够变为计算机可理解可处理的模态。

对人工智能而言，要更为精准和综合的观察并认知现实世界，就需要尽可能向人类的多模态能力靠拢，我们将这种能力称为多模态学习MML (Multi-modal Learning)，其中的技术分类及应用均十分多样。我们可以简单将其分为**跨模态理解**（例如通过结合街景和汽车的声音判断交通潜在危险、结合说话人的唇形和语音判定其说话内容）和**跨模态生成**（例如在参考其他图画的基础上命题作画；触景生情并创作诗歌等）。

多模态学习细分技术

- 归一 — 确定某一实体对应的各类多模态表示
- 转化 — 模态间信息翻译
- 翻译对齐 — 寻找成分间的关系
- 融合 — 联合多个模态信息，统一进行预测
- 协同学习 — 针对同一实体，模态信息之间互补，提升其泛化性能

多模态学习落地场景

- 自动标注字幕
- 自动生成图片解说/视频标题
- 图文互相索引/检索
- 基于图片的问答推理
- 视觉定位
- 视频抠图
- 多模态互动

Transformer架构的跨界应用成为跨模态学习的重要开端之一。

Transformer架构的核心是Self-Attention机制，该机制使得Transformer能够有效提取长序列特征，相较于CNN能够更好的还原全局。而多模态训练普遍需要将图片提取为区域序列特征，也即将视觉的区域特征和文本特征序列相匹配，形成Transformer架构擅长处理的一维长序列，对Transformer的内部技术架构相符合。与此同时，Transformer架构还具有更高的计算效率和可扩展性，为训练大型跨模态模型奠定了基础。

Vision Transformer将Transformer架构首次应用于图像领域。该模型在特定大规模数据集上的训练成果超出了ResNet。随后，谷歌的VideoBERT尝试了将Transformer拓展到“视频-文本”领域。该模型能够完成看图猜词和为视频生成字幕两项功能，首次验证了Transformer+预训练在多模态融合上的技术可行性。基于Transformer的多模态模型开始受到关注，ViLBERT、LXMERT、UNITER、Oscar等纷纷出现。

CLIP模型的出现，成为跨模态生成应用的一个重要节点。

CLIP, Contrastive Language - Image Pre-training, 由OpenAI在2021年提出，图像编码器和文本编码器以对比方式进行联合训练，能够链接文本和图片。可以简单将其理解为，利用CLIP测定图片和文本描述的贴切程度。

自CLIP出现后，“CLIP+其他模型”在跨模态生成领域成为一种较为通用的做法。以Disco Diffusion为例，该模型将CLIP模型和用于生成图像的Diffusion模型进行了关联。CLIP模型将持续计算Diffusion模型随机生成噪声与文本表征的相似度，持续迭代修改，直至生成可达到要求的图像。

除去图像领域，CLIP后续还在视频、音频、3D模型等领域扮演了关联不同模态的角色。例如入选CVPR2022，基于文本生成3D图像的Dreamfields（类似工作还包括CLIP-Forge）。不过目前，已经出现了在所需数据量和算力上表现更为优秀的匹配模型。例如南加州大学的TONICS。



在此基础上，大型预训练模型的发展重点开始向横跨文本、图像、语音、视频的全模态通用模型发展。通过计算策略、数据调用策略、深度学习框架等方法提升模型效果成为目前研究的进展关键。与此同时，覆盖更多模态的训练数据同样值得关注。例如，MultiBench提供了包括10个模态的数据集，PanoAVQA提供了360度视频数据，X-World提供用于自动驾驶的各类模态数据。目前，华为诺亚方舟已经开源了全球首个亿级中文多模态数据集“悟空”。

跨模态大型预训练模型的代表包括：开启了跨模态预训练模型的Open AI DALL·E及CLIP、NVIDIA GauGAN2、微软及北大 NÜWA女娲、NVIDIA PoE GAN、DeepMind的Gato、百度ERNIE-ViLG、Facebook及Meta的AV-HuBERT（基于语音和唇语输出文本）及Data2vec（横跨CV、NLP和语音）、中科院“紫东太初”、哥大及Facebook开发的VX2Text（基于视频、音频等输出文本）。



量子位智库认为，多模态能力的提升将成为A真正I实现认知智能和决策智能的关键转折点。在未来1-2年，“文字—图像”的生成将快速落地。目前，“文字—视频”的生成也已有相对理想的实验效果，三个模态的跨模态生成也已经开始尝试。

接下来，我们将区分具体模态，对跨模态生成领域的代表模型进行介绍

• 文字生成图像

2021年，OpenAI的CLIP和DALLE开启了AI绘画重要的一年。同年，CVPR2021收录的VQGAN也引发了广泛关注。

2022年被称为“AI绘画”之年，多款模型/软件证明了基于文字提示得到效果良好的图画的可能性，Diffusion Model受到广泛关注。

首先，OpenAI推出了GLIDE。GLIDE全称Guided Language to Image Diffusion for Generation and Editing，是一种扩散模型，参数仅35亿。支持CLIP引导（经训练后的噪声感知64×64 ViT-L CLIP模型）和无分类器引导，支持部分P图和迭代生成。

随后为Disco Diffusion，该免费开源项目搭载在Google Colab上，需要一定的代码知识，更擅长梦境感的抽象画面，在具象生成和较多的描述语句上效果较差。随后，Disco Diffusion的作者之一推出了AI绘画聊天机器人Midjourney。该软件搭载在Discord上，商业化和产品化更为成熟，并提出了明确的分润模式（商业变现达到两万美金后需要20%分润）。

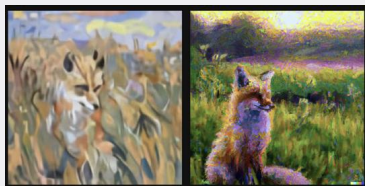
类似的软件及公司包括Big Sleep、StarryAI、WOMBO Dream。国内相关软件则包括Timmat，以及百度文心ERNIE-ViLG、小冰框架、悟道文澜、阿里M6等跨模态生成模型。

更擅长具象、对文本指令还原度更高的DALL·E2和Imagen AI证实了AI绘画的实际应用价值。但需要注意的是，两者的技术思路并不相同。尽管扩散模型等引发了巨大关注，但不同的技术思路同样呈现出了较好效果。目前尚无法确定未来AI绘画的关键技术里程碑。

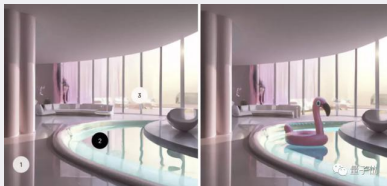
DALL·E 2

由OpenAI推出，目前尚未对外开放。整体而言，已经具备了相当的实用性。

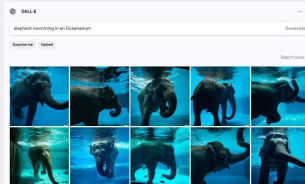
技术上是CLIP模型和Diffusion模型的结合。在这个名为unCLIP的架构中，CLIP文本嵌入首先会被喂给自回归或扩散先验，以产生一个图像嵌入。而后，这个嵌入会被用来调节扩散编码器，以生成最终的图像。相对于第一版DALL·E整体绘画水平有明显提升，画质提升为之前的四倍，并支持更细粒度的文本—图像生成功能（类似部分P图），渲染时间从数小时提升到了不足一分钟。



DALL·E (左) 和DALL·E (右) 同主题对比



文本P图时会注意倒影等细节



DALL·E 自主生成图像

Imagen AI

由谷歌推出，目前尚未对外开放。就公开作品而言，在写实场景上似乎表现更为出色。

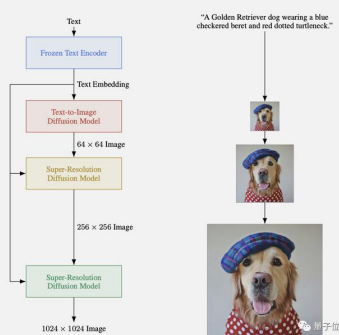
Imagen没有采取CLIP+图像模型的常用思路，而是由纯语言模型（T5-XXL），由图像生成模型完成文本到图像的转换工作。



Prompt: 一只非常快乐的毛茸熊猫打扮成了在厨房里做面团的厨师的高对比度画像，他身后的墙上还有一幅画了鲜花的画



Prompt: 带着宇航员头盔，在深夜眺望窗外的浣熊



• 文字生成视频

在一定程度上，文本生成视频可以看作是文本生成图像的进阶版技术。我们预估，AI绘画和AI生成视频将分别在3年和5年后迎来较为广泛的规模应用。

一方面，两者的本质比较接近。文本生成视频同样是以Token为中介，关联文本和图像生成，逐帧生成所需图片，最后逐帧生成完整视频。而另一方面，视频生成会面临不同帧之间连续性的问题。对生成图像间的长序列建模问题要求更高，以确保视频整体连贯流程。从数据基础来看，视频所需的标注信息量远高于图像。

