

中国计算机和软件 China Computer & Software

MaaS Model as a Service 模型即服务

MaaS Model as a Service

观点聚焦 Investment Focus

股票名称	评级
商汤集团	Outperform
科大讯飞	Outperform

(Please see APPENDIX 1 for English summary)

大模型是人工智能的发展趋势和未来。大模型是“大算力+强算法”结合的产物。目前，大模型生态已初具规模。大模型能够实现 AI 从“手工作坊”到“工厂模式”的转变，大模型通常是在大规模无标注数据上进行训练，学习出一种特征和规则，基于大模型进行应用开发时，将大模型进行微调，或者不进行微调，就可以完成多个应用场景的任务；并且，大模型具有自监督学习能力，不需要或很少需要通过人工标注的数据进行训练，降低训练成本，因而能够加速 AI 产业化进程，降低 AI 应用的门槛。另外，随着大模型不断地迭代，大模型能够达到更强的通用性以及智能程度，从而使得 AI 能够更广泛地赋能各行业应用。我们认为，大模型是人工智能的发展趋势和未来。

模型公司将是 AI 世界的基础类公司。大模型在能力泛化与技术融合方面的优势，使其在应用支撑方面具有先进性；同时大模型做到了相对标准化，能够降低 AI 技术应用于下游场景的成本。因此，**大模型在 AI 产业链中将承担“基础设施”式的功能，作为底座将 AI 技术赋能千行百业。**我们认为，模型公司能够发挥大模型作为 AI 基础设施与底座的能力，基于大模型“大规模预训练 + 微调”的范式满足 AI 产业链下游各行业客户的应用需求，模型公司将成为 AI 世界中的基础类公司。

模型公司将对科技公司竞争格局带来颠覆。2022 年 11 月 30 日，OpenAI 发布 ChatGPT，一款人工智能技术驱动的自然语言处理工具，能够通过学习和理解人类的语言来进行对话和互动，甚至能完成撰写邮件、视频脚本、文案、翻译、代码等任务。**ChatGPT 对搜索领域或带来巨大冲击。**由于 ChatGPT 能够与用户进行交流明确需求并具备文本生成能力进行回复，其相对于传统搜索引擎在输入端和输出端都具有难以替代的优势。因而 ChatGPT 可能对搜索带来一个重大变化：**用户将会转向聊天机器人寻求帮助，而不是通过谷歌提供的网站进行过滤。**同时，技术上，ChatGPT 也可能降低搜索引擎的门槛。可以说，ChatGPT 已经真真切切地改变了搜索领域，对众多科技公司产生了巨大的挑战。

MaaS 将成为人工智能公司的核心商业模式。模型是 MaaS 的最重要基座；单点工具，如基于 GPT 3.5 的 ChatGPT，是大模型的直接产品；可以预见，随着未来大模型的发展持续进行，大模型以及其产生的产品的应用场景将会持续扩展。MaaS 公司的收费已经开始，ChatGPT 采取**订阅制收费模式**，并且收费版的 ChatGPT Plus 的确拥有广泛的市场空间；



资料来源: Factset, HTI

Related Reports

全面实行股票发行注册制改革正式启动，或带来证券 IT 全新增量需求 (Full Implementation of Stock Issuance Registration System Reform Officially Launched, May Bring New Incremental Demand for Securities IT) (2 Feb 2023)

商汤作为富时全球指数成分股候选将被评估，美国投资者或可投资商汤 (Sense Time Group Will Be Evaluated as an FTSE Global Index Constituent Candidate, U.S. Investors May Be Able to Invest in Sense Time Group) (2 Feb 2023)

计算机行业跟踪周报 295 期: 计算机行业春季行情综述 (Computer Industry Weekly Report (295): Computer Industry Spring Market Preview) (29 Jan 2023)

郑宏达 Nathan Zheng
hongda.zheng@htisec.com

杨林 Lin Yang
lin.yang@htisec.com

洪琳 Lin Hong
lin.hong@htisec.com

类似的 AI 聊天公司 Jasper 分三档同样实施订阅服务的收费。而除了模型本身进行订阅收费以外，**嵌入其他产品获得引流式收入**也是模式之一，例如，微软宣布推出高级付费版 Microsoft Teams Premium，订阅者可享用由 OpenAI GPT-3.5 提供支持的“智能回顾（Intelligent Recap）”这一重要功能，该功能将提供自动生成的会议记录、推荐任务和个性化标亮；微软旗下 Dynamics 365 产品线宣布旗下客户关系管理软件 Viva Sales 也将集成 OpenAI 的技术，通过人工智能帮助销售人员完成许多繁杂且重复的文字工作。其它的收费模式还包括**不同的 Model 应对不同的客户需求和客户定价**，如 OpenAI 在推出 ChatGPT 以前，其主要收入来源于对外部客户开源自身大模型，通过提供 API 接口调用来收费。客户的要求越高，使用的模型越好，那么大模型的收费也将会更高。

