

2023 年 02 月 06 日

看好

相关研究

"计算机主线会持续吗？AIGC FAQ 几大问题！- 计算机行业周报 20230130-20230203" 2023 年 2 月 4 日

"数字经济+chatGPT+港股科技，均预示科技全面机会！-计算机行业周报 20230116-20230120" 2023 年 1 月 29 日

证券分析师

洪依真 A0230519060003
hongyz@swsresearch.com
林起贤 A0230519060002
linqx@swsresearch.com
刘洋 A0230513050006
liuyang2@swsresearch.com

联系人

洪依真
(8621)23297818x
hongyz@swsresearch.com



申万宏源研究微信服务号

ChatGPT/AIGC 九问九答

——AI 行业系列点评

本期投资提示：

- 近期市场对于 AIGC/ChatGPT 关注较高，投资者常见以下问题。
- **Q：大规模预训练模型（大模型）与小模型的核心区别？相比小模型，大模型有什么优势？** 首先最核心区别在于参数量，例如 GPT-3 参数量达到了 1,750 亿个，是传统深度学习小模型参数的至少一万倍以上。通过模型的扩大，可以带来提升包括：1) GPT-2 等大模型舍弃了小模型中常见的调参过程，转向无监督训练；2) 在 NLP 中常见小模型需要对不同的任务使用不同模型分别训练解决，而 GPT-3 等大规模预训练模型不再规定任务，而是对不同任务都有较好效果；3) GPT-3 先使用海量数据预训练大模型，得到一套模型参数，然后用这套参数对模型进行初始化，再进行训练。大幅降低后续对数据量的需求。
- **Q：GPT-3 等大模型和 Transformer 的关联？** GPT/Bert 等大模型都在自然语言领域应用。而 Transformer 是目前 NLP 领域效果较好的深度学习模型，因此 GPT/Bert 等都使用了 Transformer。Transformer 最大的特点是，可以让每个单元都可以捕捉到整句的信息，这也是 RNN/LSTM 等更早期 NLP 模型无法做到的一点。
- **Q：Transformer 对于 CV 机器视觉可以有类似 NLP 的提升吗？** 实际上 2017 年 transformer 推出之后，2018 年就有学者将其应用于 CV 图像领域，但当时应用效果未发现相比传统 CNN 等算法有较大提升。但在 2021 年后，随着 swin transformer、mae 等论文推出，transformer 在 CV 和跨模态等应用增加。因此国内擅长 CV 的 AI 公司，在 2022 年前对于基于 transformer 的大模型关注投入较少，但是我们认为随着 MIT 等效果提升，应用类的 CV 公司也会开始关注并投入大模型。
- **Q：国内 transformer 大模型做的怎样？如何看待中美技术差距。** 首先需要承认这一类大模型底层创新在美国开始。中国在 2021 年后加速追赶。此前国内大部分视觉类 AI 公司都以小模型路线为主，投入大模型研发较多的包括百度、华为等。百度推出文心大模型，基本实现跨模态和多下游应用。华为推出最大中文语言预训练模型盘古大模型。
- **Q：怎样的公司有机会参与大模型底层技术研发？** 我们认为大模型对存储、算力要求极高，且对于开发者对深度学习框架使用也有要求，普通机构难以复现。
- **Q：上市公司中有哪些机会呢？** **计算机：**泛 AI 领域是相关主线。稳健标的如 AI 领军、科大讯飞、万兴科技、虹软科技，弹性标的为云从科技、汉王科技、博彦科技、商汤科技(H)。**互联网：**看好国内 AIGC 市场，持续跟进 AIGC 的技术发展和文本、音频、图像等单模态的应用落地。重点关注技术驱动的初创代表企业、科研院所，以及与应用场景关联的 A 股上市公司。相关标的：中文在线、汉仪股份、视觉中国、蓝色光标。
- **风险提示：**大模型技术中美仍存在差异，部分技术尚处于早期实验室阶段，存在落地风险；实体清单等可能对训练硬件产生影响。

