

互联网行业专题报告：AI 大模型

ChatGPT 取得突破性进展，国内大模型争相发布

超配

核心观点

大模型提升机器理解能力，优化人机交互方式。 AI 大模型是实现通用人工智能（AGI）的重要方向，包含自然语言处理（NLP）、计算机视觉（CV），多模态大模型等。ChatGPT 推出两个月 MAU 突破 1 亿，是自然语言处理领域突破性的创新，大力出奇迹后出现涌现能力，更理解人类语言。大模型“预训练+精调”即可对下游应用赋能。我们认为大模型优化人类与机器交互方式，是效率的革命。大模型是“大数据+大算力+强算法”结合的产物，对比国内外大模型，算力储备上国内并无短板；算法上 OpenAI 有先发优势；前期数据的丰富度和量对大模型的训练至关重要。

百度文心大模型：包含 NLP、CV 等在内的系列大模型。 文心大模型包含 NLP 大模型、CV 大模型、跨模态大模型、生物计算大模型、行业大模型等。与 Bing 类似，文心一言有望优化 C 端用户搜索、创作体验；ToB 方面，百度已开放大模型 API 接口，在文案、AI 作画、开放域对话方面赋能企业。对于具体行业，百度推出文心行业大模型，以“行业知识增强”为核心特色。

阿里巴巴通义大模型：由通义-M6 模型融合语言模型和视觉模型组成，率先应用在硬件终端天猫精灵和软件通义千问。 通义大模型包括统一底座“M6-OFA”，三大通用模型“通义-M6”“通义-AliceMind”“通义-视觉大模型”，以及行业层面的不同垂直领域专业模型。在应用上，天猫精灵基于通义大模型推出拟声助手“鸟鸟分鸟”；对话式通义千问已经开始内测。

腾讯混元大模型：采用热启动降低训练成本，文字视频等多领域表现优异，已在广告游戏等多场景落地。 目前腾讯混元大模型已在腾讯广告、内容创作、游戏、对话式智能助手等方面实现落地，大幅提升工作效率并降低成本。

华为盘古大模型：基于 ModelArts 研发设计的系列模型，在物流、药物研发、气象预测等多领域已实现落地。 目前盘古 CV 大模型已覆盖了物流仓库监控等领域；NLP 大模型覆盖了智能文档检索、智能 ERP 和小语种大模型；科学计算大模型则应用于气象预报、海浪预测等方面。

字节跳动 AI 探索基础扎实，在语言大模型和图像大模型初步布局。 字节跳动 AI 大模型将从语言和图像两种模态发力，预期在今年年中推出大模型。字节跳动在算力、算法、数据方面并无短板。目前模型可用于图文、视频内容生成等，飞书将推出智能 AI 助手“My AI”。

投资建议：当前，各家公司开展新一轮 AI 模型竞赛，希望借助大模型开展“模型即服务”的范式。算力、算法、数据是大模型能否成功的重要因素，重点关注在大模型领域有技术储备的国内公司**腾讯控股、阿里巴巴、百度**等。

风险提示：AI 技术迭代进度不及预期；AI 应用落地进度不及预期；政策风险；供应链风险等。

重点公司盈利预测及投资评级

公司代码	公司名称	投资评级	昨收盘 (港元)	总市值 (百万, 港元)	EPS		PE	
					2023E	2024E	2023E	2024E
0700.HK	腾讯控股	买入	365.4	3,497,064	15.7	18.0	20	18
9988.HK	阿里巴巴-SW	买入	94.8	2,007,891	6.7	7.6	12	11
9888.HK	百度集团-SW	增持	129.6	362.437	7.9	9.3	14	12

资料来源：Wind、国信证券经济研究所预测

行业研究 · 海外市场专题

互联网 · 互联网 II

超配 · 首次评级

证券分析师：谢琦

021-60933157

xieqi2@guosen.com.cn

S0980520080008

联系人：王颖婕

021-60375476

wangyingjie1@guosen.com.cn

联系人：陈淑媛

021-60375431

chenshuyuan@guosen.com.cn

市场走势



资料来源：Wind、国信证券经济研究所整理

相关研究报告

《2023年3月电商数据：-消费复苏强劲，线上渗透率持续提升》

——2023-04-18

《互联网行业动态点评》——2023-03-16

《互联网行业动态点评-2022年12月电商数据：国内消费逐步回暖，零售线上化率进一步提升》——2023-01-31

《互联网行业动态点评-2022年11月电商数据：消费受疫情短期冲击明显，线上消费反弹快韧性强》——2022-12-21

《互联网行业动态点评-电商双11：行业增长趋缓，综合电商提升体验，直播电商继续抢占份额》——2022-11-24

内容目录

大模型概述：大模型提升机器理解能力，优化人机交互	5
AI 大模型是实现通用人工智能（AGI）的重要方向	5
大模型应用：“预训练+精调”即可对下游应用赋能，优化人类与机器交互方式	6
大模型三要素：算力、算法、数据	9
百度文心大模型	12
模型简述：已拥有 NLP、CV、跨模态、行业等多个 AI 大模型场景	12
模型 toC 应用：文心一言有望优化 C 端用户搜索、创作体验	13
模型 toB 应用：开放 API 接口赋能企业，精调行业模型	14
阿里巴巴通义大模型	16
模型简述：由通义-M6 模型融合语言模型和视觉模型组成	16
模型应用：率先应用在硬件终端天猫精灵和软件通义千问	17
腾讯混元大模型	18
模型简述：热启动降低训练成本，文字视频等多领域表现优异	18
模型应用：已在广告游戏多场景落地，大幅提升效率并降低成本	20
华为盘古大模型	22
模型简述：基于 ModelArts 研发设计的系列模型，NLP、CV 等方向性能优越	22
模型应用：在物流、药物研发、气象等多个场景实现落地，极大提高行业效率	23
字节跳动大模型	24
模型简述：AI 探索基础扎实，语言大模型和图像大模型初步布局	24
模型应用：可用于图文、视频内容生成等，飞书将推出智能 AI 助手“My AI”	25
投资建议	27
风险提示	27
免责声明	28

图表目录

图 1: 消费级应用达到一亿用户花费时长(月)	5
图 2: ChatGPT 对话演示	5
图 3: ChatGPT 能力实现解析	6
图 4: 训练大模型“预训练+精调”模式	6
图 5: OpenAI 宣布推出插件功能	7
图 6: Segment Anything Model (SAM) 演示	8
图 7: 大语言模型产品应用价值链与已渗透部分应用	8
图 8: 生成式 AI 发展历程与 ChatGPT 的突出能力	9
图 9: 主要数据集大小汇总 (部分, GB)	10
图 10: 百度文心全景图	12
图 11: 百度 NLP 大模型	13
图 12: 中国大模型市场 2022 年评估结果——百度	13
图 13: 文心一言对话演示	14
图 14: 百度文心大模型 API	14
图 15: 百度文心行业大模型原理	15
图 16: 文心行业大模型案例	16
图 17: 阿里巴巴通义大模型系列架构	16
图 18: 阿里巴巴通义-M6 发展历史	17
图 19: 阿里巴巴深度语言模型 AliceMind	17
图 20: 阿里巴巴通义-视觉大模型	17
图 21: 鸟鸟分鸟基于通义大模型系列	18
图 22: 鸟鸟分鸟 15 天训练流程	18
图 23: 阿里巴巴通义千问	18
图 24: 阿里巴巴通义千问百宝袋	18
图 25: HunYuan-tvr 在 5 个公开数据集上排名第一	19
图 26: HunYuan-NLP 1T 模型在 CLUE 总榜、分类榜和阅读理解榜登顶	19
图 27: HunYuan 大模型及解决方案	19
图 28: 腾讯广告多媒体 AI 技术研究与应用情况	20
图 29: 腾讯智能创作助手功能一览	20
图 30: 腾讯游戏 AI 路网生成模型	21
图 31: 腾讯混元助手项目组织架构	21