建议关注。 商汤集团，云从科技，科大讯飞，百度，华为。

风险提示。 行业应用不及预期，AIGC 技术发展不及预期的风险。

1. 大模型是人工智能的发展趋势和未来

大模型，又称为预训练模型、基础模型等，是“大算力+强算法”结合的产物。大模型通常是在大规模无标注数据上进行训练，学习出一种特征和规则。基于大模型进行应用开发时，将大模型进行微调，如在下游特定任务上的小规模有标注数据进行二次训练，或者不进行微调，就可以完成多个应用场景的任务。

迁移学习是预训练技术的主要思想。当目标场景的数据不足时，首先在数据量庞大的公开数据集上训练基于深度神经网络的 AI 模型，然后将其迁移到目标场景中，通过目标场景中的小数据集进行微调，使模型达到需要的性能。在这一过程中，这种在公开数据集训练过的深层网络模型，即为“预训练模型”。使用预训练模型很大程度上降低了下游任务模型对标注数据数量的要求，从而可以很好地处理一些难以获得大量标注数据的新场景。大模型正是人工智能发展的趋势和未来。

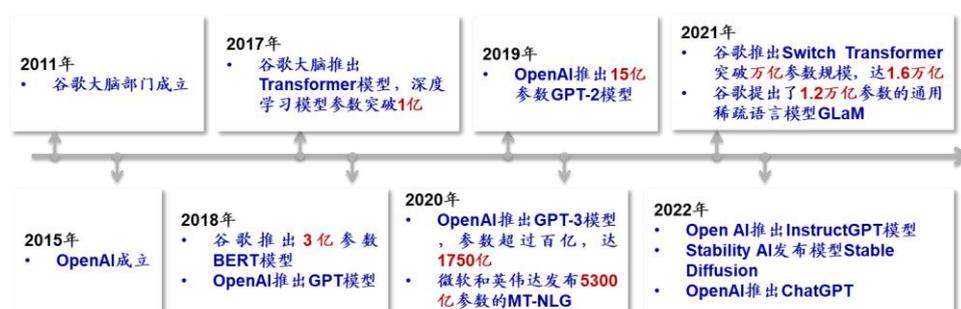
1.1 国内外公司与机构竞相研发，大模型生态已初具规模

从参数规模上看，AI 大模型先后经历了预训练模型、大规模预训练模型、超大规模预训练模型三个阶段，参数量实现了从亿级到百万亿级的突破。从模态支持上看，AI 大模型从支持图片、图像、文本、语音单一模态下的单一任务，逐渐发展为支持多种模态下的多种任务。

国外的超大规模预训练模型起步于 2018 年，并在 2021 年进入“军备竞赛”阶段。2017 年，Vaswani 等提出 Transformer 架构，奠定了当前大模型领域主流的算法架构基础；Transformer 结构的提出，使深度学习模型参数达到了上亿的规模。2018 年，谷歌提出了大规模预训练语言模型 BERT，该模型是基于 Transformer 的双向深层预训练模型，其参数首次超过 3 亿规模；同年，OpenAI 提出了生成式预训练 Transformer 模型——GPT，大大地推动了自然语言处理领域的发展。此后，基于 BERT 的改进模型、ELNet、RoBERTa、T5 等大量新式预训练语言模型不断涌现，预训练技术在自然语言处理领域蓬勃发展。