近期市场对于 AIGC/ChatGPT 关注较高，投资者常见以下问题。

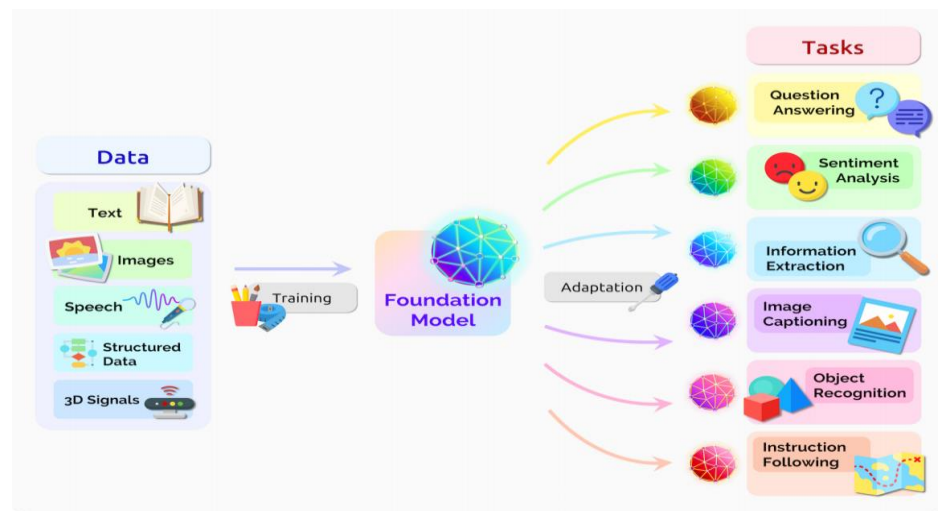
Q1：大规模预训练模型（大模型）与小模型的核心区别？相比小模型，大模型有什么优势？

首先最核心区别在于参数量，例如 GPT-3 参数量达到了 1,750 亿个，是传统深度学习小模型参数的至少一万倍以上。通过模型的扩大，可以带来提升包括：

1) GPT-2 等大模型舍弃了小模型中常见的调参 Fine-Tuning 过程，转向容量更大、**无监督训练**。

2) 在自然语言理解 NLP 中，常见任务包括翻译、问答、文本填空等，常见小模型需要对不同的任务使用不同模型分别训练解决 而 GPT-3 等大规模预训练模型**不再规定任务**，而是对以上不同任务都有较好效果；可以理解为，一个饱读诗书的人，应该会写作、对话、阅读理解等多种技能，而不需要对单独技能训练；

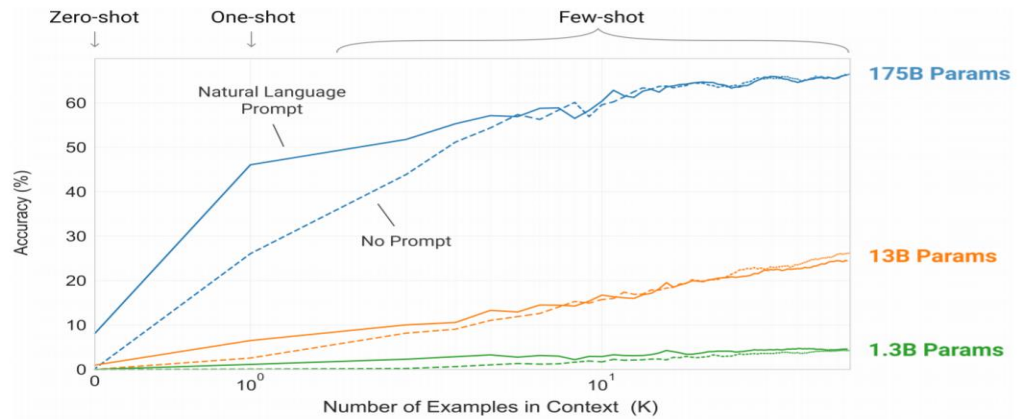
图 1 预训练大模型在综合大量信息后可以用于多范围任务



资料来源：《On the Opportunities and Risks of Foundation Models》、申万宏源研究

3) 传统的模型训练方式是反向传播算法，先对网络中的参数进行随机初始化（预训练大模型中不是随机初始化的），再利用随机梯度下降等优化算法不断优化模型参数，这种方式下对数据需求量较大。GPT-3 先使用海量数据预训练大模型，得到一套模型参数，然后用这套参数对模型进行初始化，再进行训练。**大幅降低后续对数据量的需求。**

