



## 传媒

优于大市（维持）

证券分析师

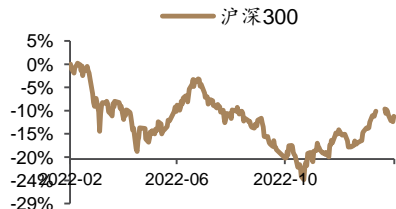
马笑

资格编号：S0120522100002

邮箱：maxiao@tebon.com.cn

研究助理

市场表现



相关研究

- 1.《传媒互联网行业周报：ChatGPT加速商业化，AIGC或带来泛娱乐和互联网的革新》，2023.2.5
- 2.《分众传媒（002027.SZ）首次覆盖报告：当风轻借力，一举入高空》，2023.2.1
- 3.《腾讯控股（0700.HK）22Q4业绩前瞻及观点更新：宏观经济波动致收入端承压，降本增效成果释放带来利润端高增速》，2023.1.31
- 4.《百度集团-SW（9888.HK）22Q4及2022年全年业绩前瞻：疫情扰动或致经营短期承压，展望后续收入业绩双端修复》，2023.1.30
- 5.《春节档期票房超65亿，行业持续复苏估值业绩双升可期》，2023.1.28

# AIGC 专题一：探析 AIGC 的技术发展和应用

投资要点：

- **AIGC 成为新的内容生产方式，跨模态生成值得重点关注。**区别于 PGC 与 UGC，AIGC 是利用人工智能技术自动生成内容的新型生产方式。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，其中，跨模态生成需要重点关注。
- **自然语言处理（NLP）赋予了 AI 理解和生成能力，大规模预训练模型是 NLP 的发展趋势。**NLP 的两个核心任务分别是自然语言理解（NLU）和自然语言生成（NLG）。以 ELMo、BERT、GPT 为代表的预训练模型，降低了 NLP 的技术门槛。ELMo 解决了“一词多义”的问题；BERT 通过 MLM（类似于完形填空）和 NLP（判断句子是否相连）进行预训练，增强了上下文的理解能力。GPT 通过预测下一个词，获得了生成能力；GPT-3 在此基础上使用了更大的数据和更大模型，无需针对下游任务进行传统的微调，并且采用了小样本学习提升生成效果。
- **ChatGPT 是 NLP 发展中具有里程碑式意义的模型之一。**ChatGPT 是 OpenAI 从 GPT-3.5 系列中的模型进行微调产生的聊天机器人模型。它能够通过学习和理解人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流，甚至能完成撰写邮件、视频脚本、文案、翻译、代码等任务。
- **生成模型赋予了 AI 创造力，扩散模型是最前沿的技术之一。**AIGC 的快速发展归功于生成算法领域的技术积累。GAN 的核心思想是“生成”与“对抗”，相比传统的深度神经网络，GAN 能产生更好的生成样本，但是仍需解决应用中的问题。**扩散模型较 GAN 更接近人的思维模式，是基于马尔科夫链，通过学习噪声来生成数据。**扩散模型实现了跨模态应用，包括 OpenAI 的 GLIDE 和 DALL·E 2、谷歌的 Imagen、Stability AI 的 Stable Diffusion 等。
- **人工智能由单模态智能，向多种模态融合方向发展。**建立统一的、跨场景、多任务的多模态基础模型或将成为人工智能发展的主流趋势之一。CLIP 模型将语言信息和图像信息联合训练，能够链接文本和图片，成为跨模态生成应用的一个重要节点，“CLIP+其他模型”在跨模态生成领域成为一种较为通用的做法。2022 年，微软提出的 BEiT-3 多模态基础模型，在视觉-语言任务处理上具备出色表现，包括视觉问答、图片描述生成和跨模态检索等。多模态提高了基础模型的灵活性，使其在其他模态的应用中发挥新的潜质。
- **未来，值得关注的技术要素包括：**长文本生成、开放式文本生成、NeRF 模型、扩散模型、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等。
- **投资建议：把握 AIGC 技术催化和商业落地的投资机会。**技术发展有望促进生产效率提升，并进一步创造新的消费和需求，有利于文娱内容和互联网行业。**在 AIGC 和 ChatGPT 方面，我们建议持续关注技术发展和应用情况，把握技术催化和商业化落地带来的投资机会：**1) 具备 AIGC 和 ChatGPT 的技术探索和应用的公司：百度集团-SW、商汤-W、万兴科技、拓尔思等；2) 具有海量内容素材且具有 AIGC 探索布局的，图片/文字/音乐/视频内容及平台公司腾讯控股，阅文集团、美图公司、视觉中国、中文在线、汉仪股份、昆仑万维、天娱数科、风语筑等。
- **风险提示：**技术发展不及预期、监管政策变化、知识产权相关问题等。

## 内容目录

1. AIGC 的跨模态生成值得重点关注 .....	5
2. 基于大模型预训练的 NLP 赋予了 AI 理解和生成能力 .....	5
2.1. ELMo 将输出的词向量作为特征，解决了“一词多义” .....	7
2.2. BERT 通过上下文预训练，提高了理解能力 .....	8
2.3. GPT-3 利用超大数据和超大模型，省去了微调的过程 .....	9
2.3.1. ChatGPT 趋近人类价值观及意图，有望进一步商业化 .....	11
3. 基于大模型的主流生成模型赋予了 AI 创造力 .....	12
3.1. GAN 是基于“生成”与“对抗”提出的生成模型 .....	13
3.2. 扩散模型更接近人的思维模式，实现了跨模态应用 .....	13
3.2.1. 扩散模型实现了从文本到图像的跨模态应用 .....	14
4. 人工智能由单模态智能，向多种模态融合方向发展 .....	16
5. AIGC 技术不断发展，有望带来更多商业化价值 .....	16
6. 投资建议：把握 AIGC 技术催化和商业落地的投资机会 .....	18
7. 风险提示 .....	19

## 图表目录

图 1: 内容生产模式的四个发展阶段 .....	5
图 2: 自然语言处理的发展历史 .....	6
图 3: 大规模预训练模型大幅提升研发效率 .....	6
图 4: BERT 的每一层都是双向模型; GPT 为单向模型; ELMo 为单项模型叠加 .....	7
图 5: 大规模预训练语言模型的发展趋势之一是参数量不断增加 .....	7
图 6: ELMo 基于 RNN, 双向 RNN 可以利用前后文信息 .....	8
图 7: ELMo 将词嵌入作加权和, 给出最终的词向量 .....	8
图 8: ELMo 解决了一词多义问题, 且兼顾语义与词性 .....	8
图 9: BERT 的结构是 Transformer 中的 Encoder 部分 .....	9
图 10: BERT 采用了预训练+微调的两阶段模型 .....	9
图 11: NSP 任务中输入向量生成示意图 .....	9
图 12: GPT 模型结构 .....	10
图 13: GPT-3 将例子直接作为模型输入 .....	10
图 14: Few Shot 较 Zero Shot 在模型增大后表现提升更显著 .....	10
图 15: GPT-3.5 的进化树 .....	11
图 16: RLHF 工作原理 .....	11
图 17: ChatGPT 的训练模式 .....	11
图 18: ChatGPT 能够进行对话, 甚至能完成撰写代码 .....	12
图 19: ChatGPT 上线 2 个月后月度用户数量破 1 亿 .....	12
图 20: OpenAI 推出付费订阅项目 ChatGPT Plus, 价格\$20/月 .....	12
图 21: 不同生成模型之间的区别 .....	13
图 22: GAN 的基本思想 .....	13
图 23: 训练后的 GAN 生成案例 .....	13
图 24: Diffusion Model 基于马尔科夫链, 通过学习噪声来生成数据 .....	14
图 25: GLIDE 可以进行文本到图像的生成 .....	14
图 26: GLIDE 可以执行图像修复 .....	14
图 27: unCLIP 模型示意图 .....	15
图 28: DALL·E 2 的生成案例 .....	15
图 29: Image 的架构示意图 .....	15
图 30: Image 的生成案例 .....	15
图 31: Stable Diffusion 生成效果 .....	16
图 32: Stable Diffusion 2.0 引入了图像深度扩散模型 .....	16