按照技术生成难度和生成内容，可以区分为拼凑式生成和完全从头生成两种方式。

拼凑式生成的技术是指基于文字（涉及NLP语义理解）搜索合适的配图、音乐等素材，在已有模板的参考下完成自动剪辑。这类技术本质是“搜索推荐+自动拼接”，门槛较低，背后授权素材库的体量、已有模版数量等成为关键因素。目前已经进入可商用阶段，国外有较为成熟的产品。代表公司/产品方面，2C的包括百度智能视频合成平台VidPress、慧川智能、Gliacloud、Synths.video、lumen5，2B端代表公司为Pencil。

Gliacloud：输入文本链接，自动对其中的标题和文字进行区分表示。自动配素材和文字的出现方式，形成说明式的视频。能够增加10倍的视频产量。



Pencil：能够基于客户的品牌和资产自动生成副本、视频并完成相关广告创意



完全从头生成视频则是指由AI模型基于自身能力，不直接引用现有素材，生成最终视频。

该领域目前仍处于技术尝试阶段，所生成视频的时长、清晰度、逻辑程度等仍有较大的提升空间。以Cogvideo为例，该模型基于预训练文本-图像模型CogView2打造，一共分为两个模块。第一部分先基于CogView2，通过文本生成几帧图像，这时候合成视频的帧率还很低；第二部分则会基于双向注意力模型对生成的几帧图像进行插帧，来生成帧率更高的完整视频。

由于从静态内容生成进入到了动态生成阶段，需要考虑其中时序性、连续性的问题。视频生成对于内容生成领域将具有节点性意义。同时，由于视频中会包括文本中难以表现的逻辑或尝试，相较于图像或纯文本训练，视频预训练模型有助于进一步释放语言模型的能力。

其他相关预训练模型还包括NVIDIA推出的GauGAN、微软亚洲研究院推出的GODIVA、清华及智源研究院提出的VideoGPT、TGAN、Ground Truth等。

图像/视频到文本

具体应用包括视觉问答系统、配字幕、标题生成等，这一技术还将有助于文本—图像之间的跨模态搜索。代表模型包括METER、ALIGN等。除了在各个模态之间进行跨越生成之外，目前，包括小冰公司在内的多家机构已经在研究多模态

生成，同时将多种模态信息作为特定任务的输入，例如同时包括图像内的人物、时间、地点、事件、动作及情感理解、甚至包含背后深度知识等。以保证生成结果更加精准。

策略生成

主要指AI基于特定问题和场景，自主提出解决方案的过程。此类技术在游戏、自动驾驶、机器人控制等领域有极高的应用价值。此类AI模型普遍采用深度强化学习技术（早期曾采用决策树及监督学习），其本质原理是让AI感知环境、自身状态并基于特定目标决定当下需要执行的动作，然后基于下一个状态给到系统的反馈进行奖励或惩罚，最终使系统在不断的“强化”过程中优化“策略”。

Game AI是决策生成目前应用最为广泛和明确的场景。游戏本身为AI提供了最佳的研发场景，有大量数据和贴近实际场景（真人玩家玩游戏时所面对的环境）。而在其他业务场景中，还需先搭建能够充分还原现实因素的虚拟环境，并合理设置Reward等关键因素，目前距离现实应用仍较远。据悉，国内多家AI企业已在尝试这一方向，但如何精准完成环境学习仍然具有关键意义。

因此，AI的决策生成目前还在Game AI为主体应用。据采访，**预计策略生成功能大约在2年后可以在其他业务场景中稳定可靠落地。**

1.3 技术及落地场景介绍——综合赛道

在这部分，我们选取了Game AI及虚拟人生成这两个涉及了各方面AIGC能力的综合赛道，并认为其将是近期体系化发展最快的AIGC行业化赛道。

GameAI

以腾讯AI Lab在游戏制作领域的布局为例，人工智能在游戏前期制作、游戏中运营的体验及运营优化、游戏周边内容制作的全流程中均有应用。

我们将其中的核心生成要素提炼为AI Bot、NPC相关生成和相关资产生成。



腾讯AI Lab在游戏领域的全栈布局

• AI Bot，也即游戏操作策略生成

可以将其简单理解为人工智能玩家，重点在于生成真实对战策略。2016年Deepmind AlphaGO在围棋中有所展示，随后，AI决策在Dota2、StarCraft2、德扑、麻将等游戏领域中均展现出了良好的实力。

技术关键在于强化学习方案优化设计，体现为多智能体使用、可适应游戏/环境复杂度、具体策略多样性等。目前，包括网易在内的我国主流游戏公司已经形成共识，除去直接以陪玩等形式服务C端玩家，也在通过游戏跑图、平衡度等方式服务于游戏开发/运营等B端。

相关业务场景



前期平衡性测试

游戏策划会根据具体的游戏内容，为角色本身的属性、技能、状态等，以及道具、环境、货币等参数，设定一系列的数值。平衡性测试能够充分地模拟玩家在某一套数值体系下的游戏体验，提出优化策略，为玩家带来更加平衡的多样性游戏交互。之前需要在测试服上邀请人类玩家试玩1-2个月后才能得到结果。现在由AI Bot直接在内部完成相关工作即可。



游戏跑图/功能测试

通过AI bot针对性的找出游戏中所有交互的可能性，通过发现潜在漏洞辅助游戏策划



对局陪伴

包括平衡匹配、冷启动、玩家掉线接管等



特定风格模拟

在绝悟中，AI通过模仿职业选手，掌握他们的典型个人风格，玩家则感觉像在与真实的职业选手对抗



基于玩法教学的新型人机互动

在游戏内“绝悟试炼”玩法中，在玩家发出各种指令后，AI会根据血量、距离等实际情况，评估指令的合理性，选择执行或拒绝，身兼队友及老师，与玩家在真实对战环境中交流协作，并在过程中向玩家传授职业级的策略与操作技术，帮助玩家迅速熟悉英雄操作与游戏玩法。在引入王者绝悟AI教学后，玩家单局游戏主动沟通的次数有明显提升，提高了PVE玩法的可玩性

代表机构：腾讯AI Lab（腾讯“绝悟”）

「绝悟」AI通过强化学习的方法来模仿真实玩家，包括发育、运营、协作等指标类别，以及每分钟手速、技能释放频率、命中率、击杀数等具体参数，让AI更接近正式服玩家真实表现，将测试的总体准确性提升到95%。

目前腾讯绝悟在环境观测、图像信息处理、探索效率等方面的创新算法已经突破了可用英雄限制（英雄池数量从40增至100），让AI完全掌握所有英雄的所有技能并达到职业电竞水平，能应对高达10的15次方的英雄组合数变化。基于绝悟，王者荣耀的数值平衡性偏差已经从1.05%下降到0.68%，其所涉及的“多智能体”决策过程中，可以在高达10的20000次方种操作可能性的复杂环境中进行决策。

目前，腾讯AI Lab还与腾讯 AI Lab 还与王者荣耀联合推出了AI开放研究平台「开悟」，并积极举办相关赛事。

代表公司：超参数

估值已达独角兽，业内率先实现在3D FPS（游戏的帧数）领域的大规模商业化落地，服务对象包括数千千万级日活的游戏产品。超参数科技的AI服务已经为数千万流水超过10亿元的游戏产品贡献了巨大的商业价值，涵盖沙盒、开放世界、FPS、MOBA、休闲竞技等多个品类。

其AI Bot支持玩家陪玩、多人团队竞技、非完美信息博弈AI，并提供了自研小游戏《轮到你了》中的虚拟玩家。目前，AI bot已在多款千万日活的产品中上线，每日在线峰值值将近百万。游戏AI平台“Delta”已具备跨云调度超过50万核的计算能力，承载超过50万个AI并发在线，每天服务全球40余个国家的上亿玩家、提供数千亿次调用。

在我国，网易伏羲、商汤科技也已在业务布局中提及该部分业务。

• NPC逻辑及剧情生成，也即由AI生成底层逻辑

此前，NPC具体的对话内容及底层剧情需要人工创造驱动脚本，由制作人主观联想不同NPC所对应的语言、动作、操作逻辑等，这种动态的个性化匹配背后依旧是不同的静态分支，创造性及个性化相对有限。

而以rct AI的智能NPC为例，其NPC能够分析玩家的实时输入，并动态地生成交互反应，从而构建几乎无限且不重复的剧情，增强用户体验并延长游戏生命周期。特别是在养成类游戏中，AI所提供的个性化生成能够带来画面、剧情及具体交互的个性化全新游戏体验。而实时剧情生成则有助于在特定框架内生成全新的可能性，增加游戏整体的叙事可能性。

目前，NPC逻辑自动生成技术已经应用在《黑客帝国：觉醒》（Mass AI，车辆行人等将独立于玩家操控变动）、Red Dead Redemption 2、Monster Hunter: World等大型游戏中。长期来看，NPC的灵活自主将引领其成为重要的社交节点，有效填充元宇宙的内部架构。