表1：大模型比较	11
表2：国内大模型概述	11
表3：盘古系列模型应用场景和领域	23
表4：相关公司盈利预测及估值	27

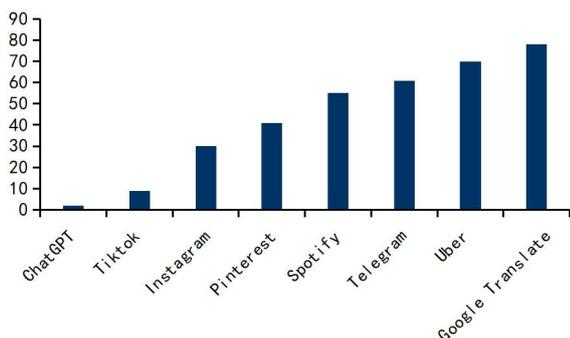
大模型概述：大模型提升机器理解能力，优化人机交互

AI 大模型是实现通用人工智能（AGI）的重要方向

AI 大模型是实现通用人工智能（AGI）的重要方向。AI 大模型是基于海量多源数据打造的模型。AI 大模型具备通用、可规模化复制等诸多优势，是实现 AGI（通用人工智能）的重要方向。AI 大模型当前包含自然语言处理(NLP)、计算机视觉（CV）等，统一整合的多模态大模型等。

ChatGPT 推出两个月 MAU 突破 1 亿，是自然语言处理领域突破性的创新，进一步理解了人类语言。大语言模型（Large Language Model, LLM）是一种使用了大量数据训练的深度学习算法，构建出一个能够理解人类语言并自动生成语言的模型。ChatGPT 在 2022 年 11 月底推出后，2023 年 1 月的月活跃用户数已达 1 亿，成为历史上用户增长最快的应用。ChatGPT 的自然语言处理能力超越了以往的自然语言处理模型，可以应对各种自然语言处理任务，包括机器翻译、问答、文本生成等。

图1：消费级应用达到一亿用户花费时长(月)



资料来源：Yahoo Finance，国信证券经济研究所整理

图2：ChatGPT 对话演示



资料来源：ChatGPT，国信证券经济研究所整理

ChatGPT 由 OpenAI 研发，GPT 模型已更新至 GPT-4，大力出奇迹后出现涌现能力。GPT 全称是 Generative Pre-trained Transformer，是 OpenAI 开发的一系列延伸自转换器架构（Transformer）的自然语言生成模型。2018 年，OpenAI 推出 GPT-1。2022 年 11 月，OpenAI 推出了对话交互式的 ChatGPT。ChatGPT 在 GPT-3.5 之上用基于人类反馈的监督和强化学习 (RLHF) 进行微调。人类反馈的引入，使机器更理解人类语言，让 GPT 获得更逼真的结果。ChatGPT 在逻辑推理、上下文理解等方面的能力，是模型规模达到量级后“涌现”出的。

图3: ChatGPT 能力实现解析

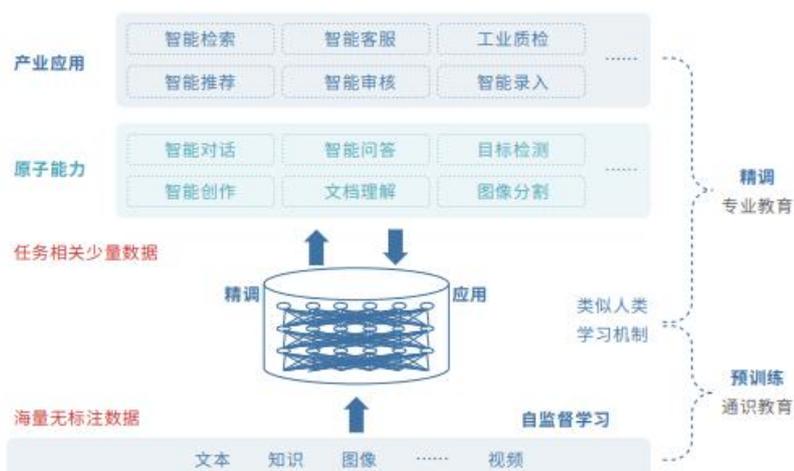


资料来源：《Natural Language Processing with Deep Learning, Jesse Mu》，艾瑞咨询，国信证券经济研究所整理

大模型应用：“预训练+精调”即可对下游应用赋能，优化人类与机器交互方式

大模型借助“预训练+精调”等模式，用比较大模型更少量的数据即可对下游应用赋能。预训练大模型基于海量数据的完成了“通识”教育。在具体应用场景下，借助“预训练+精调”等模式，应用模型用比较大模型更少量的数据即可进行相应微调，高水平完成细分应用的任务。

图4: 训练大模型“预训练+精调”模式



资料来源：IDC，百度，国信证券经济研究所整理

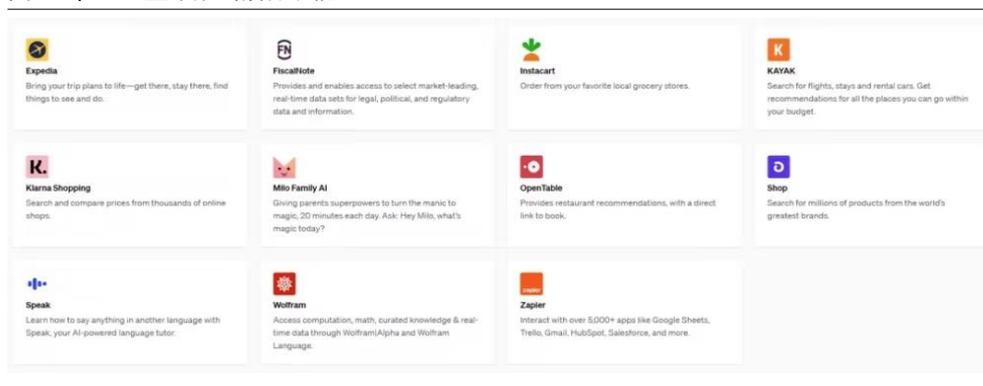
大模型增强机器理解力，优化人类与机器交互方式，提升信息处理效率。

1) 自然语言类大模型：人类可以用自然语言方式，与机器形成交互。ChatGPT 为自然语言方面的大模型应用，语言文本处理是人类世界最重要的工作内容之一，如果 AI 能

够懂的人类自然语言，并与人类形成交互，将协助人类大幅提升语言文本处理效率，带来生产力效率提升。例如，ChatGPT 可以依托人类自然问答的方式，帮助人类完成回答问题、虚拟助手（计划旅行、预定餐厅、购买产品等）、创作、文本、编程、摘要等工作。

我们也可以通过语言类大模型调用机器工具，大模型相当于人类助手。根据微软论文《Sparks of Artificial General Intelligence: Early experiments with GPT-4》，GPT-4 能够根据人类指令推断出需要哪些工具，有效地解析这些工具的输出并适当地做出回应，而无需任何专门的训练或微调。2023 年 3 月 23 日，OpenAI 宣布推出插件功能，帮助 ChatGPT 访问最新信息、进行计算或使用第三方服务。据 OpenAI 官网，第一批支持 ChatGPT 第三方插件的网站共 11 个，包含电商、预定航班、订购食材、检索实时信息、检索知识库信息等领域。

图5: OpenAI 宣布推出插件功能



资料来源：OpenAI 官网，国信证券经济研究所整理

2) **视觉类大模型**：与语言模型类似，视觉类模型需要完成对图片的理解。2023 年 4 月 6 日，MetaAI 在官网发布了基础模型 Segment Anything Model (SAM) 并开源，其本质是基于 Transform 模型架构，可以对图像中的一切对象进行分割。交互方面，SAM 可使用点击、框选、文字等各种输入提示，指定要在图像中分割的内容。精准分割为大模型图像理解的基础。图像理解可以在智能驾驶、图像识别、安防（人脸识别）等方面进行利用。