2019 年，OpenAI 继续推出 15 亿参数的 GPT-2，能够生成连贯的文本段落，做到初步的阅读理解、机器翻译等。紧接着，英伟达推出了 83 亿参数的 Megatron-LM，谷歌推出了 110 亿参数的 T5，微软推出了 170 亿参数的图灵 Turing-NLG。2020 年，OpenAI 推出了超大规模语言训练模型 GPT-3，其参数达到了 1750 亿，在两年左右的时间实现了模型规模从亿级到上千亿级的突破，并能够实现作诗、聊天、生成代码等功能。此后，微软和英伟达在 2020 年 10 月联手发布了 5300 亿参数的 Megatron-Turing 自然语言生成模型（MT-NLG）。2021 年 1 月，谷歌推出的 Switch Transformer 模型以高达 1.6 万亿的参数数量成为史上首个万亿级语言模型；同年 12 月，谷歌还提出了 1.2 万亿参数的通用稀疏语言模型 GLaM，在 7 项小样本学习领域的性能超过 GPT-3。可以看到，大型语言模型的参数数量保持着指数增长势头。这样高速的发展并没有结束，2022 年，又有一些常规业态大模型涌现，比如 Stability AI 发布的文字到图像的创新模型 Diffusion，以及 OpenAI 推出的 ChatGPT，ChatGPT 是由效果比 GPT3 更强大的 GPT-3.5 系列模型提供支持，并且这些模型使用微软 Azure AI 超级计算基础设施上的文本和代码数据进行训练。

图1 国外大模型发展历程



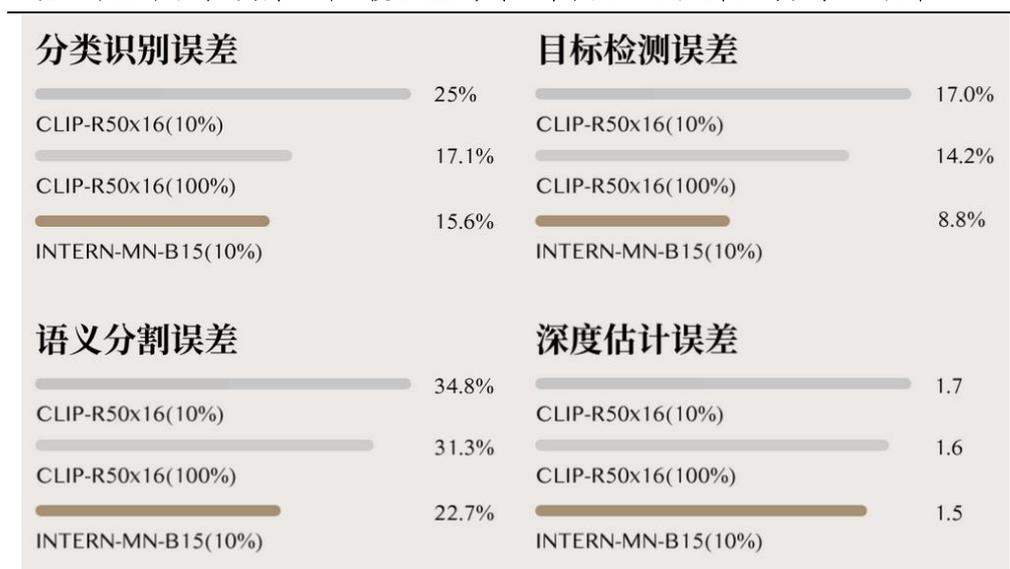
资料来源：IT之家，人工智能学家公众号，腾讯新闻，HTI

而在国内，超大模型研发异常迅速，2021年成为中国AI大模型的爆发年。2021年，商汤发布了书生（INTERN）大模型，拥有100亿的参数量，这是一个相当庞大的训练工作。在训练过程中，大概有10个以上的监督信号帮助模型，适配各种不同的视觉或者NLP任务，截至2021年中，商汤已建成世界上最大的计算器视觉模型，该模型拥有超过300亿个参数；同年4月，华为云联合循环智能发布盘古NLP超大规模预训练语言模型，参数规模达1000亿；联合北京大学发布盘古 α 超大规模预训练模型，参数规模达2000亿。阿里达摩院发布270亿参数的中文预训练语言模型PLUG，联合清华大学发布参数规模达到1000亿的中文多模态预训练模型M6；7月，百度推出ERNIE 3.0知识增强大模型，参数规模达到百亿；10月，浪潮信息发布约2500亿的超大规模预训练模型“源1.0”；12月，百度推出ERNIE 3.0 Titan模型，参数规模达2600亿。而达摩院的M6模型参数达到10万亿，将大模型参数直接提升了一个量级。2022年，基于清华大学、阿里达摩院等研究成果以及超算基础实现的“脑级人工智能模型”八卦炉（BAGUALU）完成建立，其模型参数模型突破了174万亿个，完全可以与人脑中的突触数量相媲美。