代表公司：rct AI

rct AI属于业内较为少见的可以在逻辑侧进行AI生成开发的公司，其核心产品包括个性化且动态的剧情探索、智能NPC/ AI MetaBeings/ 智能环境，提供真人般的游戏对抗体验。

通过简单设计并调整不同的参数，rct AI的混沌球（Chaos Box）算法就可以在游戏中大规模地轻松生成具有智能意识的虚拟角色。他们的行为和对话不会重复，皆为动态生成。在游戏场景中，部署具有不同性格的智能 NPC，通过对话、行为等动态交互，增加玩家的游戏时长，同时提供新的变现途径。具体包括性格化NPC（不同的对话、行为等动态交互）、对抗式 AI（对战类场景中个性化对手）、互动式AI（具体涉及探索世界、技能训练、参与对抗、完成目标、闲聊陪伴、协作建造六种互动场景）、大规模智能 NPC 部署、智能留存及智能运营策略等。

• 数字资产生成，也即结合其他模态生成能力，落地为具体场景

我们将其进一步区分为NPC相关资产生成、场景生成和地图/关卡生成。

NPC相关资产生成

据腾讯AI Lab介绍，AIGC能够使相关资产的生成效率提高5-10倍，并同时增加用户内容的个性化

• 人物外观生成

不同的NPC需要有不同的面孔、服饰、声音甚至性格特征。传统方法生成NPC成本高，需要逐个进行原型设计、多次建模及渲染等。借助AIGC有助于实现低成本大规模生产NPC，实现千人千面。

• 人物动画生成

腾讯AI Lab整合文字转语音的前沿研究，生成自然语音，同步驱动嘴型、表情等面部变化，达到高度逼真。传统方法成本极高，且制作一分钟的人物动画需要耗费数天，基于AI生成技术可以将原本数周工作量减少到几个小时。

• 人物动作生成

利用相位神经网络技术控制人物运动，将原本数周工作量减少到几小时。AI在大量的运动数据集上学习人体动作，将捕获的运动数据与场景地形数据相匹配，完成准确、流畅的动作反应，如行走、跑步、跳跃等。

地图/关卡设计

创建游戏关卡也称为程序内容生成（PCG）。这些技术采用复杂的 AI 算法来生成巨大的开放世界环境、新游戏关卡和其他游戏资产。在腾讯AI Lab自研的2D小游戏里，只需要不到一百个训练关卡作为输入，AI即可在几个小时内生成一千多个可玩的、高质量、多样化的关卡。AI生成3D关卡相较之下较为复杂（AI生成3D关卡内容则要复杂得多，3D关卡涉及地形起伏、元素之间的关联、不同分辨率的元素、路线与区域规划等，更考验AI的生成能力），但目前腾讯AI Lab已实现基于约200个训练关卡及人类的标注编辑的输入，生成超过一万个3D关卡。

场景生成

具体包括3D环境生成等。例如 GAN Theft Auto中汽车移动时的阴影均是由AI生成的。该部分目前的技术仍有较大进步空间。

量子位智库关于Game AI领域有以下判断：

AIGC在游戏领域的应用可能会对游戏行业的商业模式产生影响，例如引入更多In-Game Purchase，通过游戏主线之外的其他设计增加用户留存时间，引导用户对游戏的部分内容而非整体进行付费。此外，这可能会导致游戏发行方与用户间的关系从产品交付（游戏本身）到全生命周期的个性化服务交付（提供更多个性化埋点）。

GameAI后续的行业发展关键在于**游戏行业能否接受，并从底层架构开始支持提供包括客户数据在内的相关接口**。目前来看，国内相关游戏厂商合作意愿良好，普遍存在内部开发或外部测试需求。

虚拟人生成

虚拟数字人指存在于非物理世界(如图片、视频、直播、一体服务机、VR)中，并具有多重人类特征的综合产物。

目前“深度合成+计算驱动”型的虚拟人，综合运用文本、图像、音频等生成技术，打造综合外观、面部表情、发声习惯等产出全面拟人化的数字内容，属于AIGC领域。此种多模态生成技术的聚合应用在虚拟偶像、虚拟主播等领域已有广泛应用。在《量子位虚拟数字人深度产业报告》中，我们将虚拟人按照产业应用划分为两种，即服务型虚拟人及身份型虚拟人。

计算驱动型/AIGC型虚拟人制作流程

1. 设计形象。扫描真人形态及表演、采集驱动数据

利用多方位摄像头，对通用/特定模特进行打点扫描，采集其说话时的唇动、表情、面部肌肉变化细节、姿态等数据。

2. 形象建模，进行绑定

设计所需的模型，或基于特定真人进行高还原度建模，进行关键点绑定。关键点绑定的数量及位置影响最终效果。

3. 训练各类驱动模型

决定最终效果的核心步骤 利用深度学习，学习模特语音、唇形、表情参数间的潜在映射关系，形成各自的驱动模型与驱动方式。充足的驱动关键点配合以精度较高的驱动模型，能够高还原度的复原人脸骨骼和肌肉的细微变化，得到逼真的表情驱动模型。

4. 内容制作

基于输入的语音（或由输入文本转化的语音），预测唇动、表情等参数 核心的技术流程是基于输入的语音，或首先基于TTS技术（Text-to-speech，语音合成技术），将输入的文本转化为语音。基于语音，结合第3步得到的驱动模型，并利用生成对抗模型GAN选出最符合现实的图片，推理得到每帧数字人的图片。通过时间戳，将语音和每帧的数字人图片进行结合。

5. 进行渲染，生成最终内容

直播时进行实时渲染。为保证在特定场景下能够实现实时低延迟渲染，计算框架的大小、算力供给等技术问题同样会影响到虚拟数字人的最终生成效果

综合来看，我们认为虚拟人生成代表着从文本/音频等低密度模态向图像/视频/实时交互等信息密度更高的模态的转化。其中，视频是短期的发展重点，而长期来看，乃至在元宇宙阶段，通过实时交互成为社交节点，都将是虚拟人重要的应用场景。有关虚拟人的更多产业分析信息，请参考量子位智库此前出品的《虚拟人深度产业报告》。

在AIGC领域，我们将虚拟人生成分为**虚拟人视频生成**和**虚拟人实时互动**。

虚拟人视频生成是目前计算驱动型虚拟人应用最为广泛的领域之一，不同产品间主要的区分因素包括：唇形及动作驱动的自然程度、语音播报自然程度、模型呈现效果（2D/3D、卡通/高保真等）、视频渲染速度等。我们在此关注到了小冰公司与每日财经新闻合作的虚拟人实时直播，除虚拟人的自动生成外，还包括了摘要、图示、表格等的自动生成，在虚拟人的基础上，交付了更为完整的AIGC内容播报产品。此外，倒映有声的TTSA除虚拟人外，还包括整个画面中的素材呈现，相较于市面上嘴形、面部和身体律动的有限覆盖，虚拟人播报的整体效果也有所提升。

代表公司：倒映有声

一家以技术为核心的创新型公司和无人驱动数字分身技术解决方案供应商，通过自研神经渲染引擎和TTSA技术，实现基于文本实时生成高质量语音（音频）和动画（视频）。

在试用了倒映有声的产品后，我们发现其虚拟人自然度高于市面产品，倒映有声将其归结于神经渲染（Neural Rendering）、TTSA（基于文本和语音合成实时生成音频和视频）、ETTS（富情感语音合成）、Digital Twin（数字孪生）。通过神经渲染技术快速构建AI数字分身，通过语音+图像生成技术，生成和驱动数字分身的唇形、表情、动作、肢体姿态，创造表情自然、动作流畅、语音充满情感的高拟真度数字分身IP。

而虚拟人的实时互动则广泛应用于可视化的智能客服，多见于APP、银行大堂等。在AIGC的虚拟人领域，由于**更能够体现AI在个性化、高并发性等方面的优势，我们更强调虚拟人的实时交互功能**。我们可以将这一功能理解为以人为单位的数字孪生，其中会进一步涉及思维及策略相关的生成。但由于文本生成的局限性，该场景目前只能适用于特定行业。

该领域的代表公司包括：**HourOne.ai、Synthesia、Rephrase.ai、小冰公司、倒映有声、数字王国、影谱科技、科大讯飞、相芯科技、追一科技、网易伏羲、火山引擎、百度、搜狗等**。

除了基于NLP进行问答外，以小冰公司和腾讯AI Lab（AI虚拟人艾灵）为代表，部分公司也在尝试将不同的生成能力融合在虚拟人下，使虚拟人能够更好的融入现实世界。

以小冰公司的小冰框架为例，虚拟人不仅在人格化形式上涉及了虚拟面容生成、虚拟语音定制、交互等，并进一步被赋予了写诗、绘画、演唱、音乐创作等AI内容创作能力，以虚拟人为接口，对外提供全栈式的AIGC能力。