---

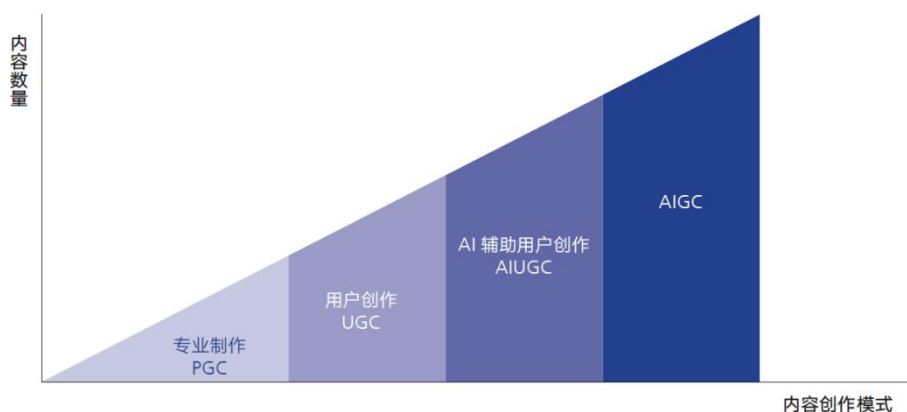
图 33: AIGC 相关技术场景及成熟度分类.....	17
表 1: ELMo、BERT 与 GPT 系列的对比 .....	6
表 2: 国外主要 AIGC 预训练模型一览表.....	17
表 3: 部分公司在 AIGC 领域的布局 (截至 2023/2/6) .....	18

## 1. AIGC 的跨模态生成值得重点关注

根据中国信息通信研究院发布的《人工智能生成内容（AIGC）白皮书》，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生产方式”。传统 AI 关注对于现有数据的分析能力，而 AIGC 注重生成创造新事物的能力。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，其中，跨模态生成需要重点关注。

自然语言处理（NLP）与生成模型是 AIGC 发展的重要技术支柱。自然语言处理增强 AIGC 的认知能力，使得人类语言与计算机语言之间实现互通；生成模型升级 AIGC 内容创作能力，使生成的作品更接近人类思维与偏好。

图 1：内容生产模式的四个发展阶段



资料来源：《AIGC 发展趋势报告 2023》，德邦研究所

AIGC 近年来的快速发展依赖于生成模型、预训练模型和多模态等技术的融合。生成模型从 GAN 发展到后续的扩散模型，不断趋近人的思维模式。预训练模型提高了 AIGC 技术的灵活性和效果质量。基于大量数据训练的巨量参数模型可以轻松应用到下游任务中，降低了使用门槛和技术成本，推动了自然语言处理、计算机视觉等技术的发展。多模态技术提高了 AIGC 的通用性，使得图像、声音、语言等互相融合。

## 2. 基于大模型预训练的 NLP 赋予了 AI 理解和生成能力

自然语言处理（NLP）主要研究用计算机来理解和生成自然语言的各种理论和方法，属于人工智能领域的一个重要甚至核心的分支。自然语言处理（NLP）的两个核心任务分别是自然语言理解（NLU）和自然语言生成（NLG）；四大任务是序列标注（分词、语义标注等）、分类任务（文本分类、情感计算等）、句子关系判断、生成式任务（机器翻译、文本摘要等）。人工智能应用领域的快速拓展对自然语言处理提出了巨大的应用需求。

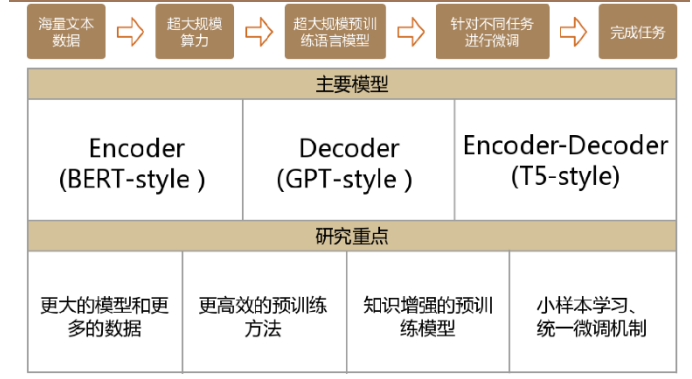
2018 年，NLP 正式进入“预训练”时代，简化了下游任务模型的接入，降低了 NLP 的技术门槛。预训练是指，首先在一个源任务上训练一个初始模型，然后在下游任务（也称目标任务）上继续对该模型进行精调，从而达到提高下游任务准确率的目的。预训练的优势在于，一是对庞大的文本语料库进行预处理，可以学习通用的语言表征，帮助完成下游任务；二是预训练提供了更好的模型初始化，带来更好的普适性并加速与目标任务的衔接；三是，预训练可以看作是一种正则化方法（regularization），以避免对小数据的过度拟合。

图 2：自然语言处理的发展历史



资料来源：车万翔等《自然语言处理新范式：基于预训练模型的方法》，德邦研究所

图 3：大规模预训练模型大幅提升研发效率



资料来源：澜舟科技，德邦研究所

2018 年 3 月，AllenNLP 提出了 ELMo 模型，属于 Feature-Based 模型，使用预训练模型产出的词向量作为输入，接入下游的独立的模型进行预测。

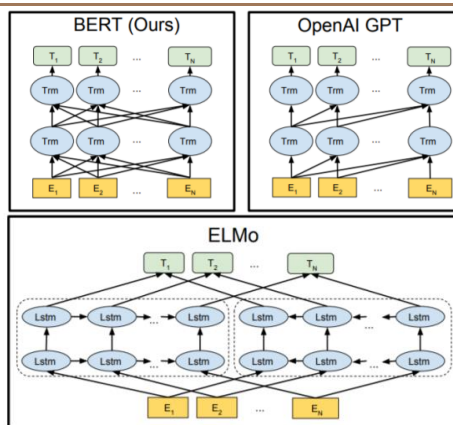
2018 年 6 月，OpenAI 提出了第 1 代 GPT 模型，开启了大规模预训练语言模型时代。GPT 属于 Finetune-Based 模型，在目标任务上精调整个模型，而不是只将模型的输出结果作为固定的词向量特征。由于预训练模型自身非常复杂，因此接入的下游任务模型可以非常简单，很大程度上降低了 NLP 的技术门槛。

表 1：ELMo、BERT 与 GPT 系列的对比

名称	时间	机构	模型				应用		
			底层结构	参数量	学习方向	NLP 学习方法	主要应用领域	优势	劣势
ELMo	2018/3	AllenNLP	RNN	94M	从左到右叠加从右到左	Feature-Based: 使用预训练模型产出的词向量作为输入，接入下游的独立的模型进行预测	ELMo 用于特征抽取，不同下游任务需要使用不同的对应模型	解决了一词多义的问题；兼顾语义与词性	RNN 模型较 Transformer 效果较弱；上下文的特征融合能力较差
BERT	2019/6	谷歌	Transformer 的 Encoder	340M	双向	Finetune-Based: 将原来的模型在下游任务上进行训练（预训练+微调）	NLP 研究	可以出色完成 NLP 除生成任务以外的其他任务，包括阅读理解，QA 等	生成任务较弱
GPT	-1	OpenAI	Transformer 的 Decoder	110M	从左右到右	Finetune-Based: 将原来的模型在下游任务上进行训练（预训练+微调）	自然语言推理、问答与常识推理、语义相似度识别、文本分类	具备生成能力	需要进行微调
	-2					无监督零样本学习	着重生成任务，例如写作、问答、翻译、写摘要	生成文本段落的能力强，内容语法工整，语义一致，效果较好；无微调，泛化能力更强	模型效果有提升空间
	-3					零样本学习、单样本学习或小样本学习	NLP 的绝大部分任务，例如面向问题的搜索、阅读理解、语义推断、机器翻译、文章生成和自动问答等	小样本学习的准确率更高，训练效果更好	逻辑推理能力有待提升