部分中国公司虽然目前还没有正式推出自身大模型产品，但是也在积极进行研究，例如云从科技，公司的研发团队高度认同“预训练大模型+下游任务迁移”的技术趋势，从2020年开始，已经陆续在NLP、OCR、机器视觉、语音等多个领域开展预训练大模型的实践，不仅进一步提升了公司各项核心算法的性能效果，同时也大幅提升了公司的算法生产效率，已经在城市治理、金融、智能制造等行业应用中体现价值。

而且，中国的超大模型并不落后于国外同类产品，在某些领域还能实现反超。以商汤科技的书生（INTERN）为例，书生（INTERN）在分类、目标检测、语义分割、深度估计四大任务26个数据集上，基于同样下游场景数据（10%），相较于同期OpenAI发布的最强开源模型CLIP-R50x16，平均错误率降低了40.2%，47.3%，34.8%，9.4%。同时，书生只需要10%的下游数据，平均错误率就能全面低于完整（100%）下游数据训练的CLIP。

图2 “书生”相较于同期最强开源模型 CLIP 在准确率和数据使用效率上均取得大幅提升



资料来源：商汤科技官微，HTI
注：CLIP 为 OpenAI 在 2021 年发布

可以看到，国内外众多公司和研究机构积极开展对大模型的研发。目前，大模型参数规模最高可达百万亿级别，数据集达到 TB 量级，且面向多模态场景（同时支持文字、图像、声音、视频、触觉等两种及以上形态）的大模型已成为趋势。大模型生态已初具规模。

表 1 国内外主要大模型梳理，大模型生态已初具规模

企业	大模型	参数	算力	数据量	模型类型
商汤科技等	书生 (INTERN+)	100 亿	商汤 AIDC, 峰值算力 3740 Petaflops	-	计算器视觉模型
商汤科技	某世界上最大的计算机视觉模型	300 亿	商汤 AIDC, 峰值算力 3740 Petaflops	-	计算器视觉模型
清华大学等	“八卦炉” (脑级 AI 模型)	174 万亿参数 (与人脑中突出数量媲美)	“海洋之光”超级计算机 (国产超算)	中文多模态数据集 M6-Corpus	多模态预训练模型
阿里	M6	10 万亿	5 12 块 GPU	1.9TB 图像和 292GB 文本	多模态预训练模型
腾讯	“混元”HunYuan_tvr	万亿	腾讯太极机器学习平台	五大跨模态视频检索数据集	多模态预训练模型
微软和英伟达	Megatron-Turing	5300 亿	280 块 GPU	3390 亿条文本数据	NLP 大模型
百度和鹏城实验室	ERNIWI 3.0 Titan	2600 亿	鹏城云脑 II 算力集群 (2048 块 CPU) 和百度飞桨深度学习平台	纯文本和知识图谱的 4TB 语料库	NLP 大模型
浪潮信息	源 1.0	2457 亿	4095 (Pflops-day) / 2128 张 GPU	5000GB 高质量中文数据集	NLP 大模型
OpenAI	GPT3.5	1750 亿	3640 (Pflops-day) / 上万块 V100 GPU 组成的 gao 带宽集群算力	超过万亿单词的人类语言数据集	多模态预训练模型
华为云	盘古系列大模型	千亿	鹏城云脑 II 和全场景 AI 计算框架 MindSpore, 2048 块 GPU	40TB 训练数据	多模态预训练模型
中科院自动化所	紫东太初	千亿	昇腾 AI 基础软硬件平台	基于万条小规模数据集	图、文、音三模态
澜舟	孟子	10 亿	16 块 GPU	数百 G 级别不同领域的高质量语料	多模态预训练模型

资料来源：商汤科技官微，天翼智库，HTI

注：Pflops-day 为算力单位，意为一天可以进行约 10^{20} 运算

1.2 大模型能给 AI 产业带来什么

1.2.1 大模型加速 AI 产业化进程，降低了 AI 应用的门槛

人工智能正处于从“可以用”逐渐走向“好用”的落地应用阶段，但目前仍处于商业落地早期，主要面临着场景需求碎片化、人力研发和应用计算成本高、长尾场景数据较少导致模型训练精度不够、模型算法从实验室场景到真实场景效果差距大等行业痛点。而大模型的出现能够在提高模型的通用性、降低训练研发成本等方面降低 AI 落地应用的门槛。