虚拟人及综合性AIGC代表公司：小冰公司

小冰是全球领先的人工智能科技公司，旗下小冰框架是全球承载交互量最大的完备人工智能框架之一，在开放域对话、多模态交互、超级自然语音、神经网络渲染及内容生成领域居于全球领先。

作为“AI being”派虚拟人。小冰的产品始终是人+交互+内容。具体包括虚拟人（夏语冰等somebody instance、虚拟男友等nobody instance和国家队人工智能裁判与教练系统观君等在垂直场景中工作的虚拟人类）、音频生成（主攻超级语言及歌声，在线歌曲生成平台与歌手歌声合成软件X studio）、视觉创造（毕业作品集《或然世界》、为国家纺织品开发中心、万事利等数百家机构提供了图案和纹样设计）、文本创造（2017年即推出小冰诗集）、虚拟社交、Game AI（Xiaoice Game Studio）等。

商业客户已覆盖金融、智能车机、零售、体育、纺织、地产、文旅等十多个垂直领域，并提出了以“人力”的逻辑去进行商业报价的虚拟人商业模式。

1.4 相关预测

我们认为，不同赛道下AIGC应用落地推广程度主要受到两方面影响，特定技术的水平状况以及在实际应用中出现的转化门槛。

特定技术的水平状况

- 操作方式与传统创作流程的契合，需要符合创作者使用思路并提供相关人工的介入环节。与底层模型架构相关
- 是否支持小样本学习，一方面是训练效率的问题，一方面也是对实际使用门槛的回应
- 生成能力的鲁棒性，确保在生成规模扩大、高并发性等实际应用需求下依然能够有稳定的产出
- 生成模型可外延和泛化能力，例如可同时涉及多重设计/创作风格
- 配套数据及现有标注情况，对于实际场景落地有重要意义
- 模态对齐工具/架构具有重要意义，CLIP的技术程度，以及潜在模态对齐模型具有重大风向标意义
- 不同元素间的连续控制情况，例如文本生成中的上下文语义承接、视频生成中不同帧的连贯性、图像生成不同像素间的细节处理等
- 具体的技术选型问题，需要相关的底层原理嵌入。但目前而言，整体的技术路线已经相对定型，需要更多在技术细节上的优化

在实际应用中出现的转化门槛

- 完成渲染、构建等相关配套环节的时间/技能/算力投入及难度，如是否搭配有相关后续平台
- 论文上指标和现实使用中的指标无法对齐
- 大量隐含信息的问题，难以直接回应现实中的应用问题

我们认为，以下技术要素值得关注：**长文本生成、开放式文本生成、NeRF模型、Diffusion模型、跨模态大型预训练模型（支持的模态数据类型、模态对齐架构设计、支持的下游应用）、小样本学习及自监督算法、强化学习及环境学习。**

以下技术场景是我们认为是短期内将有较明显爆发的：**闲聊式文本生成、个性化营销文本、富情感及细节TTS、拼凑式视频生成、基于文本的AI绘画、语音复刻。**



产业篇

2.1 价值分析

在回答AIGC的价值之前，我们需要对AIGC本身的独特性做出回答。在量子位智库看来，用AI进行内容创作的价值主要来源于五点。区别于市场观点，我们认为最后一点，也即与AI系统的个性化、实时化互动最能体现其潜在价值。尽管目前AIGC尚无法完成精准可控的生成，但我们相信这一赛道未来的技术与市场规模上限。

以下为五点主要价值，重要性逐次递增：

- **降低内容创作门槛，增加UGC用户群体**

AIGC能够代替人工完成声音录制、图像渲染等工作，使更多人员能够参与到高价值的内容创作流程中。预计这一效果在2B结构化内容生成的领域非常明显，个别场景会出现2C服务。跨模态生成成为未来重点。

- **提升创作及反馈效率，铺垫线上实时互动**

目前来看，效率提升主要体现在提升专业生产人员的生产效率。目前，用户对于能够动态交互的个性化数字内容的需求越来越高，传统的开发方式无法满足日益上升的需求，消费速度远高于制作速度，需要AIGC填补供需间的差距。

但我们认为，更为关键的是，AI同样提升了内容的反馈生成速度，对于实时交互内容有重大意义，具有将线下和真人的快速交互迁移到线上的可能，也即令AI承担真人的社交、创作、协作功能，可能会出现新的潜在场景（如社交类和探索类游戏等）。我们在Game AI板块所提及的AI bot实时玩家教学在一定程度上提供了一种互动的形式，交互教育/交互探索游戏/虚拟陪伴场景值得期待。

目前来看，内容消费者变得更容易将现实情感需求投射在虚拟世界中，预计会产生许多深入实时的互动需求，市场规模可观。

- **基于海量数据得到强创造性和开放性，有助于激发创意认知、提升内容生产多样性**

相较于人类艺术家，AI能够接触借鉴更多的数据，在基于prompt进行内容生成后，AI创作的内容会有更多的二次创造空间和自由度。例如，生成算法能基于特定条件或完全随机的生成现实中不存在的形状、色彩搭配、图案或结构等，赋予内容创作更多可能，产生“超现实感”及“未来感”，推动艺术创新。

- **对不同模态元素进行二次拆解组合，改变内容生产逻辑及形式**

通过语音克隆、编曲风格提取等手段，AIGC能够将原客体所对应的不同模态信息进行拆解，例如演讲者的面部形象、声音、演讲内容等。在重新组合之后，能够完成过往受到条件限制无法完成的工作。例如路人的声音+专业的播音逻辑、更符合特定审美的面部等，打破真人/真实场景在要素组合上具有的局限性。

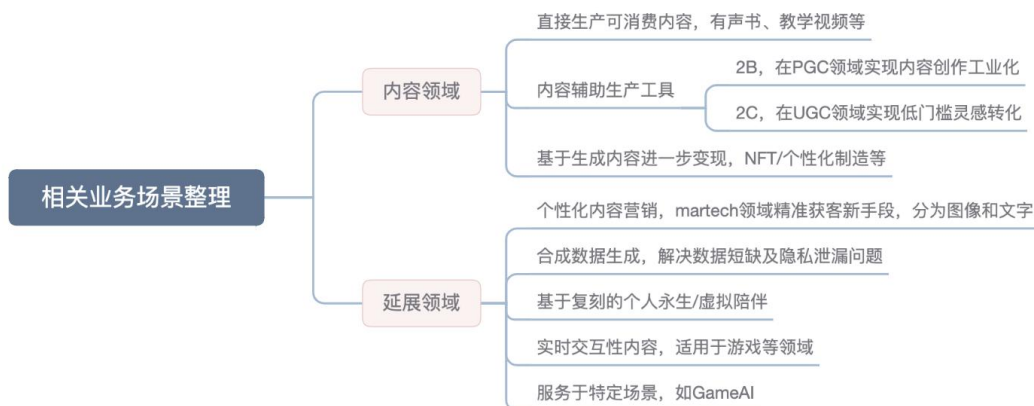
- **和其他AI系统或数据库进行联动，有实现高度个性化/高频优化**

在与特定的数据库（例如实时更新的客户数据、市场反馈数据、特定主题下的历史统计数据）或AI系统进行联动后（如个性化推荐系统等），AIGC能够在更为精准的未来预测/个性化预测基础上调整其生成内容。

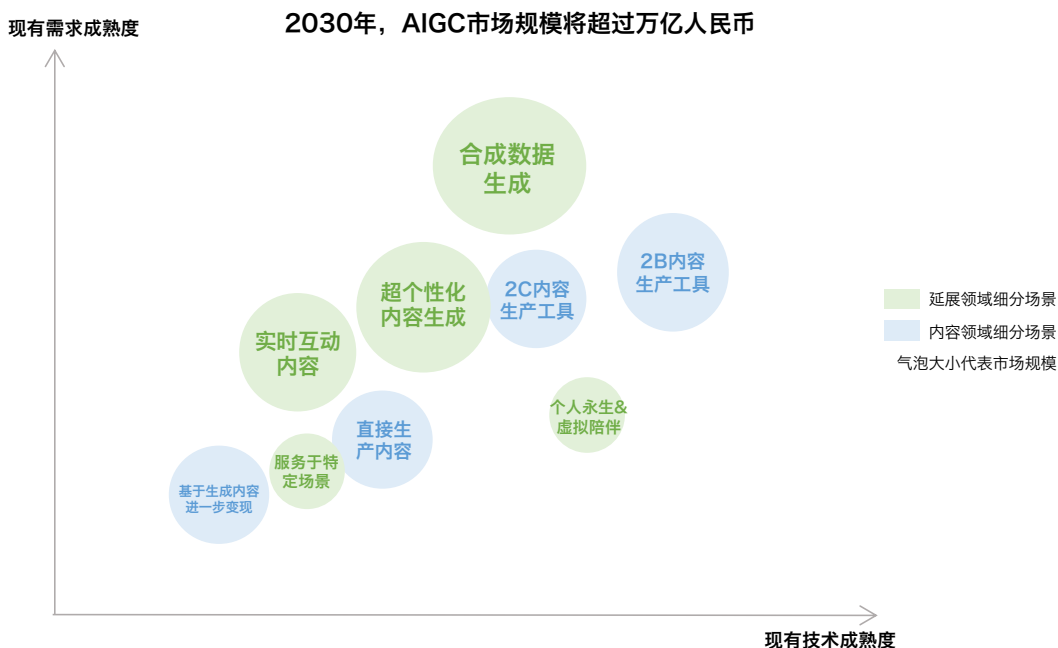
例如，根据用户习惯调整内容营销文本、根据所处渠道风格调整生成内容、参考历史数据优化生成内容等。该价值在内容营销领域有重大意义，该场景目前在国外已相对成熟，我们将在细分赛道分析中进行进一步展开。

