

AIGC 行研

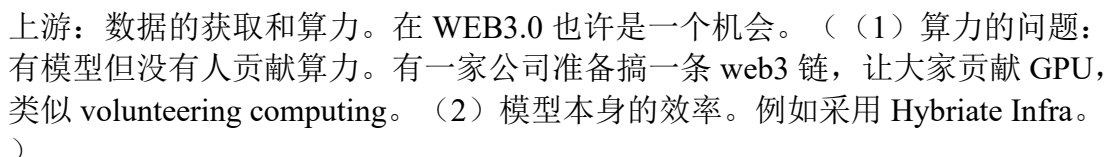
人类擅长分析事物，而机器在这方面甚至做得就更好了。机器可以分析一组数据，并在其中找到许多用例（use case）的模式，无论是欺诈还是垃圾邮件检测，预测你的发货时间或预测该给你看哪个 TikTok 视频，它们在这些任务中变得越来越聪明。这被称为“分析型 AI（Analytical AI）”，或传统 AI。

但是人类不仅擅长分析事物，我们也擅长创造。我们写诗，设计产品，制作游戏，编写代码。直到最近，机器还没有机会在创造性工作上与人类竞争——它们被降格为只做分析和机械性的认知工作。但最近，机器开始尝试创造有意义和美丽的东西，这个新类别被称为“生成式 AI（Generative AI）”，这意味着机器正在生成新的东西，而不是分析已经存在的东西。

生成式 AI 正在变得不仅更快、更便宜，而且在某些情况下比人类创造的更好。从社交媒体到游戏，从广告到建筑，从编程到平面设计，从产品设计到法律，从市场营销到销售，每一个原来需要人类创作的行业都等待着被机器重新创造。某些功能可能完全被生成式 AI 取代，而其他功能则更有可能在人与机器之间紧密迭代的创作周期中蓬勃发展。但生成式 AI 应该在广泛的终端市场上解锁更好、更快、更便宜的创作。人们期待的梦想是：生成式 AI 将创造和知识工作的边际成本降至零，产生巨大的劳动生产率 and 经济价值，以及相应的市值。

生成式 AI 可以处理的领域包括了知识工作和创造性工作，而这涉及到数十亿的人工劳动力。生成式 AI 可以使这些人工的效率和创造力至少提高 10%，它们不仅变得更快和更高效，而且比以前更有能力。因此，生成式 AI 有潜力产生数万亿美元的经济价值。

1.1 行业上下游



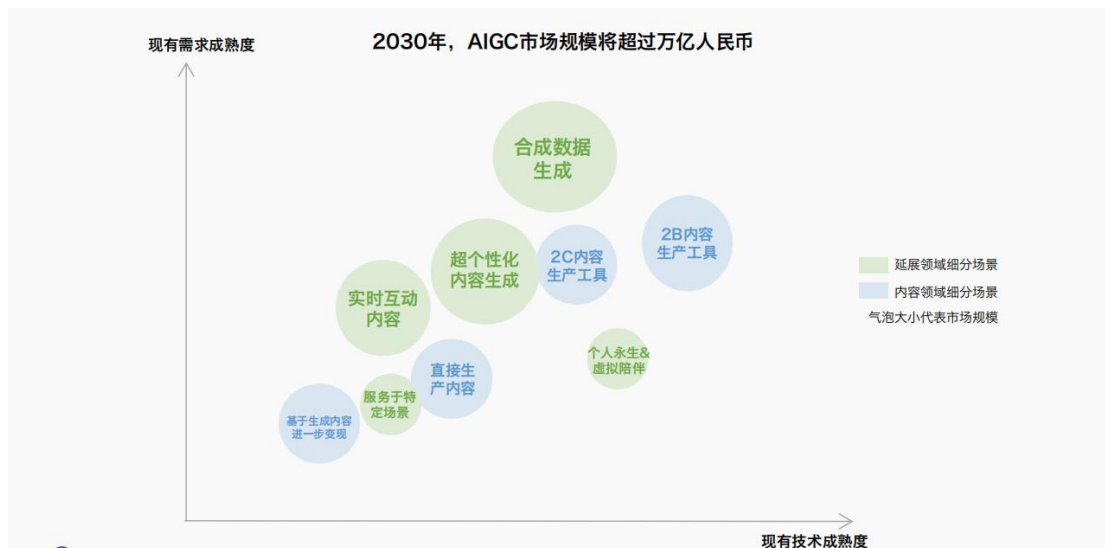
相关研究课题组包括中国科学院高林教授（计算机图像学领域，如 3D 内容生成、自动填色等）、清华大学黄民烈教授（长文本生成、对话生成）、清华大学计昊哲团队、北京大学彭宇新教授、中科大先研院-跨媒体智能计算联合实验室等。