图6: Segment Anything Model (SAM) 演示

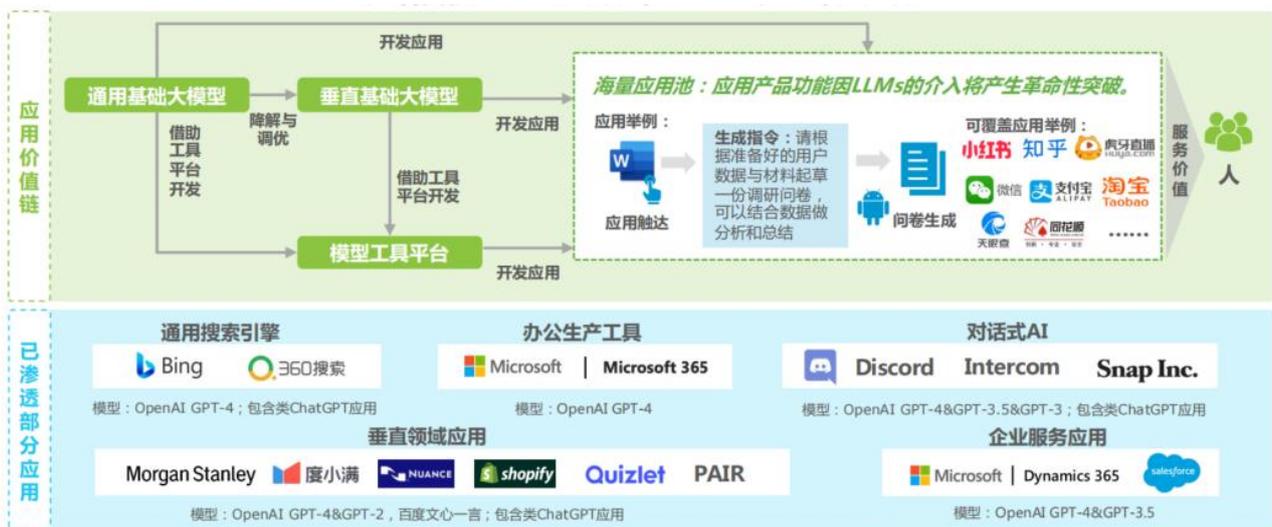


资料来源: Meta, 国信证券经济研究所整理

3) **多模态融合**: 多模态为语言、图片、音频等多个模态的感知和认知融合。多模态大模型能够让机器结合环境因素来模拟人与人之间的交互方式, 让图像、文本、语音等模态之间的统一表示和相互生成。多模态融合也是各大模型着重努力的方向。

当前, 已有的大模型可在文本、图像等方面赋能, 已有的渗透应用包括搜索引擎 (Bing 等)、办公工具 (Microsoft 365)、企业服务应用、垂直领域应用 (金融、电商等) 等等。根据 2023 阿里云峰会, 阿里巴巴集团董事会主席兼 CEO 张勇表示“所有行业、所有应用、所有软件、所有服务都值得基于新型人工智能技术、基于 AIGC 各方面技术支撑、大模型支撑重做一遍”。

图7: 大语言模型产品应用价值链与已渗透部分应用



资料来源: 艾瑞咨询, 国信证券经济研究所整理

大模型三要素：算力、算法、数据

大模型是“大数据+大算力+强算法”结合的产物。

1) **算力是 AI 发展的基础设施，芯片至关重要。**算力的大小代表着对数据处理能力的强弱。芯片性能越好，大模型的处理能力越快。比如，黄仁勋在 2023 年 2 月财报会中表示“过去十年，通过提出新处理器、新系统、新互连、新框架和算法，并与数据科学家、AI 研究人员合作开发新模型，已使大语言模型的处理速度提高了 100 万倍。”

2) **算法是 AI 解决问题的机制，源于算法理论发展、迭代优化。**不同算法可以看做解决问题的不同路径，算法的优劣可以用空间复杂度与时间复杂度来衡量。例如，GPT 是在 Transformer 模型基础上发展的，Transformer 由 GOOGLE 在 2017 年提出。Transformer 相比于传统的循环神经网络（RNN）或卷积神经网络（CNN），在处理长文本时具有更好的并行性和更短的训练时间。

图8：生成式 AI 发展历程与 ChatGPT 的突出能力



资料来源：艾瑞咨询，国信证券经济研究所整理

3) **数据是算法训练的养料，前期需要给模型喂养大量数据，形成模型理解能力，中后期数据质量决定了模型精度。**机器学习中要用标注好的数据进行训练，数据标注对未经处理的初级数据进行加工处理，转换为机器可识别信息，只有经过大量的训练，覆盖尽可能多的各种场景才能得到一个良好的模型。

- ✓ 当前，数据的丰富度和量对大模型的训练至关重要，只有大量数据的训练，大模型才有理解能力涌现的可能。当前训练数据集来源多为公开数据，比如根据 Alan D. Thompson 文章，列举的大模型的数据集包括维基百科、书籍、期刊、Reddit 链接、Common Crawl 和其他数据集等。
- ✓ 中后期，高质量数据将提升模型的精度。比如更加事实性的数据将提升模型准确性，更加通顺的中文语言将提升模型理解中文语言能力。另外，高质量反馈数据也可提高模型性能，比如 ChatGPT 采用人类强化学习 RLHF，通过更专业的问题、指令、人类反馈排序等加强模型理解人类语言逻辑。最后，也可以通过更精准的垂类数据，完成部分更细分领域模型搭建。

图9: 主要数据集大小汇总 (部分, GB)

	Wikipedia	Books	Journals	Reddit links	CC	Other	Total
GPT-1		4.6					4.6
GPT-2				40			40
GPT-3	11.4	21	101	50	570		753
The Pile v1	6	118	244	63	227	167	825
Megatron-11B	11.4	4.6		38	107		161
MT-NLG	6.4	118	77	63	983	127	1374
Gopher	12.5	2100	164.4		3450	4823	10550

资料来源: Alan D. Thompson 《What' s in my AI paper》, 国信证券经济研究所整理; 注: 确定的数据以斜体表示。仅原始训练数据集大小。

我们从以上三个维度, 分析当前发布的部分大模型:

1) 算力: 算力布局主要来源于芯片的采购布局, 算力基础设施的投入阻挡了部分小公司的入局。当前数据训练需要较高性能的芯片完成对整体模型神经网络的训练构建, 供应厂商包括英伟达等。应用层面只需调用整体大模型里的部分神经网络, 故对芯片要求没有训练模型时那么高。

2) 算法: OpenAI 仍有先发优势, GPT-3 之后并未开源。每家公司都有自己实现大模型的路径算法, 由于 GPT 在成立之初就坚持自己的算法路径 (采用单项 Transformer 架构等), 并在此基础上运用大量数据训练使得模型产生的涌现能力, 取得了一定的成功, 拥有先发优势。自 GPT-3 开始 (2020 年), OpenAI 便不再公布大模型相关训练及技术细节。虽然当前 OpenAI 的成功提供部分借鉴, 但其他公司仍需时间去研发追赶并进行模型迭代。

除此以外, 部分公司在特定领域拥有算法积累的先发优势, 比如搜索类公司对语料的切割和识别有先发优势, 视觉类公司在图像领域识别有先发优势。

3) 数据: 当前中文语料规模与质量上不如英文语料。根据 W3Techs, 中文网站占全球网站数 1.3%, 英文网站占比 63.6%。除此以外, 由于中国移动互联网发展较为成熟, 大量的中文数据资源被存于各家企业或机构里, 较难共享。

表1: 大模型比较

企业	模型	参数	算力	数据	RLHF	进度
OpenAI	GPT3	1750 亿	上万块 GPU (V100)	45TB 预训练	✓	22. 11. 30 日发布 ChatGPT
Google	LaMDA	1370 亿		1. 56TB	×	23. 3. 22 日发布 Bard
Meta	LLaMA	最大 650 亿	RSC 超算平台	1. 4T token	×	23. 2. 24 日发布 暂未公布产品
微软	Megatron-turing	5300 亿	Azure 云平台	3390 亿文本	×	
百度	文心大模型 (ERNIE 3. 0)	2600 亿	飞桨深度学习平台, 百度云	4TB (文本与知识图谱)	×	23. 3. 16 日发布文心一言, 23. 3. 27 日发布文心千帆大模型平台
阿里	M6	超过 10 亿	阿里云	1. 9TB 图像 292GB 文本	×	23. 4. 11 日发布通义千问
腾讯	混元	2000 亿	太极机器学习平台, 腾讯云	五大跨模态检索数据集	×	暂未公布产品
华为	盘古	200B	鹏程云脑、ModelArts	40TB	×	21. 4 发布盘古大模型

资料来源: OpenAI, Google, Meta, 微软, 文心大模型官网, 通义千问官网, 阿里巴巴官网, 腾讯太极机器学习平台, 华为云, 国信证券经济研究所整理