2.2 细分赛道分析

总体而言，我们将AIGC产业划分为**内容领域**和**延展领域**。内容领域是目前已经能够、但并未有效实现商业化的领域，预估未来1-2年之内会有相对完整的产业生态发展。而延展领域，由于对AI生成的自由度和稳定性有更高的技术要求，在国内的应用落地相对内容领域会更晚。分析师认为，**由于能够更加充分体现AIGC系统联动及要素拆解两大关键优势，延展领域后期将成为AIGC创造价值的主阵地，提供AIGC的想象空间天花板。**



由于AIGC目前产业化程度有限，大量业务场景尚未成功变现，我们目前尚无法对市场规模进行精准测算。在参考相关创作软件、已有成功商业化案例、AIGC预估功能等因素的基础上，我们对各细分场景的商业化潜力进行了如下估算。



内容领域

• 总体评价

该领域大部分是在和人力生成内容进行竞争。其中的存量价值来源于同类内容的降本增效，而增量价值则来源于跨模态的内容生成以及AI本身带来的科技感。

比较通用的路径为用户上传关键元素（文本的关键、图像的提示词、歌曲主旋律等），选择风格、应用场景，由AI智能完成内容加工，生成对应的内容，或者短时间内生成大量可供选择的素材，提高内容生产效率。在这个板块，由于内容供给的缺口尚不明显，我们认为**对内容渠道的把控将成为核心竞争力。发行商、内容最终消费渠道具有强的产业链话语权。**

此外，由于内容供给机制与内容平台发展紧密挂钩，**该领域的大厂具有后期跨界竞争/通过并购完成版图扩张的动力与实力**，而第三方创企公司除了完成好2B2C服务外，**需要自行关注新的内容消费场景及渠道**。

• 发展利好条件：内容行业发展的内在需求

1. 内容消费量增加，急需降低生产门槛，提升生产效率

随着内容消费的碎片化，对内容产出的多样性有了更高要求，例如不同的宣传物料，在不同的渠道/位置上都有特定的设计要求，需要区别生产，增加了内容总体的需求量。

保证内容的持续供给，这一需求进入元宇宙会更加明显，需要海量内容来填补数字世界内容的供给缺口，需要增加更多具有创作技能的人，如软件工程师、画手等。这可以通过提升原有人群的创作效率，以及赋能非专业人士完成。

2. 内容消费升维，消费者要求更为丰富的感官体验

随着相关内容消费硬件的提升，消费者对内容形态要求也会更高。从2G时代的文字，3G时代的图片，4G时代的视频，以及5G时代将充分增长的直播，通信及硬件的持续增长使得主流的内容消费形态不断变化，消费内容不断升维。无论是更高质量的视频或CG内容，还是预估将成为下一代主流的VR/AR内容，提供更加海量和丰富的感官体验已经成为一种必然。

3. 内容生成个性化和开放化，User端表达意愿有明显上升

内容消费可能从内容本身，转移到内容的参与者身上，参与者希望能够对内容本身施加更多的影响。例如具有一定不可预测性，以及和个人绑定的特殊情节/特征。在这种思路下，二次内容创作的占比将有所提升，内容在抵达每个个体时都需要通过工具进行二次改造和发散。

4. 元宇宙要求内容有新的自动生成机制

此外，AIGC也和2022年的热词元宇宙紧密相关。据rct AI评价，元宇宙需要永续性、自增长和大规模。特别是自增长，在没有人力不断保证内容增长的时候，他需要有自己驱动发生的底层机制。要维持他不断的成长，就一定需要一个背后的核心运行机制。过往的游戏策划等人力岗位难以满足这一需求，AI成为新的底层动力。

• 细分场景

1. 直接生产可消费内容

该场景的核心问题是除了AI自带的科技属性外，在本质上难以产生和人工生产内容的价值差异点。分析师认为，尽管短期内更易看到进展，但该部分不是AIGC主要的价值增长部分。在整个内容直接消费市场中，需要在内容生产和消耗速度之间的寻找平衡，而不是对大量生产消费内容产生过度迷信。这对该场景提出了两方面要求：

1. 对具体场景的供需状况进行谨慎分析，AI对供给速度的提升是否具有实际意义
2. 不能单纯的从内容供给总量来看，还是需要寻找个性化、实时化等其他角度，进入延展领域

据业内人士评价，其实国内各类内容的供给都相对有限。AI生成内容本身的价值相对有限，重点应当在概念包装和开发新的商品形式，例如结合NFT、盲盒等，也即我们所提及的第三种场景。

2. 作为内容辅助生产工具

该细分场景和前者其实是工具及产品之间的关系，基本是标杆案例和最终标准化赋能的区别。如果按照AI的参与程度进行区分，该赛道包括完全借助AI端到端的生成模式/由AI提供相关创作线索、补充其他创作元件（如伴奏等）/由AI完成特定操作性工作（局部特效生成、低分辨率转高分辨率等）等。大类可上分为AI端到端生成，和部分借助AI技术的人机共创模式。

目前，市面上存在部分不具备真正AIGC能力的相关产品，例如提供相关素材参考、基本模版、小范围智能改写等，我们肯定该类产品在提升生产力上的能力，但仍需关注其包装成分。该场景下有两点需要注意：

a. 该场景自身的工具属性

作为工具一定要回到整个细分内容行业的整体产业链上去，进行上下游有效联通。同时，作为专业工具类产品，在UI/UX/核心使用效果能力上需要与现有创作工具，如Unreal Engine进行看齐。

b. 平台和渠道的自身优势

作为辅助生产工具，该赛道将与相关应用，特别是内容分发及创作平台形成强绑定。互联网大厂具有更强的业务承接能力，且本身具有一定垄断性。初创公司及产品需要注意寻求能够相对独立出来的业务点或特有的功能。

在前沿研究领域，目前顶尖的AI研究机构主要由原来的互联网大厂（自设部门或收购）把持，其研究领域往往涵盖AI的全细分技术和应用范畴，推出的服务也主要应用于大厂本身的各项业务，如腾讯开发的AI Bot应用于腾讯游戏中，阿里的智能语音服务主要应用于微信，字节跳动则主要基于短视频场景研究AI赋能。

在内容辅助生产工具赛道下，我们按照目标群体，进一步将其细分为2B和2C两大方向。

2B，在PGC领域实现内容创作工业化，提高PGC内容生产的活跃度

内容的可持续性需要平台帮助降低创意门槛，从而加快专业生产内容的生产效率。AIGC有助于克服人力的不足，快速覆盖之前的长尾场景，具体包括端到端生成和提供中间线索两种方式。在综合调研市场需求后，我们认为后者将占据绝对主导地位，能否有效平衡创作者的个人意愿和AI创造的自由度值得关注，这需要从业公司在底层技术设计架构、产品流程设计上进行调整。

我们将这种场景定义为“存量市场价值上的二次开发”。尽管并非AIGC未来主要的价值增长点，但为第三方创作机构或内容创作平台提供2b服务是目前相对最为清晰的落地场景。

2C，在UGC领域中实现低门槛灵感转化，扩大UGC可覆盖的领域和潜在人群

AIGC能够激发C端用户灵感，并提供建设性的执行工具，同时降低其对专业知识的门槛。以编曲为例，人为创作大概需要7-10年的经验积累，且不同的风格需要艺术家进行大量的学习和迁移。

降低用户内容创作的难度对提升用户活跃度，乃至完善平台内容循环有关键性意义。要构建良好的UGC生态，其关键在于开发多数用户平庸的创作力，以及有效挖掘并分发用户创作出的优质内容。

3. 直接基于生成内容进行进一步变现

该细分场景目前大致分为衍生周边（如NFT等）和整体产业链设计环节（柔性制造中的个性化设计环节）两条路线。

目前，NFT已经成为国外相当重视的AI内容变现渠道，例如Rosebud.ai本身支持NFT从文字得到特定创作，Obvious和耐克合作得到3个NFT。但由于国内整体环境的影响，这一领域的发展尚不明确。