资料来源：谷歌，AllenNLP，OpenAI 等，德邦研究所

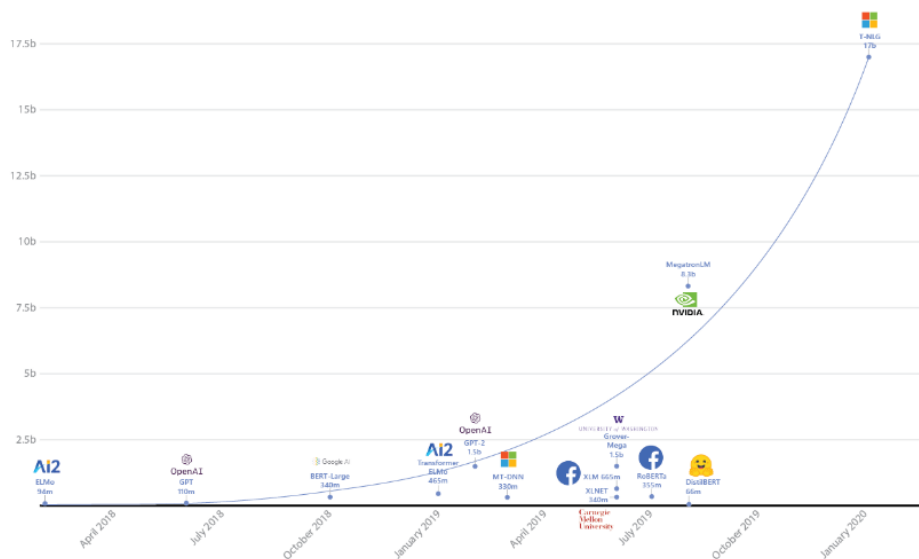
图 4: BERT 的每一层都是双向模型; GPT 为单向模型; ELMo 为单项模型叠加



资料来源: Jacob Devlin et al. 《BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding》, 德邦研究所

大规模预训练语言模型的发展趋势之一, 是使用更大规模的数据训练更大的模型。ELMo 的参数量约为 94M, GPT-1 的参数量约为 110M, BERT 的参数量为 340M, GPT-2 的参数量约为 1.5B, GPT-3 的参数量约为 175B。随着技术的不断成熟, 大模型在开发成本、易用性、开发周期、性能上会更具优势, 给产品化和商业化带来更多可能性。

图 5: 大规模预训练语言模型的发展趋势之一是参数量不断增加



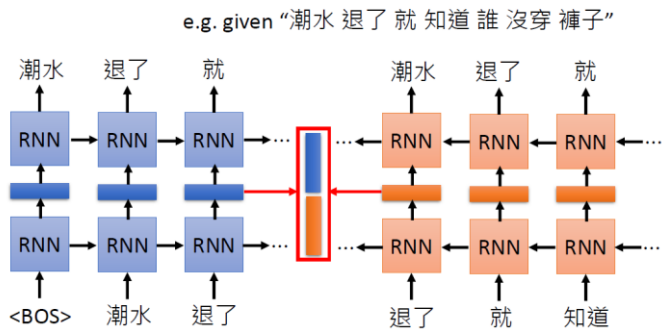
资料来源: 《Turing-NLG: A 17-billion-parameter language model by Microsoft》, 德邦研究所

## 2.1. ELMo 将输出的词向量作为特征, 解决了“一词多义”

ELMo (Embeddings from Language Models) 是由 Allen AI 于 2018 年 3 月发表的基于 RNN 的语言模型。ELMo 的主要突破在于引入词向量 (Word Embedding) 作为输出, 解决了“一词多义”的问题。

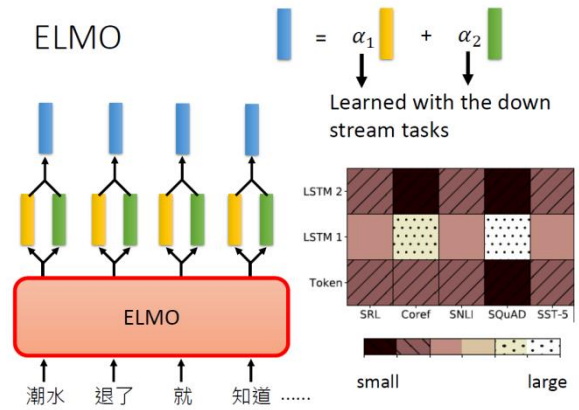
ELMo 在 RNN 的基础上, 整合上下文信息, 获得对应的词向量。RNN 所实现的是, 输入一个词汇, 就可以输出它之后所对应的词 (Token)。例如, 输入 <BOS> 表示开始, RNN 就输出“潮水”; 输入“潮水”, 就输出“退了”, 以此类推。双向的 RNN 会将正向和逆向的词嵌入接起来, 作为最终的词向量, 因此前后文的信息都可以使用到。RNN 可以有多层, 因此会产生很多个上下文词向量, ELMo 会将所有的上下文嵌入作加权和, 权重是根据不同的下游任务学习出的。

图 6: ELMO 基于 RNN, 双向 RNN 可以利用前后文信息



资料来源: 李宏毅《Putting Words into Computers》, 德邦研究所

图 7: ELMO 将词嵌入作加权和, 给出最终的词向量



资料来源: 李宏毅《Putting Words into Computers》, 德邦研究所

由于 ELMO 形成词向量是根据不同的下游任务学习出的, 因此每一个词并非对应唯一的词向量。同时, 对于 ELMO 而言, 上下文中意思相近的词拥有相近的词向量。例如, 在句子“我喜欢吃土豆”中, “土豆”的表示应该和“马铃薯”相似; 而在句子“我在土豆上看电影”中, “土豆”的表示则应该和“爱奇艺”相似。

图 8: ELMO 解决了一词多义问题, 且兼顾语义与词性

Source	Nearest Neighbors
GloVe play	playing, game, games, played, players, plays, player, Play, football, multiplayer
Chico Ruiz made a spectacular <u>play</u> on Alusik 's grounder {...}	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent <u>play</u> .
Olivia De Havilland signed to do a Broadway <u>play</u> for Garson {...}	{...} they were actors who had been handed fat roles in a successful <u>play</u> , and had talent enough to fill the roles competently , with nice understatement .

Table 4: Nearest neighbors to “play” using GloVe and the context embeddings from a biLM.

资料来源: Matthew E. Peters et al.《Deep contextualized word representations》, 德邦研究所

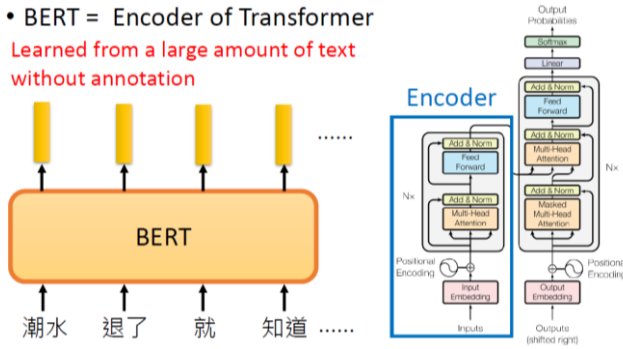
## 2.2. BERT 通过上下文预训练, 提高了理解能力

BERT(Bidirectional Encoder Representation from Transformer)是由 Google 于 2019 年 6 月发布的语言模型。BERT 的双向性 (Bidirectionality) 使其可以充分利用文本上下文, 并通过 MLM (类似完形填空) 和 NLP (判断句子是否相连) 任务进行预训练, 在 NLP 任务 (除生成任务外) 中取得了更好的效果, 例如阅读理解等。

BERT 采用了预训练和微调的两阶段模型, 属于半监督学习模型 (Semi-supervised Learning)。BERT 的预训练过程采用大量的无标签数据, 是自监督学习 (Self-supervised Learning); 微调过程采用少量的有标签数据, 因此是有监督训练 (Supervised Learning)。

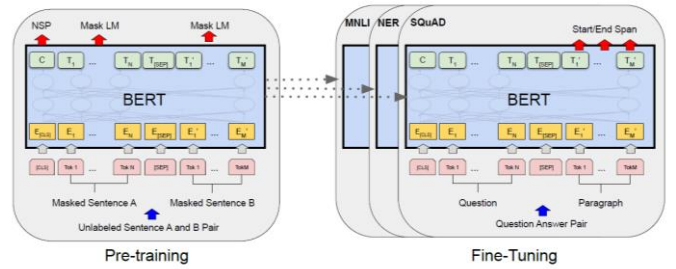


图 9: BERT 的结构是 Transformer 中的 Encoder 部分



资料来源: 李宏毅《Putting Words into Computers》, 德邦研究所

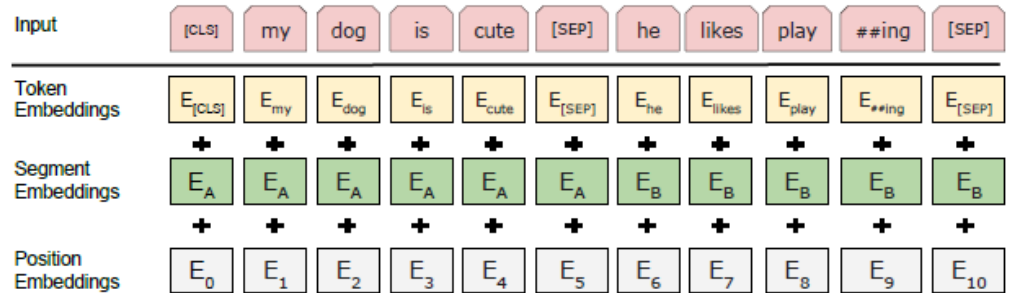
图 10: BERT 采用了预训练+微调的两阶段模型



资料来源: Jacob Devlin et al.《BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding》, 德邦研究所