红衫：AIGC 有潜力产生数万亿美元的经济价值

Gartner 预计到 2025 年，生成式人工智能将占有所有生成数据的 10%。



1.3 TAM



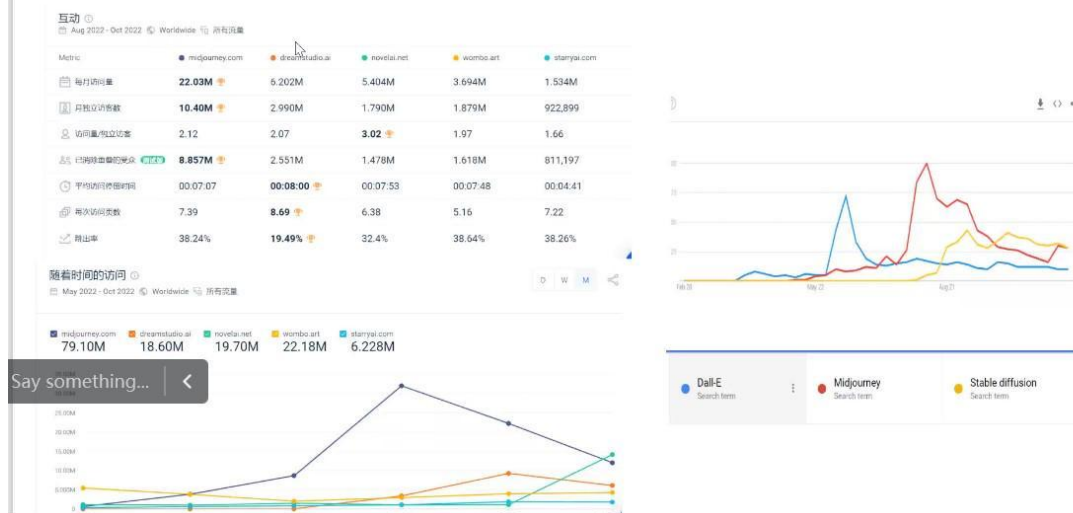
暂未找到，如何算？

1.4 Go to Market

1.To B: （1）卖 API。开源 API，通过别的公司的 API 调用获取报酬。E.g. Jasper 没有自己的 foundation model，通过调用 Open AI 的 API。（2）卖服务：替代部分重复性劳动。E.g. 电影、绘画等可以做出初稿或者替代某些简单机械的工作。（3）卖产品：虚拟人

2.To C: （1）定制化的玩法。例如在 game 中生成千人前面的玩法。（2）生产力提高工具。例如 AI 生成文字写文章等。

图像生成—2C属性更强，引流能力强，但处于商业化早期



存在客户留存问题？

二、Analysis

生成式 AI 正在变得不仅更快、更便宜，而且在某些情况下比人类创造的更好。从社交媒体到游戏，从广告到建筑，从编程到平面设计，从产品设计到法律，从市场营销到销售，每一个原来需要人类创作的行业都等待着被机器重新创造。某些功能可能完全被生成式 AI 取代，而其他功能则更有可能在人与机器之间紧密迭代的创作周期中蓬勃发展。但生成式 AI 应该在广泛的终端市场上解锁更好、更快、更便宜的创作。人们期待的梦想是：生成式 AI 将创造和知识工作的边际成本降至零，产生巨大的劳动生产率和经济价值，以及相应的市值。

三、Force (Why Now)

3.1 技术

生成式 AI 与更广泛的 AI 有着相同的“为什么是现在 (Why now)”的原因：更好的模型，更多的数据，更多的算力。这个类别的变化速度比我们所能捕捉到的要快，但我们有必要在大背景下回顾一下最近的历史。

第 1 波浪潮：小模型 (small models) 占主导地位 (2015 年前)，小模型在理解语言方面被认为是“最先进的”。这些小模型擅长于分析任务，可以用于从交货时间预测到欺诈分类等工作。但是，对于通用生成任务，它们的表达能力不够。

生成人类级别的写作或代码仍然是一个白日梦。

第 2 波浪潮：规模竞赛（2015 年-至今），Google Research 的一篇里程碑式的论文（Attention is All You Need <https://arxiv.org/abs/1706.03762>）描述了一种用于自然语言理解的新的神经网络架构，称为 transformer，它可以生成高质量的语言模型，同时具有更强的并行性，需要的训练时间更少。这些模型是简单的学习者，可以相对容易地针对特定领域进行定制。

果不其然，随着模型越来越大，它们开始可以输出达到人类水平的结果，然后是超人的结果。从 2015 年到 2020 年，用于训练这些模型的计算量增加了 6 个数量级，其结果在书写、语音、图像识别、阅读和语言理解方面超过了人类的表现水平。OpenAI 的 GPT-3 表现尤其突出：该模型的性能比 GPT-2 有了巨大的飞跃，并且从代码生成到笑话编写的任务中都提供了出色的 Twitter demo 来证明。而 Microsoft 基于 GPT-3 改进了模型，得到了 GPT-4，在多线程的连贯性方面更是取得了长足进步，目前已经通过了图灵测试。

尽管所有的基础研究都取得了进展，但这些模型并不普遍。它们庞大且难以运行(需要特别的 GPU 配置)，不能被更多人广泛触达使用(不可用或只进行封闭测试)，而且作为云服务使用成本昂贵。尽管存在这些限制，最早的生成式 AI 应用程序也已经开始进入竞争。