而在个性化制造上，消费者更多的参与到物质生成过程中，生产的门槛由柔性制造以及低成本渲染等周边工具完成，设计向的部分由AIGC完成。例如小冰与中国纺织信息中心、国家纺织产品开发中心共同推出了人工智能纺织服装图案设计平台。为企业按需定制100%原创的专属图案纹样，实现大规模、多样化、质量稳定的按需设计。

• 产业影响分析

总体而言，AIGC对于内容消费领域的整体影响相对有限，一个是AIGC目前的供给比重依旧相对较少，一个是并没有能够在场景本质中提的独特价值。特别是在商业美术领域，现有的技术水准尚无法达到所需层次。

最明显的是对内容行业原有人才结构的影响。由于前期渲染、整体调色等基本技能将交由AI完成，设计师的工作重点将会向提出prompt转移，也即如何将艺术设想清晰的描述给AI、提升自身的创意、审美、storytelling及搜索能力。在风格覆盖上，AI的效率和广泛风格将能够帮助艺术家们弥补这一缺陷。

我们进一步大胆猜测，在这个内容行业中，**最易受到AIGC影响的应当是内容创作平台。**

一方面，内容创作平台会通过提升专业人员的生产力，提升其在内容链上的供需撮合能力及效率。

而更为重要的一方面，AIGC将作为低门槛的创作工具赋能给目前因能力/时间等原因尚未接触内容创作的人群，提供“灵感—线索—作品”的快速转化路径，为内容创作平台提供更大的UGC内容供给量、用户活跃度与复购率，甚至是赋予内容创作平台新的品牌作用及定位，完成内容创作相关教育。

而长期来看，随着AIGC内容被广泛采用，版权/图库领域也将受到明显冲击。简单配图/配乐可以通过在AIGC软件中依据需求自动生成，创作者对海量图库的需求将有一定下降。

延展领域

在这一领域，量子位智库将其本质归纳为“依托人工智能，个性化数字内容的自动化构建”。

对于所有此类场景，需要把核心调整回商业本质上面去，关注和其配套数据或者底层原理是否清晰。数据问题，不仅仅是纯内容领域对素材数量的需求，也包括对于业务上下游联通数据的需求。此处的底层原理则指，除去相关的技术原理（特别是AI生成内容的自由度和稳定性上），还需要关注相关商业原理，例如个性化对于当前业务场景的重要性。

在结合量子位智库对AIGC延展价值，并调研了国内外AIGC相关企业的基础上，我们认为以下五类场景具有明显的增长潜力，顺序价值递减：

1. 个性化内容营销——最大的代表性场景，martech领域实现精准获客的新领域

随着广告投放成本持续提升、消费社交化、消费碎片化等趋势，能够使消费者体验更好的内容营销正变得更加重要。而在内容营销中，如何以现有内容最大程度引发消费者的兴趣成为重点。

在过往，由于效率限制，内容营销往往采用相对标准化的方式，由相关部门或第三方公司在洞察消费者整体画像的基础上制作有限的内容营销模板，进行分发，重点在于分发的精准个性化匹配。但发展到今天，这种方式一方面会遗漏掉部分长尾客户的需求，而另一方面，随着内容营销的普遍应用，消费者对内容的个性化有了更高的要求，如何实现更好更精准更丰富的内容营销，正在成为众多内容公司的关注方向。

在内容营销领域，我们着重关注个性化内容营销。此外，结合SEO（搜索引擎优化，在了解搜索引擎排序原理的基础上，由AI对网站进行站内及站外优化，从而提高当前网站在搜索引擎上的排名）的实时内容营销同样具有一定价值。

基于AIGC的个性化内容营销，主要是指由AI生成系统与底层的客户数据系统/营销效果反馈系统进行数据联通，实时根据相关数据调整生成需求，由AI快速迭代相关内容生成，进而提升个性化营销的效率和精准性。

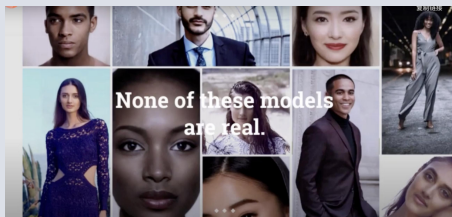
AIGC在这里的关键业务价值在于以下两点：营销内容的快速迭代更新、更为个性化的内容物料生成。优化效率与个性化相叠加，对于快消行业等对内容更迭要求高的领域有明显价值提升。

量子位智库认为，对此类业务而言，关键在于AIGC公司能够延展其服务范围，覆盖营销效果的反馈数据甚至渠道，否则AIGC公司仅承担内容生成工具，其关键竞争力在于内容的生成技术能力，护城河相对较浅。而如果能够进一步延展，公司的业务核心将变为“将营销相关洞察以有吸引力的载体呈现出来”，在交付方案和自身的议价权上都将有明显提升。就目前对国外公司的观察来看，大部分公司已经选择了后者。

如果按照具体个性化的内容进行区分，该细分赛道可以分为图像领域和文本领域，文本市场目前更为成熟。但两者均尚未在国内市场出现。

图像领域，与快速迭代紧密挂钩

代表公司为Rosebud.ai和Generated Photos。



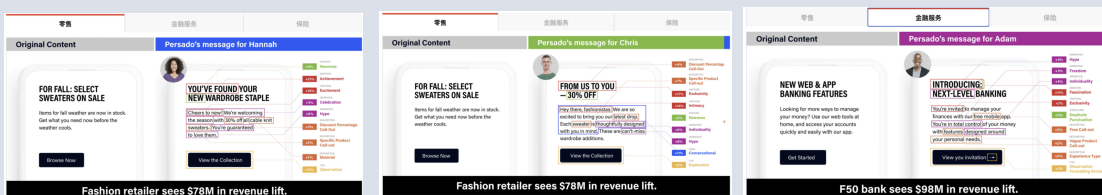
Rosebud.ai为不存在的人建模图像，将模型衣服穿到虚拟真实模型上，支持不同的模特面孔，可以根据详细的受众人口统计数据制作无限变化的模型并使用各种视觉效果来定位您的客户。该公司声称，利用他们的人工智能生成模型的第一个活动显示点击率增加了22%。

Generated Photos业务类似，有助于帮助公司运行多个版本的营销活动并快速迭代。目前已支持API接入ppt、模型及产品。

文本领域，国外布局广泛，商业化价值已得到极为明确的验证

针对不同的用户画像，产出不同的对应文本。这种以营销为目的的文本生成在国外更为常见，国外是更多以商业化为目标。

最为经典的是Persado，该公司的平台将营销创意分解为六个关键要素：叙事、情感、描述、号召性用语、格式和文字定位。该平台使用这些元素的多重组合来提出与每个客户交谈的最佳信息。许多公司利用 Persado 的平台通过促销电子邮件、社交媒体广告、店内展示横幅甚至广播内容来提高品牌参与度和转化率。



其他相关公司包括：Acrolinx、Atomic Reach、Automated Insights、AX Semantics、Anyword、Brightedge、BuzzSumo、Concured、Curata、Conversion.ai、Frase.io、HubSpot、Keywee、MarketMuse、Narrative Science、Publishthis、Pencil、Phrasee、Persado、Qordoba、Scoop.it、Showpad、Stackla、Uberflip、Vennli、Yello、Yseops

除去根据模态划分之外，实际可动态进行个性化的内容非常多，例如：博客文章、案例研究、资源、页面部分、标题、图像、数据、正文、CTA。可以从营销的全流程，包括沉睡唤醒、短信、推送、广告词等都用这种方式生成。但这会导致整个网站的背后内容架构都高度个性化，对于网站本身、第三方服务机构和算力来说都是一种巨大消耗。

我们认为，在内容营销领域，AIGC并不应当作为单独的服务对外提供，还是隶属于“创意供给—内容生产—内容管理（素材库与数据库）—内容分发—数据方案”中的内容生产部分。由于品牌主最终是以整体的广告营销效果为标准，要使AIGC最终能够得到理想效果，基础素材、营销策略设定、技术生成、评估优化、数据回流等缺一不可，因此前后的数据及流程打通有重要意义，对AIGC公司的生态建设能力提出了较高要求。

AIGC的引入可以使广告创意从推荐分发个性化进入到生成个性化、更加千人千面，可能会产生新的主流服务模式和计价模式。但目前由于国内公司尚无明确的价值证明，只有当对个性化的要求已经超出了模版积累可以满足的程度（效率、细粒度要求），我们认为AIGC成为内容营销的主力仍需要2-3年时间。

需要强调的是，对于个性化内容营销而言，目前尚不需要过高的生成技术能力。我们将在行业发展阶段中对该部分进行详细展开。

2. 合成数据生成

通过AI技术（主要为对抗生成网络GAN）从现有内容，如图片、音频、文本等，生成完全虚构但保持原始数据特定属性的全新数据，本质上是希望通过生成技术克服AI模型训练的“原始数据短缺问题/实现不同机构间的数据转移及更新以避免隐私泄露/完成特定算法的训练及数据分析，且能够避免手工标注的高昂成本。