目前, 国内已有百度、阿里巴巴、腾讯、华为等公司对 AI 大模型进行开发, 各模型系列主要的 NLP 语言大模型、CV 大模型、多模态大模型也已推出并实现部分应用落地。此外, 据 36 氪消息, 字节跳动也在语言和图像大模型上有所布局, 预期将于今年年中推出字节自研大模型。具体来看:

- ✓ 百度在 AI 方面布局多年, 具有一定大模型先发优势。当前, 申请文心一言 API 调用服务测试的企业已突破 6. 5 万。在行业大模型上, 已经与国网、浦发、吉利、TCL、人民网、上海辞书出版社等均有案例应用。
- ✓ 阿里通义大模型在逻辑运算、编码能力、语音处理方面见长, 集团拥有丰富的生态和产品线, 在出行场景、办公场景、购物场景和生活场景均有广泛应用。
- ✓ 腾讯混元大模型已经在广告投放、游戏制作投入使用, 目前集团在研究对话式智能助手, 预计投入使用后将对 QQ 和微信生态有一定优化。
- ✓ 华为与 B 端合作紧密, 预计未来应用以 ToB 为主。此外, 华为在算法、算力上储备较为丰厚。比如“鹏城云脑 II”获全球 10500 排行五连冠, 拥有强大的 AI 算力和数据吞吐能力。华为云 ModelArts 平台的高效处理海量数据能力, 7 天完成了 40TB 文本数据处理。盘古大模型最早已经在 2021 年 4 月正式发布, 当前盘古大模型训练文本数据高达 40 TB (GPT-3 为 45 TB)。

表2: 国内大模型概述

公司	大模型	底层支持	模型系列	应用
百度	文心大模型	飞桨深度学习平台, 百度云	文心 NLP 大模型 (ERNIE 3. 0)、文心 CV 大模型、文心跨模态大模型、文心生物计算大模型	文心一言、开放 NLP 大模型 ERNIE3. 0/跨模态大模型 ERNIE-ViLG/对话大模型 PLATO 的 TOB API 接口, 赋能下游企业文案、AI 作画、开放域对话等方面
阿里巴巴	通义大模型	阿里云	通义-M6、通义-AliceMind、通义-视觉大模型	天猫精灵、通义千问
腾讯	混元大模型	腾讯太极机器学习平台, 腾讯云	混元 NLP 大模型、混元 CV 大模型、混元多模态大模型、混元文生图大模型	腾讯广告、文涌 (Effidit)、3D 虚拟场景自动生成、对话式智能助手
华为	盘古大模型	鹏程云脑、ModelArts	盘古 NLP 大模型、盘古 CV 大模型、盘古科学计算大模型	“物的银行”、新药研发、气象预测
字节跳动	字节跳动大模型	火山引擎	语言和图像大模型	抖音特效、剪映 AI 生成、飞书 AI 助手 “My AI”

资料来源: 文心大模型官网, 通义千问官网, 阿里巴巴官网, 腾讯太极机器学习平台, 鹅厂技术派, 华为云, 36 氪, 国信证券经济研究所整理

百度文心大模型

模型简述：已拥有 NLP、CV、跨模态、行业等多个 AI 大模型场景

文心大模型拥有多个 AI 应用场景，文心一言是基于文心 NLP 大模型的对话式产品。根据官网介绍，文心大模型包含 NLP 大模型、CV 大模型、跨模态大模型、生物计算大模型、行业大模型等，分别可用于语言、图像、跨模态、生物、细分行业等。2023 年 3 月 16 日，百度正式官宣全新的生成式 AI 对话应用——文心一言，是文心 NLP 模型应用下的产品。其能够与人对话互动，回答问题，协助创作。

图 10：百度文心全景图

产品与社区	文心一格 AI 艺术和创意辅助平台		文心百中 大模型驱动的产业级搜索系统		扬谷社区 大模型创意与探索社区			
工具与平台	EasyDL-大模型 零门槛 AI 开发平台		BML-大模型 全功能 AI 开发平台		大模型 API			
	大模型套件							
	数据标注与处理	大模型精调	大模型压缩	高性能部署	场景化工具			
文心大模型	行业大模型							
	国网-百度·文心	浦发-百度·文心	航天-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		
	深燃-百度·文心	吉利-百度·文心	泰康-百度·文心	TCL-百度·文心	辞海-百度·文心			
	NLP 大模型		CV 大模型		跨模态大模型		生物计算大模型	
	医疗 ERNIE-Health	金融 ERNIE-Finance		商品图文搜索表征学习 VIMER-UMS		文图生成 ERNIE-ViLG	文档智能 ERNIE-Layout	化合物表征学习 HelixGEM
	对话 PLATO	搜索 ERNIE-Search	信息抽取 ERNIE-UIE	OCR 图像表征学习 VIMER-StrucTexT		视觉-语言 ERNIE-ViL	语音-语言 ERNIE-SAT	地理-语言 ERNIE-Geol
跨语言 ERNIE-M	代码 ERNIE-Code	图网络 ERNIE-Sage	多任务视觉表征学习 VIMER-UFO					
语言理解与生成			视觉处理多任务学习 VIMER-TCIR	自监督视觉表征学习 VIMER-CAE			蛋白质结构预测 HelixFold	
ERNIE 3.0 Tiny (轻量级)	ERNIE 3.0 (百亿级)	鹏城-百度·文心 (千亿级)	ERNIE 3.0 Zeus (任务知识蒸馏)				单序列蛋白质结构预测 HelixFold-Single	

资料来源：文心大模型官网，国信证券经济研究所整理

百度 NLP 核心是 ERNIE 模型。文心 NLP 大模型面向语言理解、语言生成等 NLP 场景。百度文心 NLP 大模型打造了 ERNIE 模型系列，ERNIE 问世于 2019 年，目前发展至 3.0 版本。其中语言理解和生成包含 ERNIE 3.0 Tiny、ERNIE 3.0、鹏城-百度·文心、ERNIE 3.0 Zeus，另外文心 NLP 大模型包含医疗、金融等细分领域 ERNIE 模型。

- ✓ ERNIE 3.0 Tiny（轻量级）：可用于通用语言理解。
- ✓ ERNIE 3.0（百亿级）：可用于智能创作、摘要生成、问答、语义检索、情感分析、信息抽取、文本匹配、文本纠错等各类自然语言理解和生成任务。
- ✓ 鹏城-百度·文心（千亿级）：模型参数规模达到 2600 亿（GPT-3 1750 亿），可用于舆情分析、智能创作、文本解析等。

图 11：百度 NLP 大模型

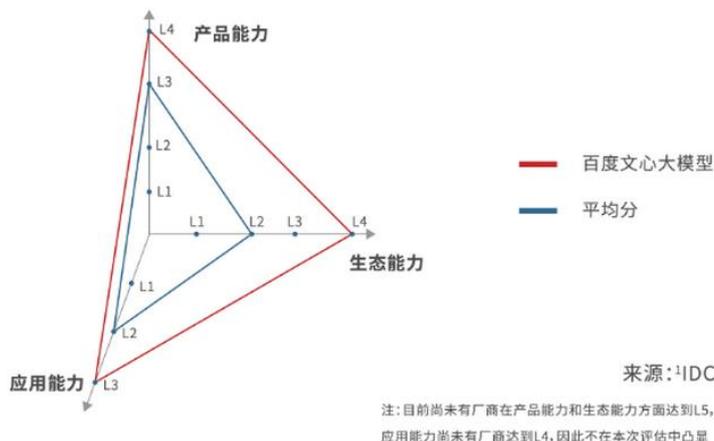


资料来源：文心大模型官网，国信证券经济研究所整理

文心大模型在国内市场格局中较为领先。根据 IDC 发布《2022 中国大模型发展白皮书》，百度文心大模型在市场格局中在产品能力、生态能力、应用能力等方面在国内较为领先。

文心一言核心优势是对中文的理解。百度作为中国语境的搜索龙头，拥有更多的中文语料数据参与训练。比如，文心一言可以针对“洛阳纸贵”这种容易产生歧义的成语，给出较为贴切的解释。可以用成语写出藏头诗，用四川话读出文章。