第 3 波浪潮：更好、更快和更便宜（2022+），算力变得更便宜，新技术，如扩散模型（diffusion models），降低了训练和运行所需的成本。研究人员继续开发更好的算法和更大的模型。开发人员的访问权限从封闭测试扩展到开放测试，或者在某些情况下扩展到开源。

核心突破-Diffusion的应用&大网络训练

- 核心思想：
 - 相比用神经网络一次直接生成图像，优化为求解一个新问题：如何基于A图生成一张图片（或特征空间），比A图像更接近目标。
 - LDM引入AutoEncoder结构，在Latent space做diffusion，在保证生成质量前提下，极大的降低了计算代价
- 模型结构：
 - Bert Tokenizer
 - transformer
 - VQGAN (UNet)
 - SR Upsampler

denoising step crossattention switch skip connection concat

	Latent Diffusion Model (LDM)	Stable Diffusion
训练数据集	LAION-400M	LAION2B-en+LAION5B清洗过的子集
文本编码	从头训练的BERT Encoder	预训练的CLIP ViT-L/14 Encoder
训练分辨率	256x256	256x256上预训练，512x512上微调
训练超参	学习率8e-4	学习率1e-4 + warmup策略

High-Resolution Image Synthesis with Latent Diffusion Models

Stable Diffusion vs Dalle2、Imagen...

	Text encoder		Image decoder		Dataset	Trained Paras
	model type	need to train	model type	need to train		
DALL-E	transformer	√	VQVAE 256	-	250M	12B
GLIDE	transformer	√	diffusion decoder 64	√	250M	3.5B
			diffusion upsampler 256	√		
DALL-E 2	clip text encoder	×	diffusion decoder 64	√	250M	1B + 3.5B + 700M + 300 M
	diffusion prior	√	diffusion upsampler 256	√		
			diffusion upsampler 1024	√		
Imagen	pretrained text encoder	×	diffusion decoder 64	√	460M + 400M	2B
			diffusion upsampler 256	√		
			diffusion upsampler 1024	√		
Parti	transformer	√	VIT-VQGAN 256	√	400M + 400M + 4E	20B
			GAN SR 1024	√		

SD使用CLIP作为Text Encoder，不可训练，可训练的仅有UNet：860M参数
数据集：485M+170M+600M 共训练了28亿图片-文本paired对

对于那些渴望接触 LLMs（Large Language Model 大语言模型）的开发人员来说，探索和应用开发的闸门现在已经打开，应用开始大量涌现。

第 4 波浪潮：杀手级应用出现（现在），随着平台层的稳固，模型继续变得更好、更快和更便宜，模型的获取趋于免费和开源，应用层的创造力已经成熟。

正如移动设备通过 GPS、摄像头和网络连接等新功能释放了新类型的应用程序一样，我们预计这些大型模型将激发生成式 AI 应用程序的新浪潮。就像十年前移动互联网的拐点被一些杀手级应用打开了市场一样，我们预计生成式 AI 的杀手级应用程序也会出现，比赛开始了。

3.2 宏观经济

目前国内的 AIGC 企业没有一线的 VC 投，做不出能够媲美 Open.AI 的公司。通过我们和海内外从事 AIGC 业务的工作人员的访谈，总结出以下原因：

- （1）中国厂商投入不够。企业层面做 AIGC 的主要是中国的大厂，例如百度、华为等，这些企业目前主要考虑的如何降本增效，AIGC 业务需要各部门的协调，而各部门由于竞争关系导致内耗较多，因此在这块业务投入不多。而代表政府的机构，如北京智源研究院，其投入约为 10 亿元，而 Open.AI 的累计投入超过 20 亿美元。
- （2）科学家和人才密度不够。中国缺少做 AIGC 方向的顶级学者，同时人才交叉密度也和美国相比有较大差距。在美国，例如 Deepmind.AI 研发团队超过 800 人，各个学科的 PHD 都有。而国内缺乏 AI 的领军人物，有一定知名度的学者也大都不写论文不带人才了。同时中国的人才交叉密度目前仍不足。最后在

创业领域，据我们了解，中国做 Nerf 和 diffusion model 的科学家有一波人出来创业了，但是他们普遍缺乏对产品的思考，而目前拥有产品管理经验的人缺少技术，因此如果能将二者结合到一起会是一个好的机会。

（3）数据和算力。中国的 AIGC 生态没有搭建起来，很多 API 都没有开源出去，因此获得的数据相比已经开源的 Open.AI 和 Stability 等公司会减少很多。同时中国企业的算力与美国也有差距，像智源等公司用于训练大模型的芯片和显卡数量都有限。

（4）资本的耐心。中国的资本比较缺乏耐心，会要求被投企业很快有回报。但 AIGC 需要投入很多研发成本，需要很多模型和算法的迭代与优化，发论文在资本眼里是不 make sense 的。

3.3 政策

中国政府的支持

四、Players

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。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。

表 4-1 国外 AIGC 现有变现方式

方式	举例
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
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，我们认为后期会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。