**BERT 的预训练有 Masked LM(MLM)和 Next Sentence Prediction(NSP) 两种方式。**MLM 类似于完形填空, 是将输入的句子中 15% 的词随机遮盖, BERT 需要猜测遮盖住的部分是什么字, 训练目标是 minimized 预测词的误差。NSP 是从训练集中拿出两个句子, BERT 需要判断两个句子是否相连。如果两个句子是连在一起的, 那么就输出 Yes, 如果不是, 那就输出 No。

图 11: NSP 任务中输入向量生成示意图



资料来源: Jacob Devlin et al.《BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding》, 德邦研究所

**BERT 的双向性使其可以更好地理解人类语言, 但在一定程度上限制了它的生成能力。**相较于单向模型仅通过上文即可进行后文续写生成, 双向模型在缺乏下文语境时生成能力受限。

### 2.3. GPT-3 利用超大数据和超大模型, 省去了微调的过程

GPT (Generative Pre-Training) 是由 OpenAI 于 2018 年 6 月提出的。**GPT 与 BERT 同样是基于 Transformer, 但是单向模型, 因此它具有更好的生成能力。**GPT 需要做的是预测下一个词是什么。例如, 输入<BOS>进入 GPT, 输出“潮水”; 输入“<BOS>潮水”, 输出“退了”; 输入“<BOS>潮水退了”, 输出“就”, 以此类推。

图 12: GPT 模型结构

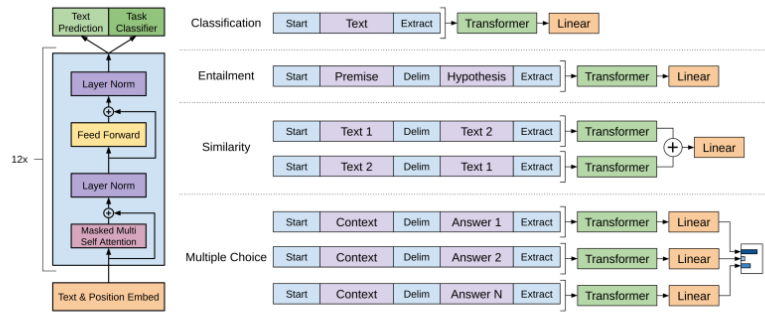


Figure 1: (left) Transformer architecture and training objectives used in this work. (right) Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

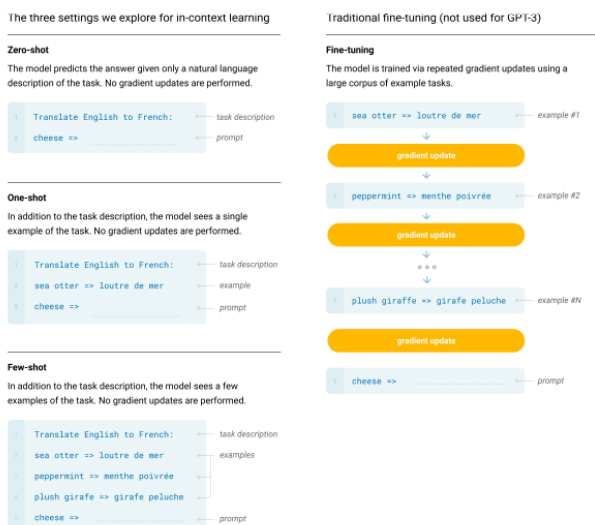
资料来源：Alec Radford et al. 《Improving Language Understanding by Generative Pre-Training》，德邦研究所

GPT-1 与 BERT 类似，分为预训练和微调的两阶段模型，属于半监督学习模型。

GPT-2 在 GPT 的基础上进行改进，通过模型扩容并学习更多数据，实现了零样本学习 (Zero-shot Learning)。GPT-2 将 Transformer 模型参数扩容，堆叠的层数增加到 48 层，从而可以保存更多的语言学知识；可学习参数量达到了 1.5B。同时，GTP-2 使用了超过 800 万个互联网网页，合计 40GB 的文本数据进行预训练。这些数据覆盖主题广，并对数据的质量也进行了筛选，因此训练出的模型具有更好的通用性。基于对大量优质样本学习，GPT-2 可直接通过无监督的方式去完成下游任务，不需要下游任务的任何标注信息，因此基本实现了训练一个模型可在多个任务中使用。

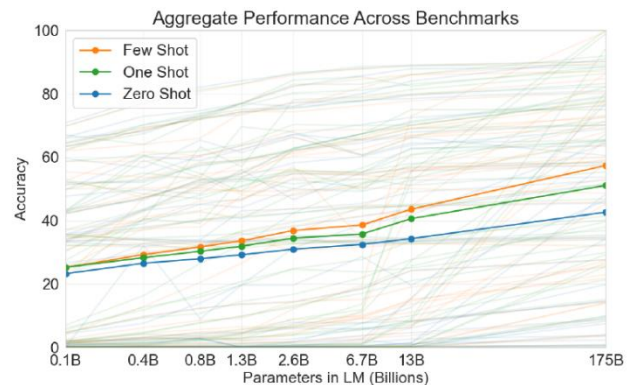
GPT-3 结构在 GPT-2 之上继续进行改进。一是使用了更大的数据量并扩大了模型。GPT-3 数据量达到 45TB，约为 GPT-2 的 1000 倍；可学习参数达到 175B，约为 GPT-2 的 100 倍。二是，GPT-3 不再追求零样本学习，而是通过小样本学习 (Few-shot Learning)，获得更好的训练效果；后续对下游任务开放微调以获得更出色的表现。不同于传统基于下游任务的微调模式，GPT-3 将例子直接作为模型输入。给 GPT 读过这些例句，它自动知道怎么解接下来的问题。2021 年 12 月 14 日，GPT-3 对下游用户开放了微调，微调后的 GPT-3 准确性大幅提高。例如，在解小学数学问题时，准确性提高 2 到 4 倍。

图 13: GPT-3 将例子直接作为模型输入



资料来源：Tom B. Brown et al. 《Language Models are Few-Shot Learners》，德邦研究所

图 14: Few Shot 较 Zero Shot 在模型增大后表现提升更显著

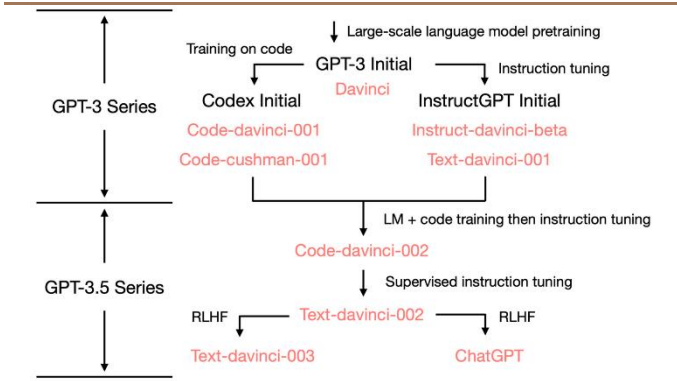


资料来源：Tom B. Brown et al. 《Language Models are Few-Shot Learners》，德邦研究所

### 2.3.1. ChatGPT 趋近人类价值观及意图，有望进一步商业化

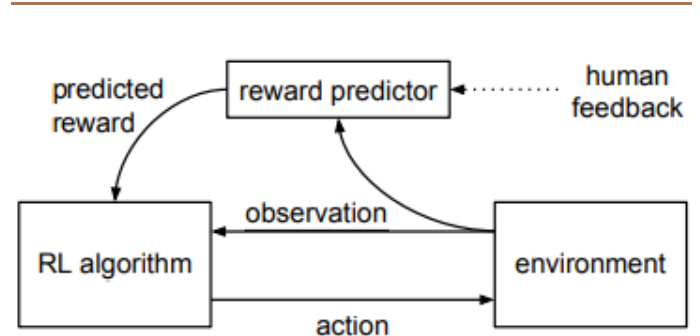
ChatGPT 是 OpenAI 使用 RLHF (Reinforcement Learning from Human Feedback) 训练出的模型，是从 GPT-3.5 系列中的模型进行微调的。RLHF 基于人类的反馈调整模型，一定程度上解决了 AI Alignment (人工智能对齐) 的问题，即要求 AI 系统的目标和人类的价值观与意图保持一致。