图 12：中国大模型市场 2022 年评估结果——百度



资料来源：IDC，国信证券经济研究所整理

模型 toC 应用：文心一言有望优化 C 端用户搜索、创作体验

百度已经推出对话式产品文心一言。文心一言可通过对话模式，帮助用户文学创作、商业文案创作、数理推算、中文理解、多模态生成等。例如文心一言发布会上案例，文心一言可以生成 2023 世界智能交通大会海报，解答“智能交通最适合

哪个城市发展”的问题，并用四川话和视频形式展现答案。我们认为，与 New Bing 类似，百度可以将对话式人工智能功能添加至搜索引擎中，优化用户搜索、创作等体验。当前，在百度 APP 内，基于文心 ERNIE 大模型已实现新闻资讯信息的自动分类，新闻标题和摘要的自动生成等功能。

图 13: 文心一言对话演示



资料来源：文心一言发布会，国信证券经济研究所整理

模型 toB 应用：开放 API 接口赋能企业，精调行业模型

目前，百度开放大模型 API 接口，在文案、AI 作画、开放域对话方面赋能开发者和企业。百度文心大模型开放了 NLP 大模型 ERNIE3.0、ERNIE-ViLG 文生图、对话大模型 PLATO。

- ✓ ERNIE 3.0 提供文案改写、开放问答、摘要、文案创作、小说创作、文本补全等文本理解与创作能力。
- ✓ ERNIE-ViLG 提供基于文本描述的 AI 作画能力。
- ✓ PLATO 提供生成式开放域对话服务，逻辑清晰、知识多元、情感丰富，闲聊能力接近真人水平。

当前已经有 650 家企业宣布接入文心一言生态，还有更多的企业在排队。据百度介绍，申请文心一言 API 调用服务测试的企业已突破 6.5 万。

图 14: 百度文心大模型 API



资料来源：文心大模型官网，国信证券经济研究所整理

对于具体行业来说，百度也推出文心·行业大模型，进行行业精调。行业大模型是在百度通用大模型的基础上学习行业特色数据与知识，建设行业 AI 基础设施。

图 15：百度文心行业大模型原理



资料来源：文心大模型官网，国信证券经济研究所整理

百度文心行业大模型以“行业知识增强”为核心特色。目前百度文心行业大模型在能源、金融、航天、制造、传媒、城市、社科以及影视等领域，与国网、浦发、吉利、TCL、人民网、上海辞书出版社等均有案例应用。我们以五个行业场景举例：

- ✓ 在燃气行业，“深燃-百度·文心”在环境巡检、安全监控等领域落地，有效解决了场景繁杂、识别困难等难题；
- ✓ 在汽车行业，“吉利-百度·文心”应用于售后服务和汽车领域知识库构建等场景；
- ✓ 在保险行业，“泰康-百度·文心”应用于医疗单据识别等任务，平均字段识别错误率相对下降；
- ✓ 在电子制造行业，“TCL-百度·文心”改变了传统质检算法流程，大幅提升了模型研发与运营效率，实现降本增效；
- ✓ 在社科行业，“辞海-百度·文心”用于版权保护、词条管理和知识挖掘等方面。

图 16: 文心行业大模型案例



资料来源: 文心大模型官网, 国信证券经济研究所整理

阿里巴巴通义大模型

模型简述: 由通义-M6 模型融合语言模型和视觉模型组成

通义大模型最早发布于 2022 年 9 月 2 日, 在阿里达摩院主办的世界人工智能大会“大规模预训练模型”主题论坛上。通义大模型实际上是由一系列模型组成, 包括统一底座“M6-0FA”, 三大通用模型“通义-M6”“通义-AliceMind”“通义-视觉大模型”, 以及行业层面的不同垂直领域专业模型。

图 17: 阿里巴巴通义大模型系列架构



资料来源: 机器之心, 国信证券经济研究所整理

通义系列前身为 M6, 是阿里达摩院研发的通用性多模态大模型, 特点是训练能耗显著降低。M6 自 2020 年初正式启动研发, 于 2021 年 1 月/3 月/5 月分别发布百亿/千亿/万亿参数多模态预训练模型。2021 年 10 月发布十万亿参数模型, 并且支持使用 512 GPU 在 10 天内即训练出具有可用水平的 10 万亿模型, 同等参数下相较于 GPT-3 能耗降低 99%。

图 18: 阿里巴巴通义-M6 发展历史



资料来源: 机器之心, 国信证券经济研究所整理

M6 融合语言模型和视觉模型, 形成当前的通义系列大模型。2022 年 1 月, 阿里 M6 发布了统一大模型 M6-OFA, 可以实现不同模型之间的模态表示、任务表示、模型结构的统一。以 M6-OFA 为统一底座, 阿里巴巴将另外两大模型“AliceMind”“视觉大模型”与 M6 进行融合, 形成了当前的通义系列架构。其中, **通义-AliceMind** 为深度语言模型体系, 包含了通用语言模型 StructBERT、生成式 PALM、结构化 StructuralLM、超大中文 PLUG、多模态 StructVBERT、多语言 VEC 等多个不同模型, 具有阅读、写作、翻译、问答、搜索、摘要生成、对话等多种能力;**通义-视觉大模型**包括文本到视觉生成、语言和视觉的模态映射两大基础模型, 可以提供文生图等能力。

图 19: 阿里巴巴深度语言模型 AliceMind



资料来源: AliceMind, 国信证券经济研究所整理

图 20: 阿里巴巴通义-视觉大模型



资料来源: 机器之心, 国信证券经济研究所整理

模型应用：率先应用在硬件终端天猫精灵和软件通义千问

硬件终端方面：天猫精灵基于通义大模型推出拟声助手“鸟鸟分鸟”。4月4日, 脱口秀演员鸟鸟在微博发布了一条关于天猫精灵的演示视频。根据公司和鸟鸟的介绍, “鸟鸟分鸟”是天猫精灵联合阿里达摩院, 基于通义大模型, 对鸟鸟录制的一个多小时音频进行学习训练而成。从形态上看, “鸟鸟分鸟”本身是一个对话机器人, 具备多轮对话能力, 并且还能利用搜索引擎等工具。个性化是“鸟鸟分鸟”最大特点, 具备拟人化的音色、语气、表达方式。我们推测, 这意味着通用大模型将有可能在表达方式层面实现较低成本的定制化。

图21: 鸟鸟分鸟基于通义大模型系列



资料来源: 机器之心, 国信证券经济研究所整理

图22: 鸟鸟分鸟 15 天训练流程



资料来源: 机器之心, 国信证券经济研究所整理

对话式通义千问已经开始内测, 包含 9 款“百宝袋”小应用。4 月 7 日, 阿里云宣布大模型“通义千问”上线, 并定向邀请企业用户进行测试。测试版的通义千问与 ChatGPT 和文心一言在界面上相似, 都是采用问答对话的形式。此外, 通义千问在对话模式之外, 制作了 9 款称为“百宝袋”的小应用, 可以分为三类: 1) 效率类: 写提纲、SWOT 分析、商品描述生成; 2) 生活类: 会放飞的菜谱、小学生作文、然后呢; 3) 娱乐类: 彩虹屁专家、写情书、为你写诗。我们认为, 未来阿里大模型有望添加至阿里系电商和钉钉等场景中, 带动用户体验和效率的提升。

图23: 阿里巴巴通义千问



资料来源: 虎嗅, 国信证券经济研究所整理

图24: 阿里巴巴通义千问百宝袋



资料来源: 虎嗅, 国信证券经济研究所整理

腾讯混元大模型

模型简述: 热启动降低训练成本, 文字视频等多领域表现优异

“混元” AI 大模型由 TEG 数据平台部和机器学习平台部联合主导, 涵盖计算机视觉、自然语言处理、多模态内容理解、文案生成、文生视频等方向。“混元” AI 大模型由腾讯 17 级研究员/杰出科学家(腾讯史上最高专业职级)张正友博士带队, 2022 年 4 月公司首次披露研发进展, 目前已更新至 2.0 版本。

性能上看, “混元”大模型在文字、视频等多个领域表现优异。其中 HunYuan-tvr 模型在文字和视频在全球最具权威的 MSR-VTT, MSVD, LSMDC, DiDeMo 和 ActivityNet 五大跨模态视频检索数据集榜单排名第一, 实现该领域的大满贯; HunYuan-NLP 1T 模型在 CLUE (中文语言理解评测集合) 总榜、分类榜和阅读理解榜排名第一。