图 2 参数量较大的模型在即使在 Few-shot 下也有很好的表现



资料来源：《Language Models are Few-Shot Learners》，申万宏源研究

Q2：因此大规模预训练思路对 AI 应用落地有什么帮助？

小模型时代，商用模型开发会综合考虑调参等研发投入量和带来的商业价值，模型开发出来后能否复用以摊薄研发成本，同时对于部分训练样本量较少的场景，没有很好解决方法。

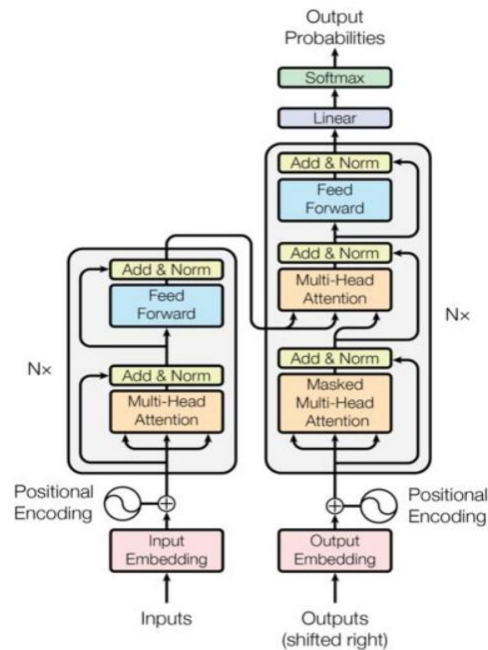
大模型可以在长尾场景应用落地，降低训练成本、减少数据要求。基于大规模预训练思路，一方面大模型可以应对多个泛化任务，大模型+细分场景微调，更适合长尾落地；另一方面，对于小样本（few shot）训练，大模型也有较好提升。

Q3：GPT-3 等大模型和 Transformer 的关联？

GPT/Bert 等大模型都在自然语言领域应用。而 Transformer 是目前 NLP 领域效果较好的深度学习模型，因此 GPT/Bert 等都使用了 Transformer。

此前 NLP 模型问题是：例如 RNN、LSTM，基本做法是把一段话拆分成多个单元，然后按前后顺序依次输入。输入的 LSTM 或 RNN 模块可以根据训练中句子的长度进行调整，较短的句子可以通过占位符进行补齐。也就是说当输入第 n 个词的时候，机器仅有前 n-1 个词的信息。即实际训练无法充分发挥这些输入之间的关系而导致模型训练结果效果极差。即使尝试用 CNN 解决这个问题，但需要叠加很多层，才能捕捉长远的信息。

图 3 Transformer 的详细结构



资料来源：《Attention Is All You Need》、申万宏源研究

Transformer 引入了 Self-Attention 自注意力机制：让机器注意到整个输入中不同部分之间的相关性。Transformer 最大的特点是，可以让每个单元都可以捕捉到整句的信息，这也是 RNN/LSTM 等更早期 NLP 模型无法做到的一点。

Q4：那 Transformer 对于 CV 机器视觉可以有类似 NLP 的提升吗？

实际上 2017 年 transformer 推出之后，2018 年就有学者将其应用于 CV 图像领域，但当时应用效果未发现相比传统 CNN 等算法有较大提升。当时常见做法是将自注意力和 CNN 一起使用，而原有结构基本不变，对于传统卷积神经网络仍有较高依赖。

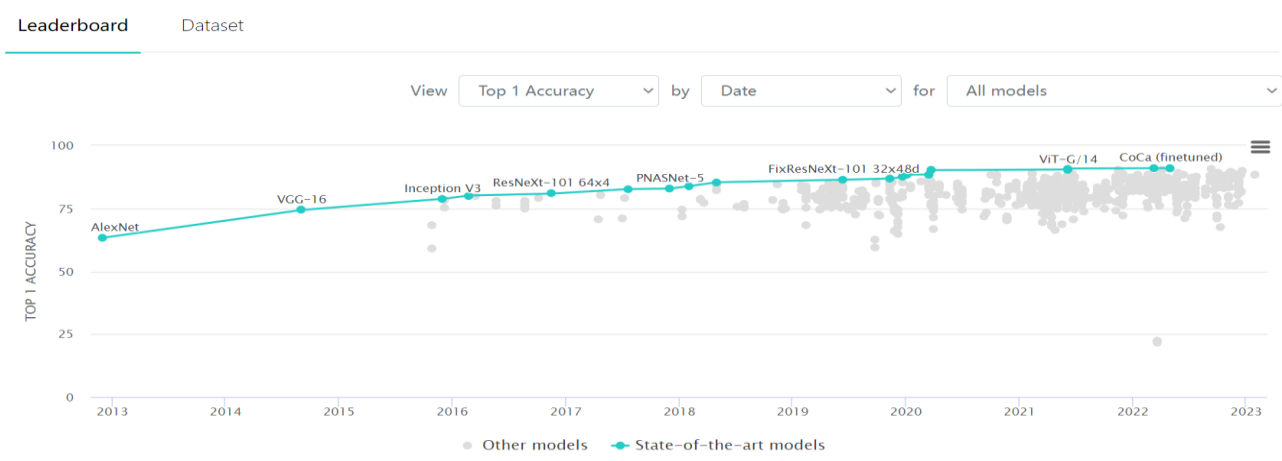
Transformers 用在 CV 上的难点。对于一个 Transformer 的 encoder，输入元素是一个个的单词，输出是一些元素。Transformers 自注意力操作中，每个元素需要和每个元素互相进行两两互动，最后算得一个 Attention 自注意力图，通过加权平均，最后得到输出。由于单元间两两互动，因此复杂度是序列长度的平方，也就是 n^2 ，在自然语言处理中，一个句子的长度一般就是几百或上千，例如 BERT 为 512 序列长度。