图 15: GPT-3.5 的进化树



资料来源：符尧《拆解追溯 GPT-3.5 各项能力的起源》，德邦研究所

图 16: RLHF 工作原理



资料来源：Paul Christiano et al.《Deep Reinforcement Learning from Human Preferences》德邦研究所

参考李宏毅教授给出的解读，Chat GPT 的模型学习方法与 Instruct GPT 类似，分为四个阶段。

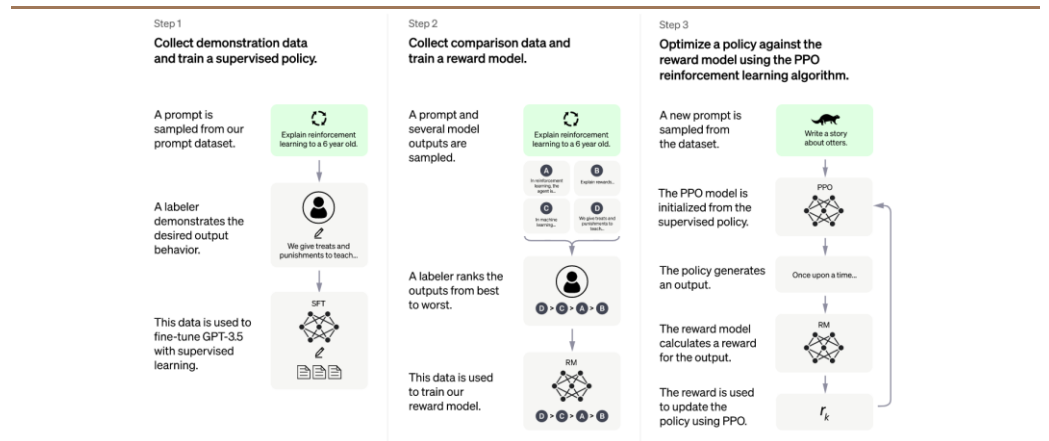
第一阶段，学习文字接龙。例如，输入“你好”，通过 GPT 后输出“美”。GPT 每一次的输出可能都是不同的，因为在网络上的语句中，“你好”后面有多种接法，例如“你好高”、“你好吗”，而输出结果是遵循概率分布的，高频组合出现的概率较高。

第二阶段，由人类老师引导文字接龙的方向，即找人来思考想问 GPT 的问题，并人工提供正确答案。不需要穷尽所有问题，只需要告诉 GPT 人类的偏好。

第三阶段，生成奖励模型 (Reward Model) 来模仿人类老师的喜好。例如，输入“台湾最高的山是哪座？”进入 GPT，输出可能有两个结果，输出一是“玉山”，输出二是“谁来告诉我呀”。两个回答都可以与问题衔接，但是人类的偏好是输出一，因此 Reward Model 给两个输出排序时，输出一是高于输出二。

第四阶段，使用奖励模型作为奖励函数，近端策略优化 (PPO) 算法微调策略，通过调整参数获得最大的奖励。

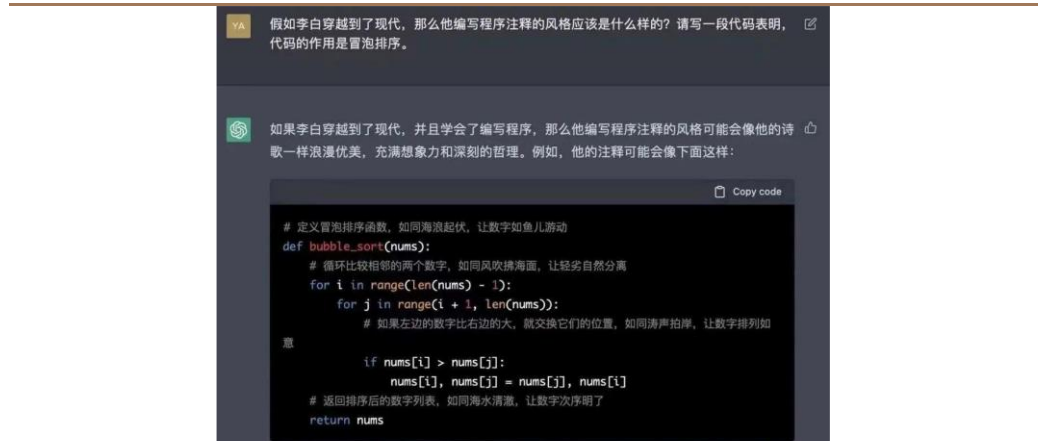
图 17: ChatGPT 的训练模式



资料来源：OpenAI，德邦研究所

ChatGPT 的应用包括但不限于，问题解答（包括解释代码块的作用）、求解数学方程式、写作文本、调试和修复（包括更正代码错误）、语际翻译、总结文本并检测文本中的关键词、分类、提出建议等。

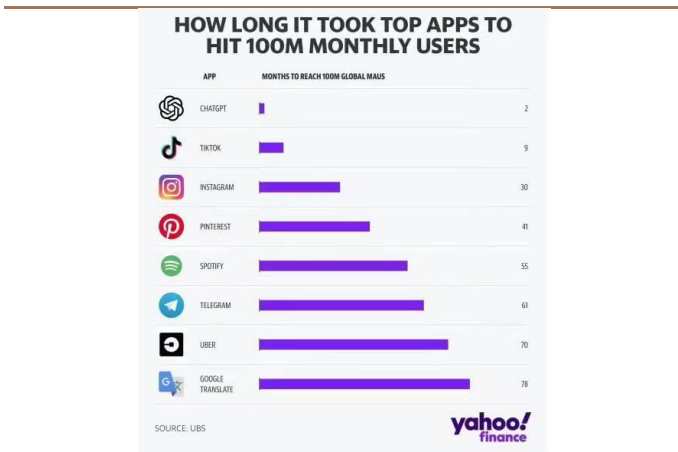
图 18: ChatGPT 能够进行对话，甚至能完成撰写代码



资料来源：量子位，德邦研究所

ChatGPT 月度用户已破亿，正在逐步探索商业化途径。根据 UBS 统计数据，ChatGPT 上线 2 个月后月度用户数量破 1 亿。根据 SimilarWeb 数据，截至 1 月 30 日，过去 28 天内日独立访客数已接近 1300 万。2023 年 2 月 1 日，OpenAI 推出付费订阅项目 ChatGPT Plus，价格为 \$20/月，目前面向美国用户。付费服务包括，在高峰时段也可常规访问 ChatGPT、更快的响应、以及优先访问新功能和改进。与此同时，ChatGPT 将继续提供免费访问。此外，OpenAI 还将很快推出 ChatGPT API，积极探索低成本计划、商业计划和数据包的选项，以提高产品的可用性。

图 19: ChatGPT 上线 2 个月后月度用户数量破 1 亿



资料来源：yahoo finance, UBS, 德邦研究所

图 20: OpenAI 推出付费订阅项目 ChatGPT Plus，价格 \$20/月

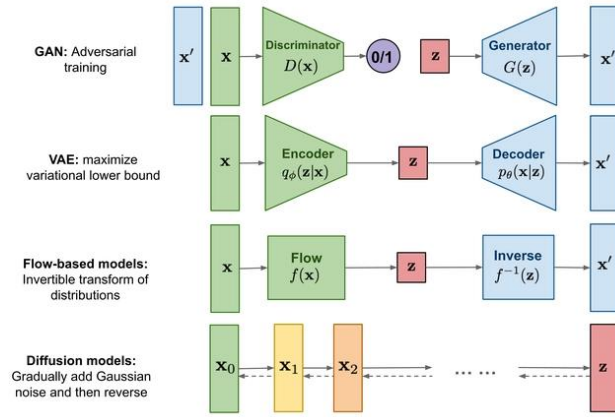


资料来源：OpenAI, 德邦研究所

### 3. 基于大模型的主流生成模型赋予了 AI 创造力

AIGC 的快速发展归功于生成算法领域的技术积累，其中包含了：生成对抗网络 (GAN)、变分自动编码器 (VAE)、标准化流模型 (NFs)、自回归模型 (AR)、能量模型和扩散模型 (Diffusion Model) 等。