图 25: HunYuan-tvr 在 5 个公开数据集上排名第一

Method	Hunyuan_tvr	DRL (阿里)	MDMMT-2 (华为)	腾讯PCG (CLIP2TV CLIP2video)	CAMoE (快手)	CLIP4Clip (微软)
MSR-VTT	55.0%	53.3%	48.8%	52.9%	48.8%	44.5%
MSVD	58.2%	50.0%	56.8%	47.0%	49.8%	46.2%
LSMDC	29.7%	26.5%	26.9%	-	25.9%	22.6%
DiDeMo	52.1%	49.0%	-	-	-	41.4%
ActivityNet	57.3%	46.2%	-	-	-	41.4%

资料来源：鹅厂技术派，国信证券经济研究所整理

图 26: HunYuan-NLP 1T 模型在 CLUE 总榜、分类榜和阅读理解榜登顶

排行	模型	研究机构	测评时间	Score1.1	认证	AFQMC	TNEWS1.1	IFLYTEK	OCNLI_50K	WSC1.1	CSL	CMRC2018	CHID1.1	C3 1.1
1	HunYuan-NLP 1T	腾讯混元AI大模型团队	22-11-26	86.918	待认证	85.11	70.44	67.54	86.5	96	96.2	87.9	98.848	93.723
2	遵义-AliceMind	达摩院NLP	22-11-22	86.685	待认证	84.07	73.47	67.42	85.87	94.33	95.03	86.8	99.208	93.969
3	HUMAN	CLUE	19-12-01	86.678	已认证	81	71	80.3	90.3	98	84	92.4	87.10	96.00
4	CHAOS	OPPO研究院融智团队	22-11-09	86.552	待认证	83.37	73.22	65.81	86.37	94.6	95.7	87.2	99.217	93.477
5	WenJin	Meituan NLP	22-10-20	86.313	待认证	84.49	73.04	64.38	86.23	94.44	95.67	86.25	98.898	93.415

资料来源：腾讯太极机器学习平台，国信证券经济研究所整理

技术上看，“混元”大模型基于腾讯太极机器学习平台研发，通过改进模型与算法，大幅降低训练成本。HunYuan 利用千亿小模型帮助万亿大模型完成启动，相较从 0 启动耗时短耗能低，仅用 256 卡，最快一天内即可完成万亿参数大模型 HunYuan-NLP 1T 的训练，整体训练成本仅为直接冷启动训练万亿模型的 1/8。

图 27: HunYuan 大模型及解决方案

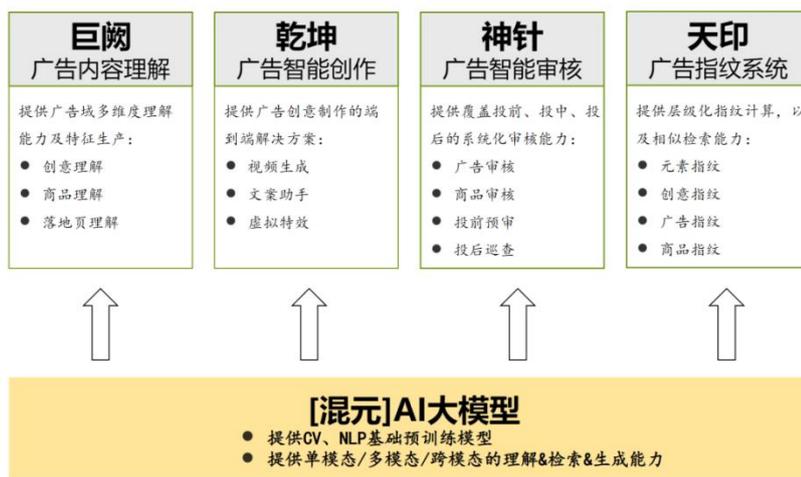


资料来源：量子位，国信证券经济研究所整理

模型应用：已在广告游戏多场景落地，大幅提升效率并降低成本

腾讯广告业务：腾讯在广告业务中应用 AI 模型，助力广告主提升广告创作效率、提升广告推荐系统的内容理解能力，大幅提升了广告效率，实现降本增效。在混元 AI 大模型基础上建立“巨阙”广告内容理解、“乾坤”广告智能创作、“神针”广告智能审核、“天印”广告指纹系统等四大技术平台，提升了广告系统的理解能力，让系统更加智能，从而提高用户体验以及广告转化效果与广告制作效率。

图 28：腾讯广告多媒体 AI 技术研究与应用情况



资料来源：腾讯云开发者，国信证券经济研究所整理

内容创作：腾讯智能创作助手文涌（Effidit）主要用途是帮助创作者开阔思路、提升创作后的文本水平和质量等，目前已经发布 2.0 版本。据官方介绍，文涌融合了知识抽取、文本理解、文本生成、大规模预训练模型、经典语言模型、搜索等技术，部分功能的实现使用“混元”系列 AI 大模型作为底层预训练模型。

图 29：腾讯智能创作助手功能一览

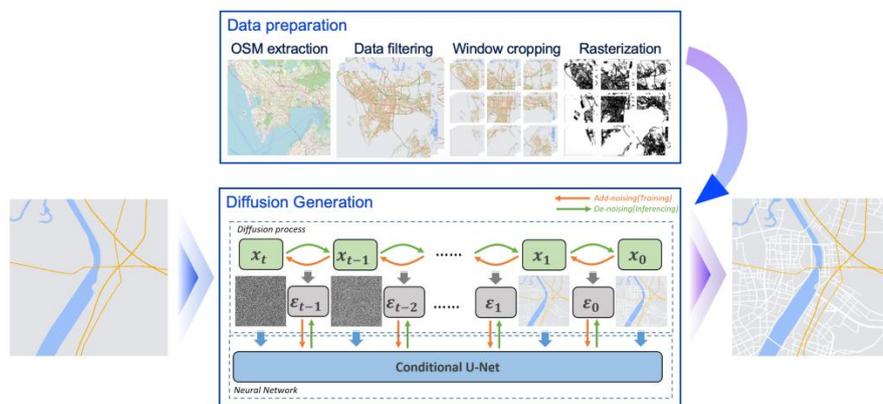


资料来源：Effidit，国信证券经济研究所整理

游戏业务：腾讯近日提出基于 AI Bot 与 AIGC 两大方向的自研 3D 虚拟场景自动生成解决方案，帮助游戏开发者以更低成本创造风格多样、贴近现实的虚拟城市，大幅提升 3D 虚拟场景的生产效率。以 3D 虚拟城市道路布局设计为例，现有技术生成的路网通常横平竖直、风格单一，要实现不同城市风格必须手动设计和反复

修改，耗时至少一周。腾讯解决方案只需用户输入城市主干道和海岸线，模型 1 分钟内即可填充合理、多样的路网细节并支持微调操作，相比现有技术下效率提升近 100 倍。

图 30: 腾讯游戏 AI 路网生成模型



资料来源：腾讯 AI 实验室，国信证券经济研究所整理

对话式智能助手：腾讯集合集团内资源与顶尖人才，正在全力打造类 ChatGPT 对话式产品。 36 氪旗下媒体消息，腾讯类 ChatGPT 对话式产品项目组——腾讯混元助手项目组（HunyuanAide）于 2023 年 2 月已经成立。该项目组联合腾讯内部多部门构建大参数语言模型，目标是通过性能稳定的强化学习算法训练，完善腾讯智能助手工具，打造腾讯智能助手，并能成为国内的业界标杆，我们预计未来 AI 智能助手将广泛应用于腾讯系社交产品和办公产品等。

图 31: 腾讯混元助手项目组织架构



资料来源：职场 Bonus，国信证券经济研究所整理

华为盘古大模型

模型简述：基于 ModelArts 研发设计的系列模型，NLP、CV 等方向性能优越

“盘古”大模型由计算机视觉 (CV) 大模型、科学计算大模型、自然语言处理 (NLP) 大模型、多模态大模型组成，该模型由华为云人工智能领域首席科学家田奇带领的华为云 EI 盘古团队主导设计。“盘古”项目自 2020 年 11 月起立项，2021 年 4 月在华为开发者大会上首次正式发布，2023 年 4 月 8 日，团队介绍了盘古模型的应用进展，包括模型在物流、药物研发、气象预测等领域的落地。