而对于一个比较常见的 224×224 大小的图片，如果直接以每一个像素点为单位进行两两互动的話，那序列长度为 50176，大概为 BERT 序列长度的 100 倍，因此复杂度极高。

但在 2021 年后，随着 swin transformer、mae 等论文推出，transformer 在 CV 和跨模态等应用增加。

图 4 图像分类任务中使用 Vision Transformer 取得了较好的结果

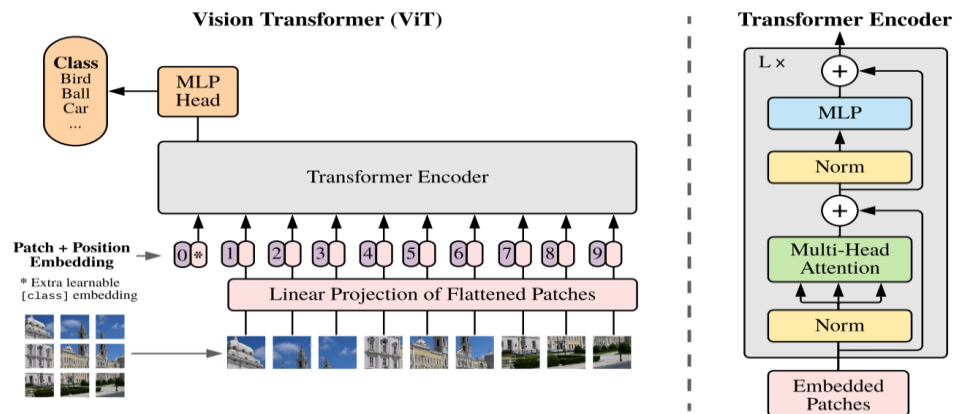
Image Classification on ImageNet



资料来源：paperswithcode.com、申万宏源研究

例如 2021 年 google 将尝试将自注意力机制和 transformer 直接应用在图像领域。在不依赖 CNN 结构的情况下，如何尽可能地讲 NLP 领域的标配——Transformer 不做修改的迁移到 CV 领域，并取得了较好的效果。

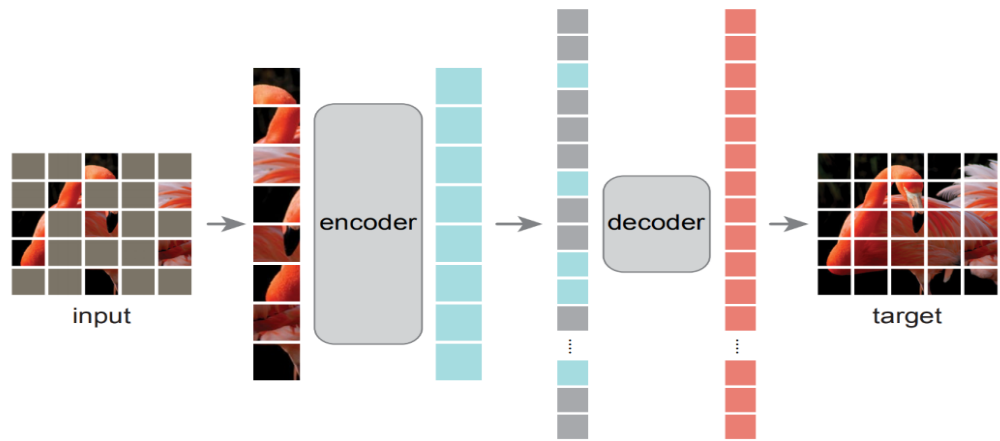
图 5 2021 年 google 的 Vision Transformer 论文的核心思路