图 21：不同生成模型之间的区别



资料来源：Lilian Weng 《What are Diffusion Models?》，德邦研究所

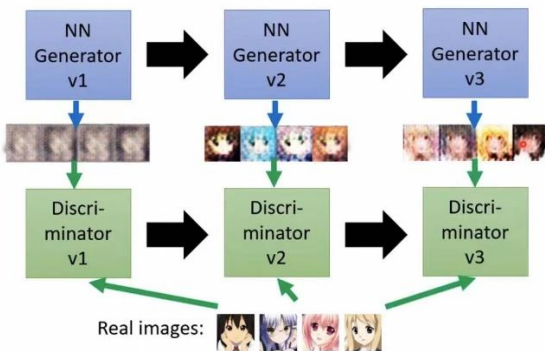
### 3.1. GAN 是基于“生成”与“对抗”提出的生成模型

2014 年，Ian J. Goodfellow 提出了 GAN，是一种深度神经网络架构，其核心思想是“生成”与“对抗”。GAN 网络结构包含两个模型，一个是生成模型（Generator, G），另一个是判别模型（Discriminator, D）。生成模型通过输入的数据生成图像；判别模型工作主要是负责判断生成模型所生成的图像是人工绘制还是 AI 生成。

以画人像为例，第一代  $G_1$  不知道如何画，因此生成的图像模糊；第一代  $D_1$  学习到，判断的方法是图片是否有眼睛；第二代  $G_2$  会生成眼睛以“骗”过  $D_1$ ；第二代  $D_2$  又学习到可以用是否有嘴巴来判别；由此第三代  $G_3$  会生成嘴巴，以此类推。直到最后，当判别模型分辨不出生成结果是否真实的时候（判别概率为 0.5），模型达到纳什平衡点，即为最理想的状态。

图 22：GAN 的基本思想

#### Basic Idea of GAN



资料来源：李宏毅 《Generation》，德邦研究所

图 23：训练后的 GAN 生成案例



资料来源：Ian J. Goodfellow et al. 《Generative Adversarial Nets》，德邦研究所

GAN 是生成模型的一大突破，相比传统的深度神经网络，能产生更好的生成样本，但是仍需解决应用中的问题。一是，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。二是，缓解 GAN 的模式坍塌（model collapse），即生成样本大量重复相似。它会造成训练结果冗余、生成图像质量差、样本单一等问题。

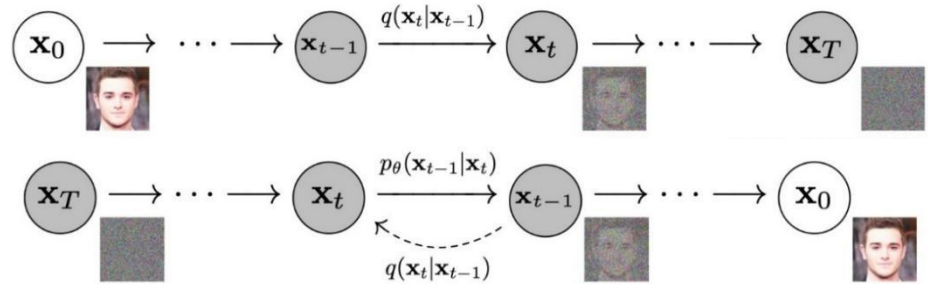
### 3.2. 扩散模型更接近人的思维模式，实现了跨模态应用

扩散模型（Diffusion Model）最早是 2015 年由 Jascha Sohl-Dickstein 等人提出的。2020 年 Jonathan Ho 等人发表了《Denoising Diffusion Probabilistic

Models》之后引发众多关注。与 GAN 等模型最大区别在于，扩散模型是基于马尔科夫链，通过学习噪声来生成数据。

扩散模型通过正向扩散过程进行学习，并通过逆扩散过程实现生成。扩散过程中，正向扩散过程，是从真实图像  $X_0$  不断学习噪声，变换成纯高斯噪声的图片  $X_T$ ，即一个马尔科夫链。模型需要通过逆扩散过程进行生成， $X_T$  通过沿着马尔科夫链向后遍历，可以重新生成新的数据  $X_0$ 。

图 24: Diffusion Model 基于马尔科夫链，通过学习噪声来生成数据



资料来源：AssemblyAI, Jonathan Ho et al. 《Denoising Diffusion Probabilistic Models》，德邦研究所

扩散模型的生成逻辑相比其他的模型更接近人的思维模式。本质上，扩散模型的工作原理是通过连续添加高斯噪声来破坏训练数据，然后通过反转这个噪声过程来学习恢复数据。训练后，可以通过简单地将随机采样的噪声传递给学习的去噪过程来生成数据。相比于其他模型，扩散模型的优势在于生成的图像质量更高，且无需通过对抗性训练，这使得其训练的效率有所提升。同时，扩散模型还具有可扩展性和并行性。

### 3.2.1. 扩散模型实现了从文本到图像的跨模态应用

基于扩散模型的示例架构包括 OpenAI 的 GLIDE 和 DALL·E 2、谷歌的 Imagen、Stability AI 的 Stable Diffusion 等。

GLIDE (Guided Language to Image Diffusion for Generation and Editing), 是 OpenAI 推出的文本引导图像生成模型。GLIDE 的训练步骤为，首先使用文本条件扩散模型作为基本模型进行预训练，利用到了 Transformer 对文本进行编码，使得文本可以对图像进行引导。然后再对模型进行微调，实现无分类器引导，保留文本引导图像生成的同时也可执行无条件图像生成。GLIDE 还支持通过文本作为提示 (prompt) 对图像进行修复，例如，将人像中的头发抹去，提示中输入“红发男子”，GLIDE 会对人像的头发进行修复。

图 25: GLIDE 可以进行文本到图像的生成

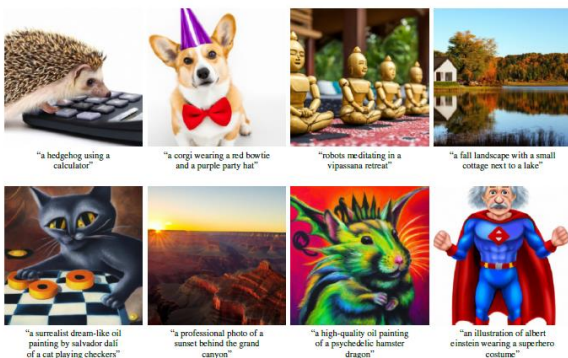
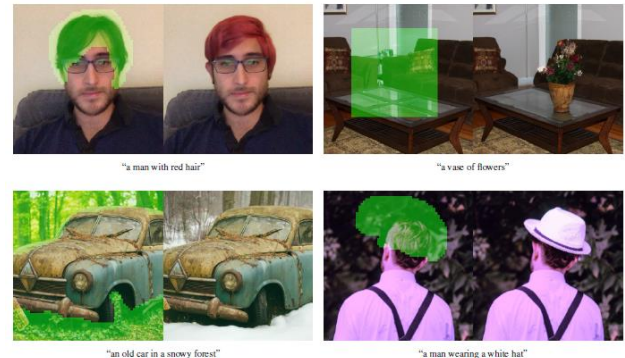


图 26: GLIDE 可以执行图像修复

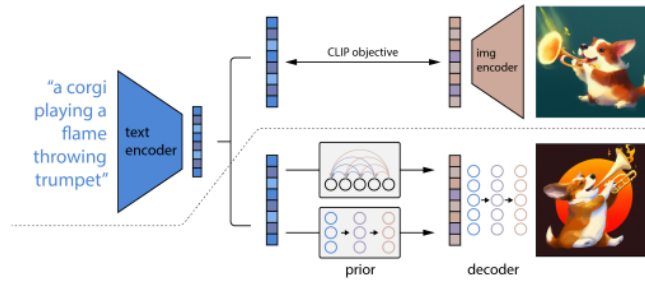


资料来源: Alex Nichol et al.《GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models》, 德邦研究所