(1) 大模型能够实现 AI 从“手工作坊”到“工厂模式”的转变

过去十年中，通过“深度学习+大算力”从而获得训练模型是实现人工智能的主流技术途径。由于深度学习、数据和算力这三个要素都已具备，全世界掀起了“大炼模型”的热潮，也催生了大批人工智能企业。但是，在深度学习技术兴起的近 10 年间，AI 模型基本上是针对特定应用场景需求进行训练的，即小模型，属于传统的定制化、作坊式的模型开发方式。传统的 AI 模型从研发到投入应用需要完成包括确定需求、数据收集、模型算法设计、训练调优、应用部署和运营维护等阶段组成的整套流程。这意味着除了需要优秀的产品经理准确确定需求之外，还需要 AI 研发人员扎实的专业知识和协同合作能力完成大量复杂的工作。

图3 传统的定制化、作坊式模型开发流程



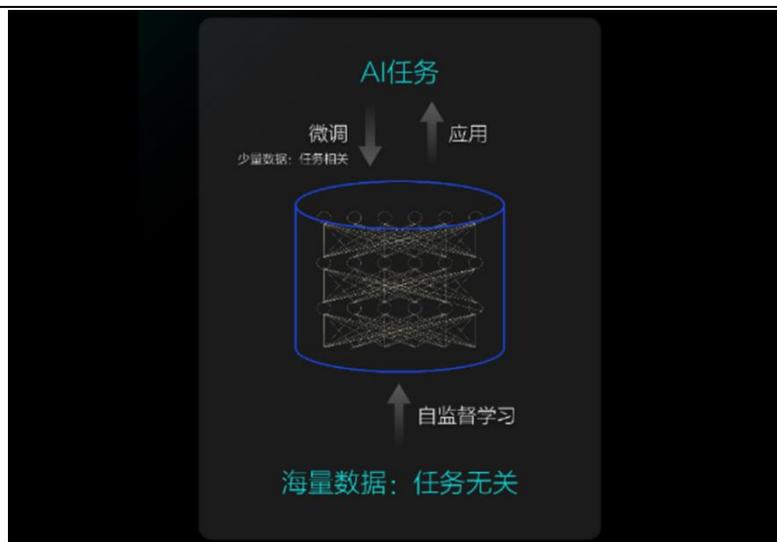
资料来源：中国科学院自动化研究所官微，HTI

传统模式中，**研发阶段**，为了应对各式各样的场景需求，AI 研发人员需要设计专网专用的个性化定制化神经网络模型。模型设计过程要求研究人员具有充分的网络结构和场景任务专业知识，且需承受人力设计网络结构的试错成本和时间成本。一种降低专业人员设计门槛的思路是通过网络结构自动搜索的技术路线，但是这种方案对算力要求很高，并且不同场景任务都要调用大量机器进行自动搜索以获取最优模型，算力成本和时间成本仍然很高。一个项目往往需要专家团队实地驻场数月才能完成，其中数据收集和模型训练评估是否能够达到指标要求往往需要多次循环迭代，人力成本很高。

落地阶段，经“一个场景一个模型”式的作坊式模型开发得到的模型在许多垂直行业场景任务不通用。譬如无人驾驶全景感知领域，常常需要多行人跟踪、场景语义分割、视野内目标检测等多个模型协同作战才行；而同样是目标检测和分割应用，在医学图像领域训练的皮肤癌检测和分割 AI 模型也无法直接应用到监控场景的行人车辆检测和场景分割。模型无法复用和积累，同样导致了 AI 落地的高门槛、高成本与低效率。

而大模型通过从海量的、多类型的场景数据中学习，并总结不同场景、不同业务下的通用能力，学习出一种特征和规则，成为具有泛化能力的模型底座。基于大模型进行应用开发或面对新的业务场景时，将大模型进行微调，例如在下游特定任务上的小规模有标注数据进行二次训练，或不进行微调，就可以完成多个应用场景的任务，实现通用的智能能力。由此利用大模型的通用能力可以有效的应对多样化、碎片化的 AI 应用需求，为实现规模推广 AI 落地应用提供可能。