资料来源：《AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE》、申万宏源研究

例如 2022 年引入的 MAE 方法。masked autoencoders(MAE)方法是随机 mask (遮罩)掉一部分像素，能极大加速模型的训练速度并提升准确率。由于语言的信息密度更高，mask 部分文字可能使得语义完全不同；而图像的冗余度就很高，因此 mask 掉部分信息可以大大减少冗余。

图 6 MAE 的核心思路



资料来源：《Masked Autoencoders Are Scalable Vision Learners》、申万宏源研究

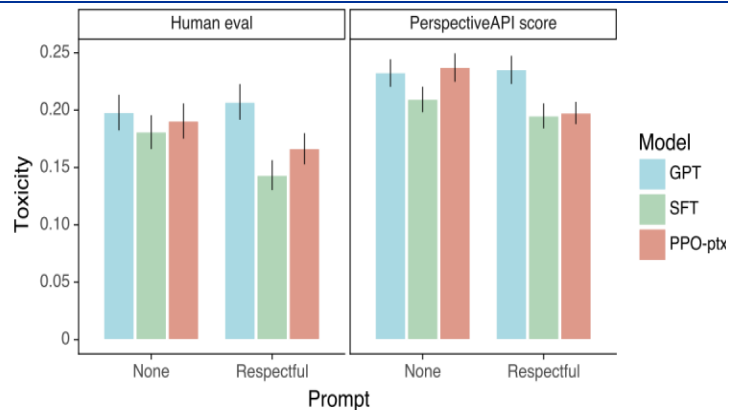
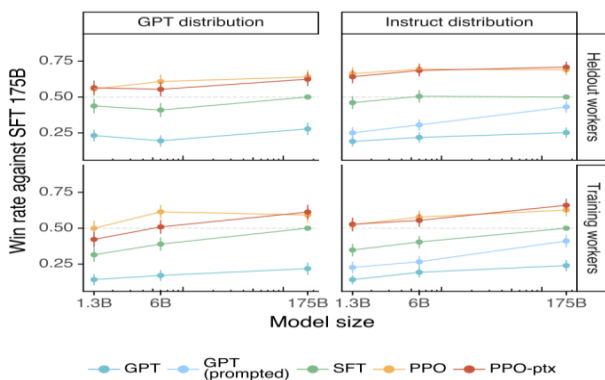
因此，国内擅长 CV 的 AI 公司，在 2022 年前对于基于 transformer 的大模型关注投入较少，但是我们认为随着 VIT 等效果提升，应用类的 CV 公司也会开始关注并投入大模型。

Q5：为何 GPT-3 在 2020 年就推出了，而资本市场近期才开始对大模型高关注？

2020 年的 GPT-3 仍有较多缺陷 其中无法商用的最大问题是输出结果有毒或不真实，OpenAI 2022 年 1 月正式公开了改良版 InstructGPT。比 GPT-3 小了 100 多倍，仅有 13 亿个参数。OpenAI 使用了一种通过人类反馈来强化学习 (RLHF) 的现有技术。OpenAI 根据用户和 API 的交互结果，对模型的多个输出进行了排名，然后再利用这些数据微调 GPT-3，大幅减少了有毒结果的输出。

图 7 InstructGPT 的输出结果比 GPT-3 提升

图 8 GPT-3 最被诟病的毒的结果 (RealToxicityPrompts) 在 InstructGPT 中减少



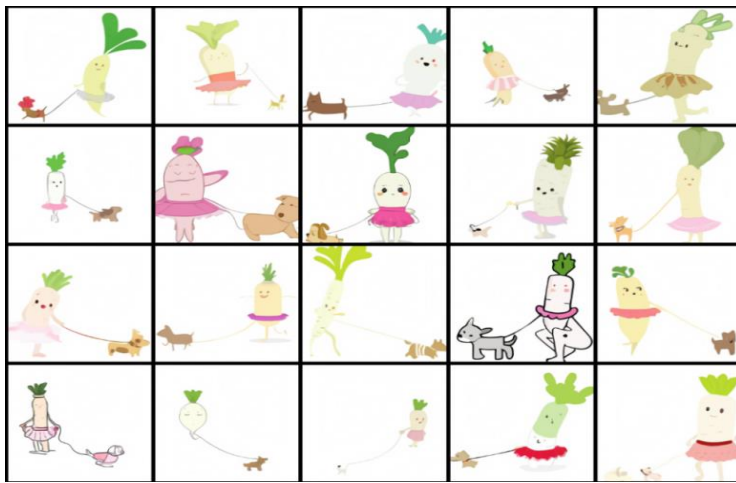
资料来源：《Training language models to follow instructions with human feedback》、申万宏源研究