盘古模型基于华为一站式 AI 开发平台 ModelArts 进行研发。ModelArts 是华为开发的 AI 生产线，具备持续构建大模型训练及推理加速能力、分布式训练能力等，并提供数据处理、算法开发、模型训练、模型管理、模型部署等 AI 应用开发全流程技术能力。

图 32：盘古模型基于 ModelArts 平台进行开发设计



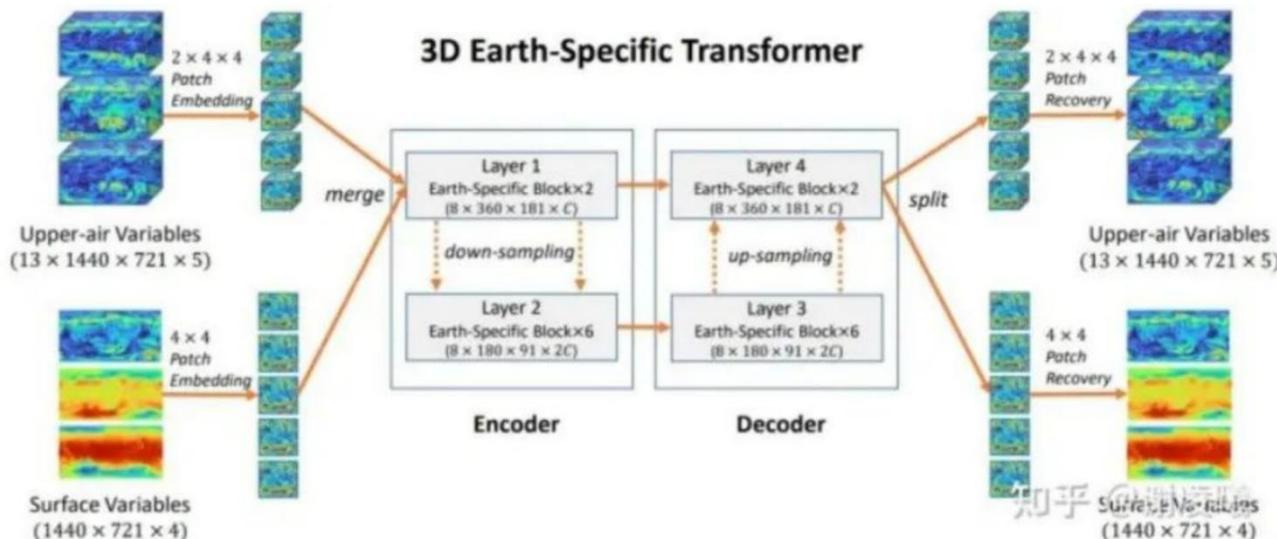
资料来源：计算机文艺复兴，华为云，国信证券经济研究所整理

模型性能优异，NLP 大模型是首个千亿参数级中文大模型，2021 年位列 CLUE 榜单三项第一。2021 年 4 月华为云盘古 NLP 大模型发布，这是当时业界首个 2000 亿参数的中文预训练模型。在 2021 年，盘古 NLP 模型在中文语言理解评测基准 CLUE 榜单中总排行榜及分类、阅读理解单项均排名第一，刷新了三项榜单世界历史纪录。

盘古 CV 大模型是业界最大的 CV 预训练模型。盘古 CV 大模型包含了 30 多亿参数和 10 亿级别的图像，已经在 100 多个场景中得到了验证。在 Image Net 1%、10% 数据集上的小样本分类精度上均达到业界最高水平。

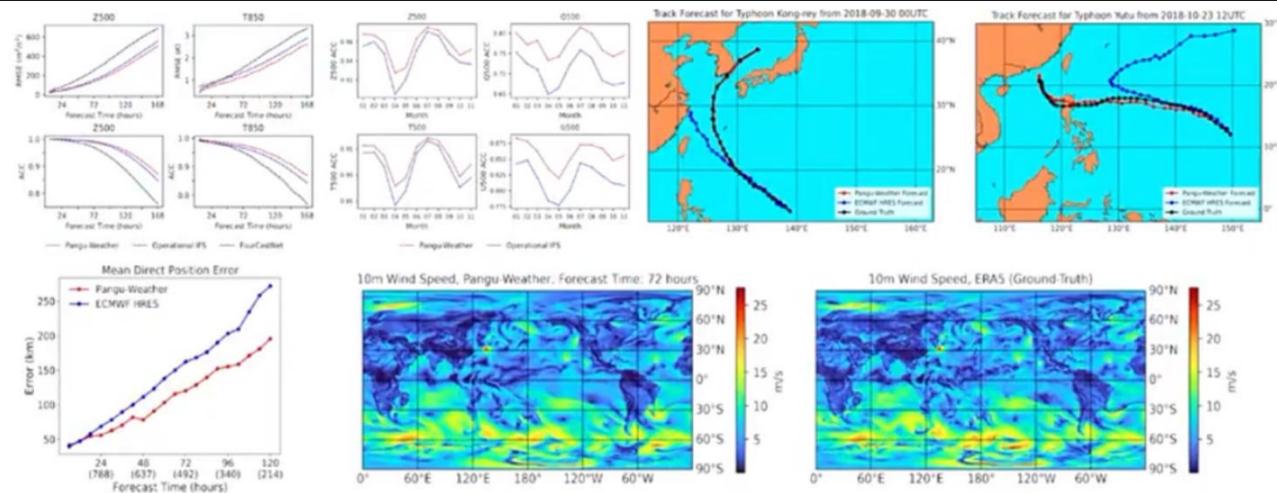
气象预测：盘古气象大模型的研究团队通过 3D Earth-Specific Transformer 的方式处理复杂的不均匀 3D 气象数据，提高 AI 大模型的预测精度。盘古气象大模型能够提供秒级的全球气象预报，其气象预测结果包括位势、湿度、风速等，其精度首次超过了欧洲气象中心的数字分析的方法，并且预测速度提升了 1 万倍以上。

图34: 3D Earth-Specific Transformer 示意图



资料来源：华为云，国信证券经济研究所整理

图35: 盘古气象大模型指标表现优异，预测精度超过传统方式



资料来源：华为云，国信证券经济研究所整理

字节跳动大模型

模型简述：AI 探索基础扎实，语言大模型和图像大模型初步布局
字节跳动 AIGC 大模型将从语言和图像两种模态发力。根据 36 氪消息，字节跳动

在大模型上已有所布局，分别在语言和图像两种模态上发力，预期是在今年年中推出大模型。

字节跳动 AIGC 大模型尚在初期探索阶段，有一定 AI 基础。根据 36 氪，在 ChatGPT 公开前，字节跳动对大模型的投入存疑。但字节跳动在工程团队和云服务团队方面并无短板。

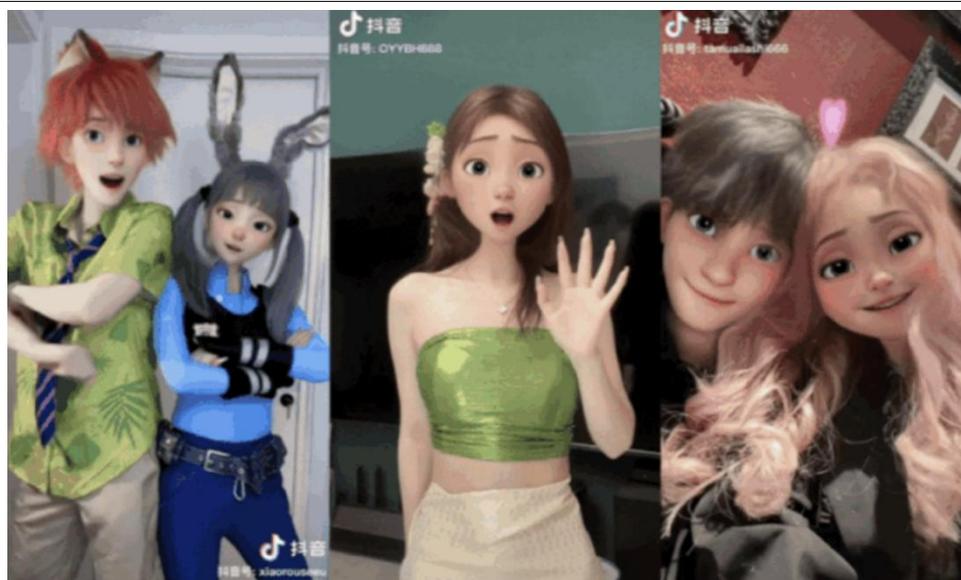
1) 算力：字节并不缺 GPU。字节旗下的企业技术服务平台火山引擎，在 2021 年正式进军基础云市场（IaaS 层），为大模型所需的算力奠定了基础。4 月 18 日，火山引擎发布自研 DPU 等系列云产品，并推出新版机器学习平台：支持万卡级大模型训练、微秒级延迟网络，弹性计算可节省 70%算力成本。同时，火山引擎宣布与字节跳动国内业务并池，抖音等业务的空闲计算资源可极速调度给火山引擎客户使用。