资料来源: Alex Nichol et al.《GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models》, 德邦研究所

**DALL·E 2 是 OpenAI 于 2022 年 4 月发布的系统。**DALL·E 2 系统背后的模型称为 unCLIP, 使用 CLIP 作为图文预训练模型, 获得文本描述和图像的嵌入, 再通过扩散模型思路构建生成过程。CLIP 是 OpenAI 发布的一个多模态模型, 基于大量图像和文本对上训练, 可用于查找最能代表图像的文本片段, 或查找给定文本查询的最合适图像。DALL·E 2 生成的图像分辨率较 DALL·E 和 GLIDE 提高了 4 倍。

图 27: unCLIP 模型示意图



资料来源: Aditya Ramesh et al.《Hierarchical Text-Conditional Image Generation with CLIP Latents》, 德邦研究所

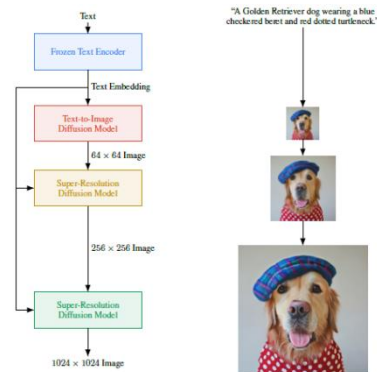
图 28: DALL·E 2 的生成案例



资料来源: Aditya Ramesh et al.《Hierarchical Text-Conditional Image Generation with CLIP Latents》, 德邦研究所

谷歌于 2022 年 6 月发布的 Imagen, 利用了大型 Transformer 语言模型在理解文本方面的能力, 以及扩散模型在高保真图像生成方面的能力。Imagen 的工作模式为, 在用户输入文本要求后, 条件扩散模型将文本嵌入映射  $64 \times 64$  的图像中; Imagen 进一步利用文本条件超分辨率扩散模型对  $64 \times 64$  的图像进行升采样为  $256 \times 256$ , 再从  $256 \times 256$  升到  $1024 \times 1024$ 。

图 29: Image 的架构示意图



资料来源: Chitwan Saharia et al.《Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding》, 德邦研究所

图 30: Image 的生成案例



资料来源: Chitwan Saharia et al.《Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding》, 德邦研究所

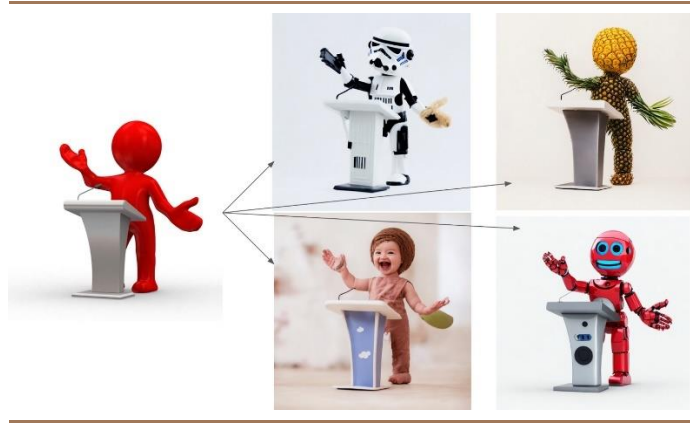
**Stable Diffusion 是 Stability AI 发布的文本到图像模型, 目前已更新到 2.0 版本。**Stable Diffusion 的生成速度较快, 在几秒钟内便可创造出艺术作品。2.0 版本更新了文本转图片扩散模型 (Text-to-Image Diffusion Models) 并加入了超分辨率扩散模型 (Super-resolution Upscaler Diffusion Models), 可以生成分辨率为  $2048 \times 2048$  甚至更高的图像。同时, 引入了图像深度扩散模型 (Depth-to-Image Diffusion Model), 扩展了图像到图像功能。

图 31: Stable Diffusion 生成效果



资料来源: Stability AI, 德邦研究所

图 32: Stable Diffusion 2.0 引入了图像深度扩散模型



资料来源: Stability AI, 德邦研究所

## 4. 人工智能由单模态智能，向多种模态融合方向发展

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，将这种能力被称为多模态学习 (Multi-modal Learning, MML)。建立统一的、跨场景、多任务的多模态基础模型或将成为人工智能发展的主流趋势之一。

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，使其能够有效提取长序列特征，将视觉的区域特征和文本特征序列相匹配。

CLIP 模型的出现，成为跨模态生成应用的一个重要节点。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。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。

2022 年，微软提出的 BEiT-3 多模态基础模型，在视觉-语言任务处理上具备出色表现，包括视觉问答、图片描述生成和跨模态检索等；同时能够更加轻松地完成多模态编码和处理不同的下游任务。

语音、视觉和多模态预训练模型将加速人工智能向通用基础模型方向演进。在这个演进过程中，深度学习与强化学习相互促进发展，融合大量行业知识，模型将具备在不断变化的环境中快速适应的灵活性。基础模型不再局限于其传统用法，有望在其他模态的应用中发挥出新的潜质。例如，通常用于游戏的强化学习 PPO (Proximal Policy Optimization)，也被应用到了 NLP 领域的 ChatGPT 中。

## 5. AIGC 技术不断发展，有望带来更多商业化价值

1 月 17 日，微软 CEO 纳德拉在世界经济论坛上表示，微软正在迅速推进 OpenAI 的工具商业化，计划将包括 ChatGPT、DALL-E 等人工智能工具整合进微软旗下的所有产品中，并将其作为平台供其他企业使用，包括且不限于 Bing 搜索引擎、Microsoft Office、Azure 云服务、Teams 聊天程序等。(来自: IT 之家)

1 月 23 日，微软公司在官方博客宣布，已与 OpenAI 扩大合作伙伴关系。作为两家公司合作伙伴关系的第三阶段，微软将向 OpenAI 进行一项为多年、价值数十亿美元的投资，以加速其在人工智能 (AI) 领域的技术突破。

未来，值得关注的技术要素包括：长文本生成、开放式文本生成、NeRF 模型、扩散模型、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等。



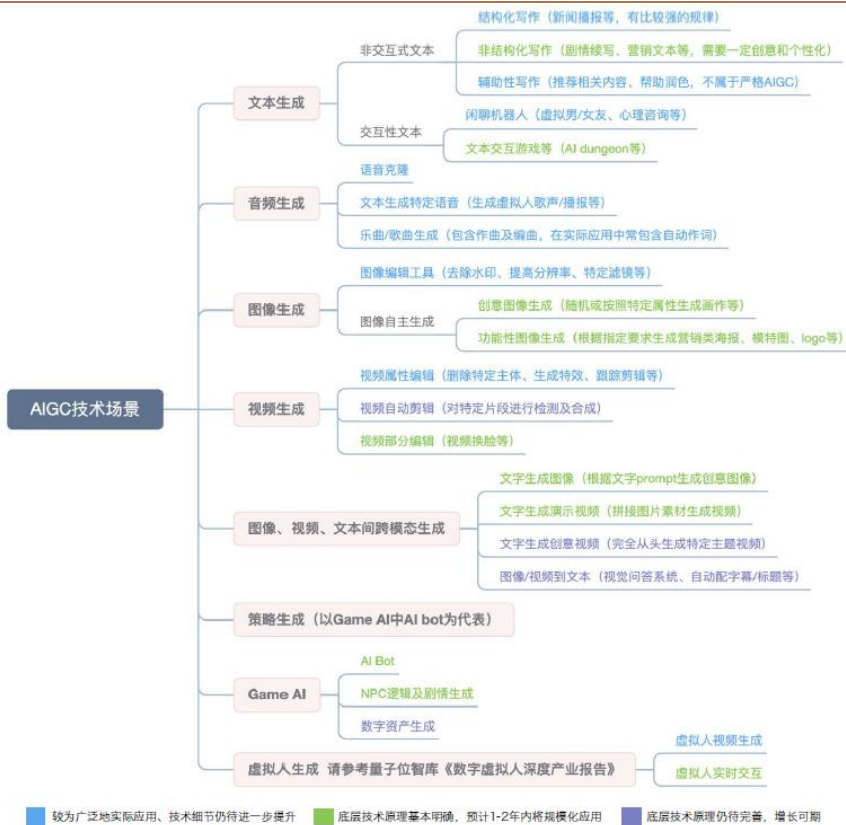
表 2: 国外主要 AIGC 预训练模型一览表

公司	预训练模型	应用	参数量	领域
谷歌	BERT	语言理解与生成	4810 亿	NLP
	LaMDA	对话系统	-	NLP
	PaLM	语言理解与生成、推理、代码生成	5400 亿	NLP
	Imagen	语言理解与图像生成	110 亿	多模态
	Parti	语言理解与图像生成	200 亿	多模态
微软	Florence	视觉识别	6.4 亿	CV
	Turing-NLG	语言理解、生成	170 亿	NLP
Facebook	OPT-175B	语言模型	1750 亿	NLP
	M2M-100	100 种语言互译	150 亿	NLP
Deep Mind	Gato	多面手的智能体	12 亿	多模态
	Gopher	语言理解与生成	2800 亿	NLP
	AlphaCode	代码生成	414 亿	NLP
Open AI	GPT3	语言理解与生成、推理等	1750 亿	NLP
	CLIP&DALL-E	图像生成、跨模态检索	120 亿	多模态
	Codex	代码生成	120 亿	NLP
	ChatGPT	语言理解与生成、推理等	-	NLP
英伟达	Megatron-Turning NLG	语言理解与生成、推理	5300 亿	NLP
Stability AI	Stable Diffusion	语言理解与图像生成	-	多模态