在**医疗和自动驾驶领域**（Waymo、Cruise、Aurora、Zoox）应用广泛。目前已经有了测试标准APAT，技术生态较完善。

相关公司包括Anyverse、Autonoma、Bifrost、CVEDIA、Cognata、Coohom Cloud、Datagen、Deep Vision Data、EdgeCase、Lexset、Mindtech、Neurolabs、Neuromation、reinvent systems、rendered AI、Scale Synthetic、Simerse、Sky Engine、Simerse、Sky Engine、Synthesis ai、syntheticaic、Synthetic Data pty、syntetik、Usearch、Vypno、Betterdata、Datomize、Diveplane: 、Facteus :、Generatrix、Gretel、Hazy、Instill AI、Kymera Labs、Mirry.ai、Mostly AI、Oscillate.ai、Replica Analytics、Sogeti、Statice、Syndata AB、Synthesized、Syntheticus、Syntho、Tonic、Ydata、Rendered.ai、Bifrost 和 Mirage。

合成数据生成可能会导致**专有数据的竞争壁垒将被逐渐削弱**，数据采用标准从**真实数据到可信**。

3. 基于复刻的个人永生/虚拟陪伴

背后有可能用作**虚拟偶像的运营逻辑**，以及用于保持一定的品牌调性，如根据用户审美和既有设计成果进行定制化训练，调整对应的设计风格。

4. 交互性内容。适用于游戏等实时变化较大、要求内容多样性的领域

已有尝试包括Melodrive电子游戏背景音乐**实时生成**、宾利宣布与 LifeScore 合作以实时生成配合引擎的音乐、实时闲聊机器人、开放式探索游戏等。

5. 以特定内容作为部分服务

重点不在于内容本身，而是服务于游戏画面设计、动漫设计等特定内容消费场景下，例如GameAI中的数字资产生成。为了避免其中出现一些不恰当的功能性内容，**需要其他的特定数据集或者基准**等。此外，也可结合精神心理治疗等相关因素，例如Brain.fm的专注音乐和聆心智能。

以Zynro为例，**该公司围绕其电商服务能力构建AIGC能力的矩阵**，同时包含了文本、图像等不同的生成能力，提供了打包的解决方案。

2.3 行业发展阶段分析

现状分析

从行业重点、关键场景、玩家分布情况等情况来看，**我国目前尚未进入真正的AIGC阶段**，在细分赛道上出现了个别优秀的先行公司及研究机构，**但目前行业尚未进入大规模验证和体系化发展的阶段**。

场景上来看，目前最关键的是**内容领域需求较为饱和**，且**降本增效尚不够迫切**。细分来看，**虚拟人行业**将成为产业中短期增长的关键动力。玩家需要寻找**容错率较高、专业度或创意要求有限的场景**。但长期而言，AIGC依旧需要有**需求更加清晰且明确的消费落地场景**（如**实时互动、实时反馈类场景**）和**高度个性化生成场景**。

玩家方面，国内目前的初创玩家数量较少，大部分细分赛道的初创玩家在5家以下，大部分为具有相关业务大厂/研究机构的内部探索项目。因为市场规模短期内相对有限，且渠道和前期资源的积累需要较长时间，我们认为后期会更容易出现高行业集中度。

技术方面，除去个别布局较早的企业外，国内AIGC领域大规模稳定生产的能力还需进一步增强，完成从**尝鲜式做法向正式和业务化的进入各类场景的转化**。据业内人士估计，**预计2-3年内，各AIGC能力有望在成本可控的前提下，实现工业化生产**。

变现方面，尽管国内的Game AI、虚拟人、TTS、结构化文本写作等领域已形成稳定的业务合作关系及收入关系，但行业整体的变现能力仍待进一步证明。且基本为项目制/长期项目制，标准化/平台化产品欠缺发育。

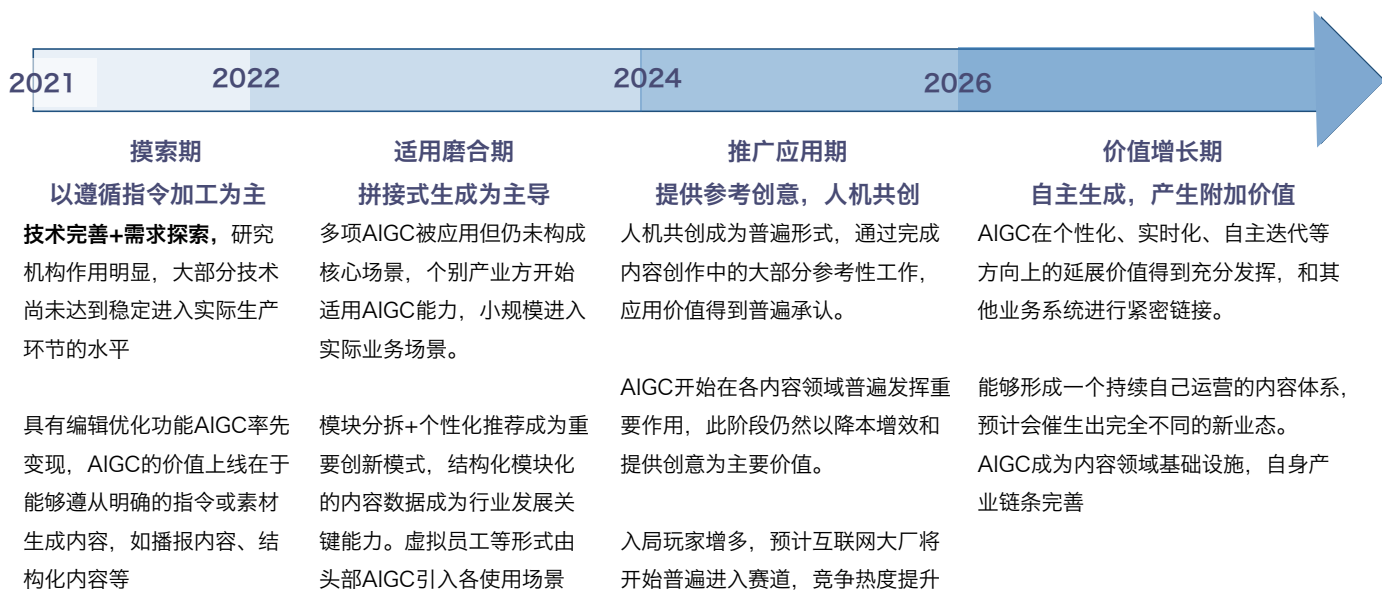
行业发展阶段划分

量子位智库认为，我国AIGC行业目前是尝试阶段，重点在于技术完善+需求探索，研究机构的作用明显。大部分情况下AIGC尚未进入业务场景，变现能力有待增长。具有编辑优化功能的AIGC将成为这段时期的发展重点。在主要场景中，AI能够遵从明确的指令或素材，完成有助于最终生成的工作，例如虚拟人播报、语音克隆等。

我们将行业发展阶段进行了如下划分。其中，关键节点包括：

- **AIGC技术能力的成熟。**体现在特定模型在大规模测试后指标能够稳定在理想状态，可以以论文发表及Github反馈等作为关键指标。但需要强调，该阶段的技术能力成熟往往举例实际应用仍有一段距离，不能直接作为判定标准。
- **产品形态的成熟。**不仅包括低代码/零代码等低门槛要求，还需要符合创作者的使用习惯，例如，能够中途及时介入修改、多重形式的prompt、有充足的介入接口等。分析师认为后者更为重要，除去产品设计之外，也与本身的技术架构相关。
- **核心场景的确定。**目前，尽管许多AIGC能力已经进入实际使用，但尚未寻找到能充分体现其核心价值的核心场景，应用价值无法得到充分发挥。
- **产业方的接纳态度。**包括底层接口开放、业务流程融通、产业基础设施完善等。可以以行业头部公司的接纳/自研态度作为重要参考指标。

需要强调的是，由于各内容领域的生成技术之间发展差距较大，该阶段评价仅基于AIGC行业整体。相关指标也可用于衡量不同AIGC赛道目前的成熟度，用于预测其未来发展路径。



在参考国外发展情况，并深入了解国内内容领域生成需求的基础上，量子位智库认为，国内即将进入注重“模块分拆+个性化推荐”的阶段。尽管尚未出现垂直的初创公司，但我们已在国内看到了相关尝试，且相信这一场景未来巨大的应用潜力。我们将对该“模块化+个性化推荐”泛AIGC思路进行具体展开。

有关“内容模块化拆解+个性化组合推荐”

分析师认为，**素材模块分拆+个性化推荐**的组合（如特赞、筷子科技、阿里鹿班）、乃至**于含有个性化推荐能力的内容编辑工具**（如爱设计，支持尺寸延展、调整模块布局等）目前更具有有落地可行性。

在实际应用效果上，尽管技术路径不完全相同、且在技术相关需求上有较大差距，但在实际应用效果上，我们认为目前这种实现方式相对于此前标准化内容/人力生产内容的效果仍然有较大突破，是一种平衡较好的方式。**对于本身已有内容丰富，且能够进行细粒度拆分的企业而言，足以支撑其背后的想象空间。**