2) 算法：阿里巴巴 M6 大模型的原带头人杨红霞现已加入字节 AI Lab（人工智能实验室），参与语言生成大模型的研发。

模型应用：可用于图文、视频内容生成等，飞书将推出智能 AI 助手“My AI”

抖音、今日头条基于 AIGC 生成图文内容。在 AIGC 的加持下，用户上传照片后，短时间内系统即可自动生成不同风格的图片。例如，抖音上线该“漫画脸”特效后，吸引了大量用户，据抖音平台数据显示，截至 2022 年 12 月 6 日，该特效已经超 2428 万人使用，并迅速飙升至特效潮流榜 TOP1。

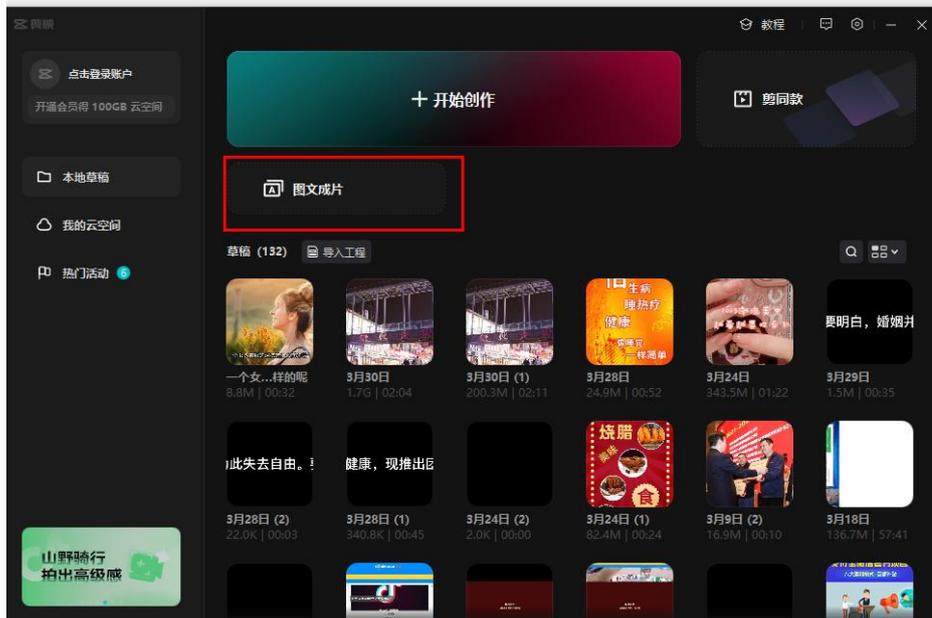
图 36: 抖音漫画脸特效



资料来源：抖音，机器之心，国信证券经济研究所整理

剪映提供 AI 生成视频功能。剪映的创作者可以通过几个关键词或一小段文字，生成一段创意小视频。除此以外，剪映可以根据文字描述智能匹配视频素材，将视频包装为更垂直的内容作品，包括财经、历史、人文等类别。

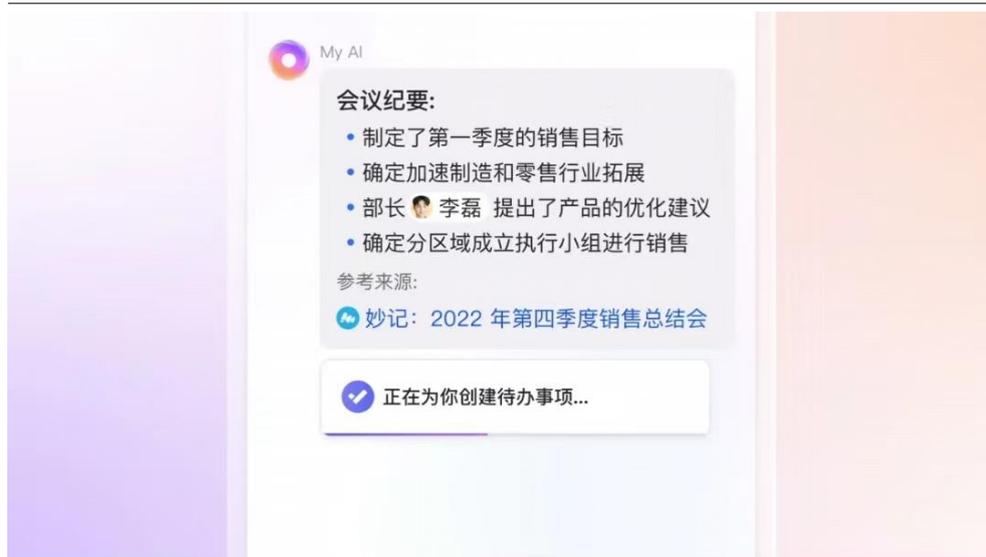
图 37: 剪映 AI 成片



资料来源: 剪映, 国信证券经济研究所整理

飞书将推出智能 AI 助手“My AI”。4 月 11 日字节跳动飞书宣布将推出智能 AI 助手“My AI”，My AI 将包括自动汇总会议纪要、创建报告、优化和续写文字内容，在飞书 app 中，My AI 也可以通过对话形式，帮助用户自动创建日程、搜索公司内部知识库等功能。

图 38: 飞书将推出智能 AI 助手“My AI”



资料来源: 飞书视频号, 国信证券经济研究所整理

投资建议

ChatGPT 的涌现能力的实现以及应用成功，使得 AI 产业发展进入新阶段。AI 大模型增强机器理解能力，可以优化人类与机器交互方式，是效率提升的革命。大模型“预训练+精调”即可对下游应用赋能。当前，各家公司开展新一轮 AI 模型竞赛，希望借助大模型能力开展“模型即服务”的范式。算力、算法、数据是大模型能否成功的重要因素，我们重点关注在大模型领域有技术储备的国内公司腾讯控股、阿里巴巴、百度等。

表4：相关公司盈利预测及估值

公司代码	公司名称	投资评级	收盘价	EPS			PE			PB (MRQ)
				2022	2023E	2024E	2022	2023E	2024E	
0700. HK	腾讯控股	买入	365.4	12.1	15.7	18.0	26	20	18	4.3
9988. HK	阿里巴巴-SW	买入	94.8	6.4	6.7	7.6	13	12	11	1.8
9888. HK	百度集团-SW	增持	129.6	7.4	7.9	9.3	15	14	12	1.4

数据来源：wind、国信证券经济研究所预测及整理

风险提示

AI 技术迭代进度不及预期；AI 应用落地进度不及预期；政策风险；供应链风险等。

免责声明

分析师声明

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

国信证券投资评级

类别	级别	说明
股票 投资评级	买入	股价表现优于市场指数 20%以上
	增持	股价表现优于市场指数 10%-20%之间
	中性	股价表现介于市场指数 \pm 10%之间
	卖出	股价表现弱于市场指数 10%以上
行业 投资评级	超配	行业指数表现优于市场指数 10%以上
	中性	行业指数表现介于市场指数 \pm 10%之间
	低配	行业指数表现弱于市场指数 10%以上

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所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或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中所提及的意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所

深圳

深圳市福田区福华一路 125 号国信金融大厦 36 层
邮编：518046 总机：0755-82130833

上海

上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层
邮编：200135

北京

北京西城区金融大街兴盛街 6 号国信证券 9 层
邮编：100032