资料来源：《Training language models to follow instructions with human feedback》、申万宏源研究

因此，我们现在看到的 ChatGPT 可以优化负面结果，比如提问“如何闯空门”，机器会回答“这并不合适”，同时连贯对话也更加流畅，达到了可以商用的效果。

同样的情况在 AIGC 图片生成领域也存在。2021 年 OPEN AI 就推出了图片生成工具包 DALL-E，但是此时图像像素仍然较低。2022 年后，OPEN AI 进一步引入 Diffusion 等方式，使得图片像素提升 4 倍以上，且同时模型真实度可以维持，并在此基础上推出全新工具包 DALL-E2。此时较高像素下生成的图片可以开始达到商用的效果。

图 9 DALL·E 设计“一颗白菜穿着芭蕾舞裙在遛狗”



资料来源：申万宏源研究

图 10 DALL·E-2 设计“一碗汤是另一个次元的入口”



资料来源：申万宏源研究

Q6：为何也有人认为 GPT 等大模型并未对 NLP 做出颠覆式创新？

传统 NLP 最大缺陷，即自然语言理解、缺乏逻辑等问题，在 GPT 中仍未被解决。GPT 等模型依赖于大规模离线收集的语料进行训练，但是对于在线提供的即时信息，往往并不能进行充分理解并且纳入自己的回复中。更进一步的，**对于对话中提供的因果关系，ChatGPT 也往往缺乏很好的理解能力。**

例如提问：韩梅梅喜欢画画，她的儿子也是。李华是韩梅梅唯一的儿子。所以李华喜欢画画吗？ChatGPT：不太清楚李华是否喜欢画画。

此外，GPT 会给出部分问题貌似合理但不正确或荒谬的答案，或者过度使用一些词句和特定表达，同时缺乏类似人类的举一反三的能力等。

Q7：国内 transformer 大模型做的怎样？如何看待中美技术差距。

首先，需要承认这一类大模型底层创新在美国开始，包括 2017 年 Google《Attention Is All You Need》首次提出了 NLP 模型 Transformer，OpenAI 的 GPT-3 等。中国在 2021 年后加速追赶，尤其是在 CV 大模型和多模态融合。

据国内 OpenBMB 开源社区统计，目前，在全球超千亿参数的大模型中，中国企业或机构贡献了 1/3，美国贡献了 1/2，世界其他国家或组织贡献了剩下的 1/6。

此前国内大部分视觉类 AI 公司都以小模型路线为主 投入大模型研发较多的包括百度、华为等，优势可能会在多模态融合。

图 11 大模型/小模型两类公司路线区别



资料来源：申万宏源研究

百度文心大模型：NLP+CV，基本实现跨模态和多下游应用。应用场景涵盖：NLP 大模型、CV 大模型、跨模态大模型、生物计算大模型、行业大模型。API 包括了：1) ERNIE 3.0 文本理解与创作：预置作文生成、文案创作、情感分析等任务提示，2) ERNIE-ViLG AI 作画大模型，3) PLATO 大模型开放域对话服务。

图 12 百度文心大模型和多行业工具平台

产品与社区	文心一格 AI 艺术和创作辅助平台	文心百中 大模型驱动的产业级搜索系统	赐谷社区 大模型创作与探索社区
工具与平台	EasyDL-大模型 零门槛 AI 开发平台	BML-大模型 多功能 AI 开发平台	大模型 API
	数据标注与处理	大模型微调	大模型套件 大模型压缩 高性能部署 场景化工具
	行业大模型		
	国网-百度文心	浦发-百度文心	航天-百度文心
	深燃-百度文心	吉利-百度文心	泰康-百度文心
	人民网-百度文心	冰城-百度文心	电影频道-百度文心
	NLP 大模型	CV 大模型	跨模态大模型
文心大模型	医疗 ERNIE-Health 对话 PLATO 跨语言 ERNIE-M ERINIE 3.0 Tiny (百亿级) ERINIE 3.0 (百亿级)	金融 ERNIE-Finance 搜索 ERNIE-Search 代码 ERNIE-Code 网络 ERNIE-Sage 信息抽取 ERNIE-UIE OCR 图像表征学习 VIMER-StrucTexT 多任务视觉表征学习 VIMER-UFO 视觉处理 多任务学习 VIMER-TCIR 自监督视觉表征学习 VIMER-CAE	文图生成 ERNIE-ViLG 文档智能 ERNIE-Layout 跨模态大模型 视觉-语言 ERNIE-ViL 语音-语言 ERNIE-SAT 地理-语言 ERNIE-GeoL
	生物计算大模型 化合物表征学习 HelixGEM 蛋白质结构预测 HelixFold 单序列蛋白质结构预测 HelixFold-Single		