图4 AI 大模型“工厂模式”的开发方式



资料来源：百度文心大模型官微，HTI

(2) 大模型具有自监督学习能力，能够降低 AI 开发以及训练成本

传统的小模型训练过程中，涉及众多调参、调优的手动工作，需要大量的 AI 专业研发人员来完成；同时，模型训练对数据要求高，需要大规模的标注数据。然而许多行业数据获取困难、标注成本高，同时项目研发人员需要花费大量的时间收集原始数据。譬如，人工智能在医疗行业病理学、皮肤病学和放射学等医学图像密集领域的影响扩展增长，但是医学图像通常涉及到用户数据隐私，很难大规模获取到用于训练 AI 模型。而在工业视觉瑕疵检测领域，以布匹瑕疵为例，市场上需要检测的布匹种类包括白坯布、色坯布、成品布、有色布、纯棉、混纺等等，瑕疵种类繁多，颜色、厚薄难以识别，需要在工厂里长时间地收集数据并不断优化算法才能做好瑕疵检测。工业视觉应用场景千变万化，每个场景都具有专业知识的个性化的需求，而不同行业的数据在另一个场景或任务可能并不适用，还需重新收集、标注数据和训练模型，造成重复造车轮子现象，研发流程虽然相通，但研发的模型却难以复用。此外，AI 模型也需要完整的、实时的支持机器自学习的数据闭环，从而能够不断迭代优化。这一方面整体也导致了当前 AI 研发整体门槛、成本较高。

而大模型通过自监督学习功能将输入的原始数据直接进行自动学习区分，合理构造适合模型学习的任务，不需要或很少需要通过人工标注的数据进行训练，很大程度上解决了人工标注数据标签的高成本、长周期和精确度的问题，并降低了训练所需要的数据规模。这在很大程度上减少了大模型训练数据获取、标注的成本，也更适合于小样本学习，从而有利于将传统的、有限的 AI 扩展到更多的应用场景。

我们认为，大模型相对于传统 AI 模型开发模式在研发时拥有更标准化的流程，在落地时拥有更强的通用性，可以泛化到多种应用场景；并且大模型的自监督学习能力相较于传统的需要人工标注的模型训练能够显著降低研发成本，共同使得大模型对于 AI 产业具有重要意义，为解决 AI 落地难、促进 AI 产业化进程这一问题提供方向。

1.2.2 大模型带来更强大的智能能力

除通用能力强、研发过程标准化程度高外，大模型最大的优势还是在于“效果好”，其通过给模型“填喂”大数据提高其自学习能力，进而具有更强的智能程度。比如在自然语言处理领域，百度、谷歌等巨头的探索表明，**基于预训练大模型的 NLP 技术的效果，已超过过去最好的机器学习能力**。OpenAI 的研究表明，2012 至 2018 年 6 年间，在最大规模的人工智能模型训练中所使用的计算量呈指数级增长，其中有 3.5 个月的时间计算量翻了一倍，比摩尔定律每 18 个月翻一倍的速度快得多。下一代 AI 大模型，参数数量级将堪比人类大脑突触的水平，并且或将不仅能处理语言模型，将更是一个能处理语言、视觉、声音等多任务的多模态 AI 模型。大模型为通往强人工智能提供了一条可能的通道。

图5 弱人工智能仍属于计算机“工具”范畴，强人工智能能自适应地完成任

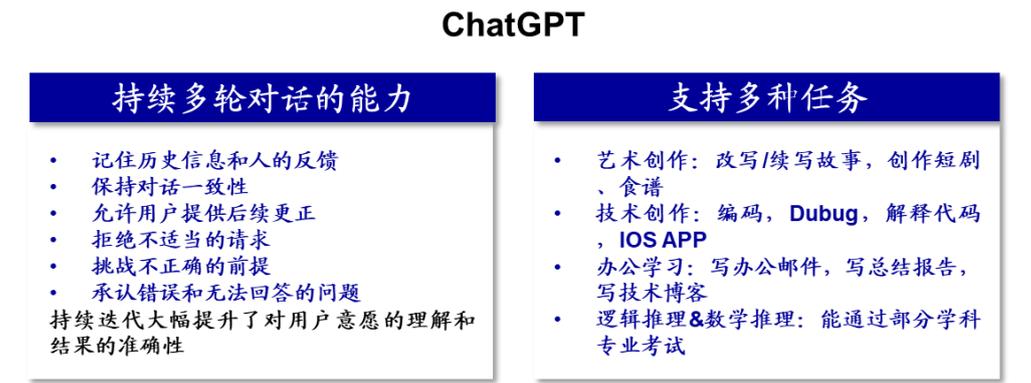


资料来源：知乎，HTI

从大模型发展历程中能够看出，多模态大模型是发展趋势之一。由于具有在无监督情况下自动学习不同任务、并快速迁移到不同领域数据的强大能力，**多模态大模型**被广泛认为是从限定领域的弱人工智能迈向强人工智能的路径探索。OpenAI 联合创始人、首席科学家 Ilya Sutskever 也曾表示，“人工智能的长期目标是构建多模态神经网络，即 AI 能够学习不同模态之间的概念，从而更好地理解世界”。将文本、语音、图像、视频等多模态内容联合起来进行学习，大模型由单模态向多模态方向发展，**能够对更广泛、更多样的下游任务提供模型基础支撑，从而实现更加通用的人工智能模型**。