典型路径是“**提供关键模版，对已有要素进行AI元素分拆、标注分析、元素个性化重组以得到标签化的内容解构结果，并联合智能创意决策及迭代——结合用户的反馈信息与市场营销效果进行迭代**”。

特赞的内容智能生成工具和阿里鹿班是满足这一思路的典型产品。特赞本身是链接内容供需双方的二级供应商，通过其模块化组合的内容智能生成，降低供给方的创作投入，带动整个平台效率提升。据其官方数据，目前已实现71倍生产效率增长，71.8%的作品可达到设计师水准。

特赞“内容智能生成”（增加设计师服务需求方的效率）



Dynamic Generation
内容智能生成
一键修改海量图片与视频尺寸、智能化套版，提升基础设计工作效率

智能工具箱
敏捷智能修图小工具，无设计背景也可轻松使用

- 智能抠图**
一键去背景、抠图商品和人物，制作白底图
- 0-1 智能生成**
主要视觉元素 + 文案 + 尺寸，一键AI智能设计物料
- 切图工具**
根据不同渠道的详情页规范，一键长图切成多个短图，方便上架

阿里鹿班（显著提高商家的内容运营效率）



快速入口 电商场景应用功能，一键批量生成

- 精品主图
- 批量白底图
- 海报拼图
- 海报切图

其他功能

- 自定义场景
- 淘宝旺铺海报
- 钻视图
- 主图中衣服检测
- 1688旺铺海报
- 直通车

可控的视觉内容生成系统，使其在短时间内完成大量banner图、海报图和会场图的设计，提高工作效率。用户只需任意输入想达成的风格、尺寸，鹿班就能代替人工完成素材分析、抠图、配色等耗时耗力的设计项目，实时生成多套符合要求的设计解决方案。主要应用于生成海报，制作商品主图等。

国内尚未出现垂直该部分业务的初创公司。由于该业务场景下的产业生态复杂。我们认为，为满足快速迭代和大量生产的需求，**该AIGC场景下的主要客群将集中在第三方机构，如第三方公关/广告代理公司、咨询公司、设计工作室。旗下SKU较多，或处于快消品等特殊行业的公司也将存在相关需求。**

基于这种核心需求，为了更好提升AIGC的生成结果能力和针对用户需求的个性化能力，**其产业上下游将包含内容撮合平台（如特赞、爱设计）、创意数据源提供方（包含特定高校、设计师等）、相关数据提供方（特定互联网平台、广告主相关数据）、内容创作者生态、第三方技术服务商等。**

从最终应用价值的角度来进行倒推，**预计在未来五年内，模板拆解+个性化组合都将成为主流的模式。**由于已有**数据结构拆解和模块化程度**，不同场景的对应时间不太一样。数据及目标能够被结构化分离、描述并打标签，是该思路下后续发展的关键。

2.4 产业链分析

由于我国的AIGC行业尚未发展成型，我们在此基于自身理解，绘制了产业链分布图。目前，在上游，我国AIGC产业还有众多欠缺，以数据标注为重点体现。



相关研究课题组包括中国科学院高林教授（计算机图像学领域，如3D内容生成、自动填色等）、清华大学黄民烈教授（长文本生成、对话生成）、清华大学计昊哲团队、北京大学彭宇新教授、中科大先研院-跨媒体智能计算联合实验室等。

2.5 玩家分析

由于我国的代表公司较少，我们已经在技术及场景部分列举了各场景下的国外代表公司，并对我国在该领域的代表公司进行了具体展开，以供参考。国外公司的完整信息清单可扫码添加量子位智库小助手进行获取。

经过对比，我们发现中外公司间存在明显差异，整体差距大约在3年左右，底层技术差距是核心原因。

在我国，由于技术发展不足以及投资环境的影响，AIGC大多被作为公司的部分业务、乃至相对边缘化的功能进行研发开发，独立运行的初创公司数量明显少于国外，大部分细分赛道的初创玩家在5家以下。这也间接导致了国内的

AIGC场景开发较少。整体上，更为常见的是研究机构对AIGC能力进行理论研究，实际效果开发较少。同时，TTS等较为成熟的AIGC能力大多被综合性的AI公司打包提供，在应用场景上缺乏明确的指向性。

其次，在分布场景上有阶段性差异。在国内，布局最多的赛道是AI写作和语音合成领域，虚拟人赛道刚刚开始兴起，基本均停留在内容领域。而在国外延展领域得到了更为充分的挖掘，例如个性化文本生成、合成数据等赛道均是重点布局领域。据相关公司披露，这两个赛道均已获得了明确业务价值和收入验证，但在国外目前进展较少。

技术部分导致了这一原因，但分析师认为，更大的原因在于此类业务拓展的综合性要求较高，需要客户方的数字化程度以及对对应行业的充分了解，预计国内外在这两个赛道上存在2-3年差距。

最后，我国的AIGC行业尚未建立起明确的变现方式。以写作机器人、自动配音等场景为例，大部分产品仍处在免费试用的“流量吸引+平台改良”阶段，收费空间相对较小。此外，多家公司提及，由于服务B端客户时话语权较弱，会考虑向2B2C领域、乃至直接2C领域延展。但作为工具，能否在互联网流量相对稳定的前提下有效接触C端用户，设计好产品转化路径，依旧是一项挑战。

国外AIGC现有变现方式

- 作为底层平台接入其他产品对外开放，按照数据请求量和实际计算量计算

GPT-3对外提供API借口，采用的四种模型分别采用不同的按量收费方式

- 按产出内容容量收费

包括DALL·E、Deep Dream Generator等AI图像生成平台大多按照图像张数收费

- 直接对外提供软件

例如个性化营销文本写作工具AX Semantics则以约1900人民币/月的价格对外出售，并以约4800欧元/月的价格提供支持定制的电子商务版本。大部分C端AIGC工具则以约80人民币/月的价格对外出售

- 模型训练费用，适用于NPC训练等个性化定制需求较强的领域

- 根据具体属性收费

例如版权授予（支持短期使用权、长期使用权、排他性使用权和所有权多种合作模式，拥有设计图案的版权）、是否支持商业用途（个人用途、企业使用、品牌使用等）、透明框架和分辨率等

从竞争方面来看，与具有渠道优势的大型平台竞争成为一项关键点。因为市场规模短期内相对有限，且渠道和前期资源的影响比较重，我们认为后期更容易出现高行业集中度。

我们认为，未来业务关联的大公司收购可能会成为主流现象（例如Loom.ai被Roblox收购、Body Labs被亚马逊收购、深极智能被字节跳动收购），或应当存在较明显的大厂扩展业务趋势。但大厂的业务扩展动机往往在于通过新卖点快速获取流量，优化核心业务，并不会过度关注AIGC本身业务价值的充分挖掘。因此，在明确的新场景出来之前，我们认为这个行业更容易分散在不同的内容消费场景下。

以下是我们所分析的行业门槛及核心竞争力：

1. 无论是内容还是延展领域，在产品上最终需要回到一体化解决方案服务能力

如果以最终生成内容为目标，AIGC本质上提供的还是内容的生成工具，在逻辑上和之前的内容辅助编辑逻辑是相同的。

采集、生产、媒资管理、分发消费等视频整个生命周期，一般都需要覆盖内容生成的全生命周期

而在特定的业务场景中，正如我们前面所说，也需要关注相关的数据联通等环节

2. 回避大厂商后期的竞争压力

作为小公司，需要先专注打磨产品的稳定性，借助快速的试错切进细分赛道。利用其高度灵活性，积极探索产品和市场可行性，新的垂直赛道和应用场景，特别是现在场景还不够明确，大公司基本上都是用这个服务于自己原来的细分场景下。再次重申，我们强调实时性和个性化场景。

3. 与行业的深度绑定关系

和行业的深度绑定关系，和相关平台或底层系统的磨合，能够有相关信任进入接口。大家一定是在某一个场景下去用这个东西的，所以始终是整体内容平台下的特定工具和原来的内容载体需要建立良好的合作关系，除去说明场景可行性外，还需要强调对方在基础架构上的配合意愿。

4. 构建业务闭环

创作型的AI工具如何得到反馈的手段，需要新的模式形成闭环。需要从“拼接式”（需要大量的人工标注数据，只能针对具体任务，不会自我成长）到“进化式”（创造特定条件和核心能力，使之能够完成通用任务并自我成长），如何与用户增加对话轮次、建立情感链接，成为当下“虚拟人”、乃至整个人工智能领域关心的话题。

这其中会涉及特别的衡量指标，所需的样本量和数据量、一些特别的衡量指标，比如开放式对话中的话题引导成功率、上下文一致性等、语音生成领域的语音舒适度

充足的数据（例如IP数据、交流数据、运行数据等）以及交叉人才同样值得关注。