资料来源：百度文心官网，申万宏源研究

华为盘古大模型：最大中文语言预训练模型。2021 年 4 月华为发布千亿参数 40TB 训练数据的全球最大中文语言 (NLP) 预训练模型，30 亿参数的全球最大视觉 (CV) 预训练模型。将 P-tuning、priming 等最新技术融入到盘古的微调框架中，提升微调效果；在样本极少的情况下，盘古的少样本学习能力远超 GPT 系列和 BERT 系列；要得到相同的 F1 结果，盘古所需的数据量仅为中文 GPT-3 的 1/9，实现了近 10 倍的生产效率提升。

Q8：怎样的公司有机会参与大模型底层技术研发？

我们认为大模型对存储、算力要求极高，且对于开发者对深度学习框架使用也有要求，普通机构难以复现。

据 NVIDIA 估算,如果要训练 GPT-3 ,即使单个机器的显存 / 内存能装得下,用 8 张 V100 的显卡,训练时长预计要 36 年;即使用 512 张 V100 ,训练也需要将近 7 个月;如果拥有 1024 张 80GB A100 ,那么完整训练 GPT-3 的时长可以缩减到 1 个月。

以微软与 OpenAI 合作建造的 Azure A 工智能算力平台为例,该算力平台投资约 10 亿美元,使用该超算中心训练一次超大模型 GPT-3 大约花费 1200 万美元。

同时解决分布式训练问题:上百台服务器之间的通信、拓扑、模型并行、流水并行等问题,模型训练是显存峰值问题。

因此 GPT-3 发布一年后,只有 NVIDIA 、微软等大企业可以复现。

Q9: 如果底层算法进入壁垒很高,目前看只有互联网巨头领先,那么上市公司中有哪些机会呢?

计算机: 泛 AI 领域是相关主线。稳健标的如 AI 领军、科大讯飞、万兴科技、虹软科技,弹性标的为云从科技(上证 e 互动平台公开披露了视觉语音 NLP 都类似 chatGPT 的“预训练模型+反馈调优”路线)、汉王科技(互动易平台公开披露了 NLP 技术积累和项目)、博彦科技、商汤科技(H)。当前,结合相关主线和已有涨幅,格灵深瞳、虹软科技、当虹科技等值得重视。

互联网: 看好国内 AIGC 市场,持续跟进 AIGC 的技术发展和文本、音频、图像等单模态的应用落地。重点关注技术驱动的初创代表企业、科研院所,以及与应用场景关联的 A 股上市公司。相关标的: 中文在线、汉仪股份、视觉中国、蓝色光标。此外,部分游戏公司也深度参与 AIGC 应用,腾讯控股 AI 领域深耕多年(优图等),内容领域龙头;网易旗下易智、伏羲、天音、有道智云等 AI 平台涵盖计算机视觉、数字孪生、人际交互、AI 音乐创作等方向;三七互娱率先应用智能化投放系统;吉比特 AI 等新方向持续探索跟踪,内容创新能力强;巨人网络拥有 AI 实验室,playtika 大数据和 AI 是其核心能力,心动公司 taptap 有望受益于 AIGC 实现内容提质增量。昆仑万维发布昆仑天工, AIGC 全系列算法与模型开源;神州泰岳 NLP 研发应用深耕多年,游戏出海排名前列。恺英网络 VR 等创新方向积极布局。

表 1: AIGC、AI、大模型相关业务的标的

公司名称	AIGC&AI 模式 应用领域	商业模式	主要进展
中文在线	主打文本	创作者辅助工具	AI 文字辅助创作功能已上线,并向公司 17K 文学平台全部作者开放;同时推出 AI 绘画功能。
汉仪股份	主打文本	形成 AIGC 字库技术储备阶段。	汉仪字库与阿里巴巴合作推出首款 AI 字体“阿里汉仪智能黑体”。
昆仑万维	主打音频	已经正式销售 AI 作曲,并与音乐公司达成歌曲代销、与车企达成 AIGC 战略合作、与游戏公司签约 BGM 等合作。	StarX 旗下音频社交领域产 StarMarker 搭建了 MusicX Lab 音乐实验室,聚焦 AI 音乐生成和 AI 作曲。发布「昆仑天工」AIGC 全系列算法与模型,并宣布模型开源。
视觉中国	主打图像	AIGC 作为创作者辅助工具	拥有高质量图片数据集,能通过 AIGC 生成艺术照片。视觉中国与百度旗下的 AI 作画平台文心一格将在创作者赋能和