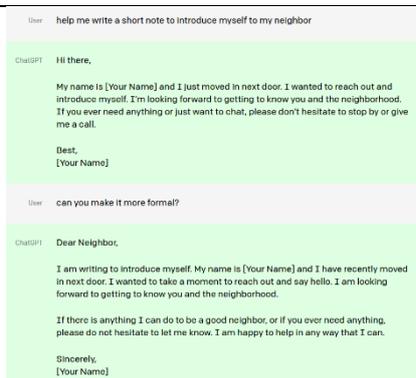
更具体来看，大模型带来的更强大的智能能力，**能够推动人工智能向更高级智能应用领域迈进**，例如 AIGC、更智能的对话客服等领域。GPT-3 等大模型在新闻文本生成、商业文本分析、法律文本分析等领域具有较高的产业应用价值。2022 年 OpenAI 发布的基于 GPT-3.5 的 ChatGPT 则能够以对话方式进行交互；除问答以外，ChatGPT 能够完成编程、写稿等众多超出人们想象的任务。甚至有研究人员表示未来 ChatGPT 这样的新兴人工智能技术很可能会影响白领的工作。在一项由医疗初创公司 Ansible Health 领导的研究中，ChatGPT 参加美国医生执照的三项考试，考试成绩可以达到或接近达到及格；ChatGPT 在法律领域的表现甚至更加优秀，研究人员发现，ChatGPT 可以取得美国多州律师考试的及格分数；ChatGPT 的能力也已经获得了大型科技企业的认可，谷歌公司指出，如果 ChatGPT 参加谷歌的面试，理论上会得到入门级程序员的聘任。

图6 ChatGPT 的功能及特点



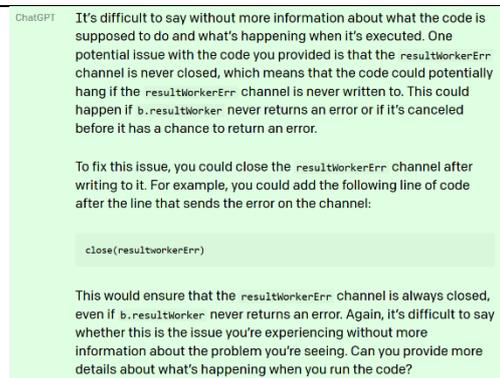
资料来源：钛媒体公众号，HTI

图7 ChatGPT 能够持续多轮对话，并具备一定写作能力



资料来源：OpenAI 官网，HTI

图8 ChatGPT 能够修改程序中的错误



资料来源：OpenAI 官网，HTI

2022 年，大模型正在成为 AIGC 领域发展的算法引擎。在大模型的能力加持下，包括以文生图以及虚拟数字人等 AIGC 类应用将快速进入到商业化阶段，并为元宇宙内容生产带来巨大的变革。大模型正在让人工智能技术从五年前的“能听会看”，走到今天的“能思考、会创作”，未来有望实现“会推理、能决策”的重大进步。

我们认为，随着大模型不断地进步、迭代、向多模态方向发展，大模型能够达到更强的通用性以及智能程度，从而使得 AI 能够更广泛地赋能各行业应用，尤其是更高级智能应用领域，这是大模型能够给 AI 产业带来的突破之一，也是大模型构成人工智能发展趋势和未来的原因之一。