资料来源:《AIGC 发展趋势报告 2023》, 德邦研究所

根据量子位, 短期内有望较明显爆发的技术场景包括: 闲聊式文本生成、个性化营销文本、富情感及细节 TTS (Text To Speech)、拼凑式视频生成、基于文本的 AI 绘画、语音复刻等。

图 33: AIGC 相关技术场景及成熟度分类



资料来源: 量子位, 德邦研究所

## 6. 投资建议：把握 AIGC 技术催化和商业落地的投资机会

技术发展有望促进生产效率提升，并进一步创造新的消费和需求，有利于文娱内容和互联网行业。在 AIGC 和 ChatGPT 方面，我们建议持续关注技术发展和应用情况，把握技术催化和商业化落地带来的投资机会：1) 具备 AIGC 和 ChatGPT 的技术探索和应用的公司：百度集团-SW、商汤-W、万兴科技、拓尔思等；2) 具有海量内容素材且具有 AIGC 探索布局的，图片/文字/音乐/视频内容及平台公司腾讯控股，阅文集团、美图公司、视觉中国、中文在线、汉仪股份、昆仑万维、天娱数科、风语筑等。

表 3：部分公司在 AIGC 领域的布局（截至 2023/2/6）

港股公司						
证券代码	证券简称	主要研究领域	相关应用/布局	总市值 (亿港元)	PE (TTM)	PS (TTM)
9888.HK	百度集团-SW	深度学习开源框架	飞桨 PaddlePaddle	3,895	81.5	2.8
		全栈 AI 基础设施	百度 AI 大底座			
		AI 应用场景全覆盖的大模型	文心大模型（NLP、CV、跨模态、生物计算、及各种行业大模型）			
0020.HK	商汤-W	AI 基础设施	SenseCore 商汤 AI 大装置	1,004	-5.2	19.2
		CV（2D/3D 关键点捕捉；3D 关键点驱动；肖像风格化；图像/视频编辑）；虚拟穿戴；数字人	-			
		AI 实验室覆盖 CV、NLP、机器学习、强化学习·等方面研究	-			
0700.HK	腾讯控股	AI 平台服务	腾讯云 TI 平台：一站式机器学习生态服务平台，涵盖数据预处理、模型训练、评估、预测全流程； 人工智能服务平台：连接云边缘设备、算法和数据，提供全栈式 AI 应用开发、部署、管理等功能； 腾讯智能对话平台：开放腾讯核心智能对话系统技术，提供人机交互应用开发最佳体验； IP 虚拟人：提供 IP 角色授权与多模态交互方案，打造“能听会说”的 AI 智能形象； 音视频字幕平台：为企业和个人提供多语言音视频字幕转写、翻译、编辑与压制一站式服务	36,016	18.4	5.9
0772.HK	阅文集团	NLP	联合微软(亚洲)互联网工程院开启 AI 赋能网络文学	403	34.7	4.1
1357.HK	美图公司	CV	美图秀秀上新 AI 绘画功能	108	-50.9	5.0
A 股公司						
证券代码	证券简称	主要研究领域	相关应用/布局	总市值 (亿元)	PE (TTM)	PS (TTM)
300624.SZ	万兴科技	CV	旗下首款 AIGC 产品“万兴 AI 绘画”将开启公测；Filmora、万兴喵影、Pixso	55	180.6	4.9
300229.SZ	拓尔思	NLP	TRS 小思智能问答机器人；语义智能技术平台“智拓”	147	58.9	13.9
300418.SZ	昆仑万维	图像、音乐、文本、编程等多模态内容生成能力	“昆仑天工”旗下模型包括天工巧绘 SkyPaint、天工乐府 SkyMusic、天工妙笔 SkyText、天工智码 SkyCode	232	263.6	5.0
000681.SZ	视觉中国	CV	发售数轮 AIGC 数字藏品；与百度旗下的 AI 作画平台文心一格达成战略合作协议、赋能创作者和版权保护。AIGC 方式创造图片内容，平台场景逻辑顺，且仍在大力投入 AI 布局。	118	142.9	17.8
300364.SZ	中文在线	NLP、CV 等	AI 文字辅助创作功能已推出，并在 17k 小说网面向所有作者开放；AI 辅助绘画功能已推出。“AI 主播”实现低成本 2 高效优质的内容生产，通过 AI 驱动复活 3D 模型虚拟数字人。	89	-108.6	7.1
301270.SZ	汉仪股份	字体生成	AI 造字	39	55.9	17.1
002354.SZ	天娱数科	多模态智能算法等	升级发布“MetaSurfing-元享智能云平台”，在原有功能基础上纳入 AIGC 功能模块	92	-834.8	4.8
603466.SH	风语筑	CV；数字人等	公司在虚拟展厅、虚拟数字人、虚拟场景以及数字艺术创作领域均有结合 AI 人工智能技术的探索与应用，有效提升了数字创意、三维设计与 3D 内容创作的效率和独特性。现阶段，上述探索仍集中在 AI 生成艺术、AI 辅助内容创意与生产等领域	89	147.3	4.5

资料来源：各公司官网，各公司公告，Wind，同花顺等，德邦研究所 注：CV（Computer Vision，计算机视觉）

## 7. 风险提示

技术发展不及预期、监管政策变化、知识产权相关问题等。

# 信息披露

## 分析师与研究助理简介

马笑，华中科技大学硕士，2022年加入德邦证券，传媒互联网&海外首席分析师，行业全覆盖。5年多二级研究经验，2年产业战略/投资/咨询经验。曾任新时代证券TMT组长，传媒/计算机首席分析师，此前担任过管理咨询顾问/战略研究员/投资经理等岗位。2018年东方财富百佳分析师传媒团队第一名；2020年wind金牌分析师。

## 分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度，独立、客观地出具本报告。本报告所采用的数据和信息均来自市场公开信息，本人不保证该等信息的准确性或完整性。分析逻辑基于作者的职业理解，清晰准确地反映了作者的研究观点，结论不受任何第三方的授意或影响，特此声明。

## 投资评级说明

	类别	评级	说明
<b>1. 投资评级的比较和评级标准：</b> 以报告发布后的6个月内的市场表现为比较标准，报告发布日后6个月内的公司股价（或行业指数）的涨跌幅相对同期市场基准指数的涨跌幅；	股票投资评级	买入	相对强于市场表现 20%以上；
		增持	相对强于市场表现 5%~20%；
		中性	相对市场表现在-5%~+5%之间波动；
		减持	相对弱于市场表现 5%以下。
<b>2. 市场基准指数的比较标准：</b> A股市场以上证综指或深证成指为基准；香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	行业投资评级	优于大市	预期行业整体回报高于基准指数整体水平 10%以上；
		中性	预期行业整体回报介于基准指数整体水平-10%与 10%之间；
		弱于大市	预期行业整体回报低于基准指数整体水平 10%以下。

## 法律声明

本报告仅供德邦证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

市场有风险，投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考，不构成投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。在法律许可的情况下，德邦证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经德邦证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容，务必联络德邦证券研究所并获得许可，并需注明出处为德邦证券研究所，且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可，德邦证券股份有限公司的经营范围包括证券投资咨询业务。