公司名称	AIGC&AI 模式 应用领域	商业模式	主要进展
			版权保护等方面展开多项合作。
蓝色光标	主打图像	创作者辅助工具	AI 视频生成平台“蓝标分身”；AI 一键抽象生成智能画平台“创意画廊”，主打“康定斯基”风格创作。
万兴科技	主打图像	在单个微信用户 5 次免费关键词创作基础上，主要通过售卖额外创作次数变现。	旗下软件“万兴 AI 绘画”已在微信小程序公测。
科大讯飞	语音起点，图像/文本/语义等综合	AI 产品+解决方案+云平台	AI 教育领域成熟有收入利润，医疗、政府等 AI 收入高增。
虹软科技	主打图像图形	AI 软件产品+版权费模式	手机领域高市占率，汽车领域开始前装量产。
云从科技	图像和音频	NLP 投入阶段，有较多 AI 项目	上证 e 互动平台公开披露了视觉语音 NLP 都类似 chatGPT 的“预训练模型+反馈调优”路线。
汉王科技	主打图像	AI 一些 NLP 项目	互动易平台公开披露了 NLP 技术积累和项目。
商汤科技 (H)	语音/图像/文本/语义等综合	是大模型领军之一。打造的 AIGC 解决方案“商汤虚拟 IP 解决方案”拥有大规模高质量 3D 动作库，通过 AI 对大量真人动作数据进行深度学习，可以实现高效率、低成本的内容生成，	已经银行等场景落地
美图 (H)	图像	升级产品功能的辅助工具	旗下应用程序美图秀秀 (Meitu) 上新 AI 绘画功能
百度 (H)	语音/图像/文本/语义等综合	大模型领军。	推出文心一格，支持图片生成；升级“生成式搜索”能力；发布了新一代 AIGC 营销内容生产工具
20230201 新增扩散标的			
金山办公	文本/图像/语义	AI 嵌入办公软件，提供具备写作素材推荐、中文校对等语义理解功能以及格式美化等办公智能功能的 AI 助手，并以订阅形式收取增值服务费用	WPS 套件均已融入 AI 能力 (素材推荐、校对、排版、美化等)，智能写作研发项目已顺利完成，并搭建基于人工智能的文档创作服务平台
海天瑞声	语音/图像等	AI 训练数据专业提供商	为 AI 产业链上的各类机构提供 AI 算法模型开发训练所需的专业数据集。发行人所提供的训练数据覆盖智能语音 (语音识别、语音合成等)、计算机视觉、自然语言等多个 AI 核心领域
拓尔思	语义	AI 产品+解决方案	公司是国内最早从事自然语言处理 (NLP) 研发的企业之一，在 NLP、知识图谱、OCR、图像视频结构化领域都具备自主可控的底层技术，处于行业领先地位
当虹科技	视频	为电视台、新媒体、互联网、泛媒体等行业等客户提供通过 AI 技术进行短视频、长视频的生产/加工等解决方案	提供 AI 智能剪辑；用 AIGC 的方式自动生成对应短视频，极大程度上简化了流程提高了效率，保证了视频上线的时效性
二六三	语义	虚拟数字人应用了基于 AIGC 技术的 NLP 问答技术，业务尚处于研究和探索阶段	公司制作的虚拟数字人可在各类活动中担任虚拟主持人、AI 讲师、AI 客服等工作。公司制作的虚拟数字人在语音交互功能中应用了基于 AIGC 技术的 NLP 问答技术
同花顺	语义/文本等综合	AI 产品+解决方案+平台	同花顺 AI 开放平台，可面向客户提供数字虚拟人、智能金融问答、智能语音、智能客服机器人、智能质检机器人、会议转写系统、智慧政务平台、智能医疗辅助系统等 40 余项人工智能产品及服务，应用于证券、基金、银行、保险、运营商、高校、企业、科研院所、政府部门等多个行业
佳都科技	语义	公司投资云从科技 (计算机视觉)、思必驰	思必驰已经掌握全链路语音及语言交互技术，涵盖语音信号

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

文档名称：【申万宏源】AI行业系列点评：ChatGPT、AIGC九问九答.pdf

请登录 <https://shgis.com/post/1758.html> 下载完整文档。

手机端请扫码查看：