2. 模型公司将是 AI 世界的基础类公司

2.1 大模型，智能时代的基础设施

大模型的“大规模”和“预训练”属性，决定了其具有能力泛化、技术融合、应用支撑三大核心作用。

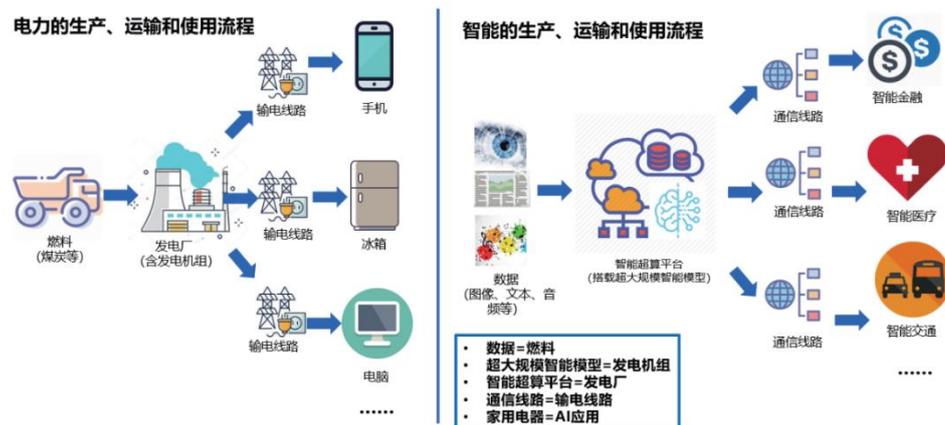
能力泛化方面，AI 大模型预先在海量通用数据上训练并具备多种基础能力，可结合多种垂直行业和业务场景需求进行模型微调和应用适配，能够摆脱传统 AI 能力碎片化、作坊式开发的束缚。AI 大模型得益于其“大规模预训练+微调”的范式，可以很好地适应不同下游任务，展现出它强大的通用性。

技术融合方面，单个 AI 大模型通过端到端联合训练调优，能有效集成自然语言处理、计算机视觉、智能语音、知识图谱等多个 AI 核心研究领域的多项技术，性能上可实现“1+1>2”的效果，可显著提升 AI 大模型的功能丰富性和性能优越性。

应用支撑方面，AI 大模型已成为上层应用的技术底座，能够有效支撑智能终端、系统、平台等产品应用落地，可解决传统 AI 应用过程中存在的壁垒多、部署难问题。

从人工智能到各行业的商业应用，可以看作是上下游的关系。而 AI 大模型因为其自身而也正是大模型在能力泛化与技术融合方面的优势，使其在应用支撑方面具有先进性；同时大模型做到了相对标准化，下游可以降低对算法的使用成本，以及商业应用的适配成本。因此，大模型在 AI 产业链中将承担“基础设施”式的功能，作为底座将 AI 技术赋能千行百业。也就是说，在基于数据的互联网时代、基于算力的云计算时代之后，将进入基于大模型的 AI 时代。在未来，基于大模型，人工智能将如供水供电一般流向终端，流向用户和企业。

图9 大模型将作为一种基础设施将 AI 赋能千行百业



资料来源：智源研究院，HTI

如果将 AI 比作电力，那么大模型则相当于“发电机”，能将智能在更大的规模和范围普及。大模型的智能能力在未来将成为一种公共基础资源，像电力或自来水一样随取随用。每个智能终端、每个 APP，每个智能服务平台，都可以像接入电网一样，接入由 IT 基础设施组成的“智力”网络，让 AI 算法与技术能够更广泛地应用于各行各业。

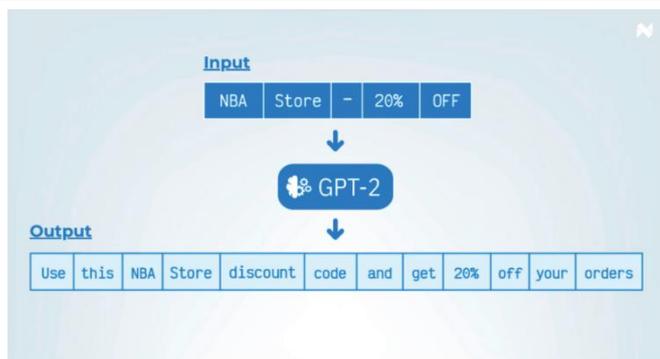
基于大模型形成的大规模智能网络和云计算完全不同。云计算尽管也将算力比作随取随用的资源，但目前仍然做不到像电力一样的方便。这是因为云计算本质上需要云和用户进行双向的资源交换，用户需要经过繁琐的数据准备、计算过程定义等过程，将数据和算法转移到云端，再由云端完成任务后传回本地。而基于大模型的大规模智能网络则不需要用户定义计算过程，只需要准备好数据，大模型强大的能力能够在少量微调甚至不微调的情况下直接完成用户所需的工作。这一智能能力生产和消费的网络，未来或将成为人类社会产业发展的主流模式。

我们认为，大模型在能力泛化、技术融合、研发标准化程度高等方面的优势让其有能力支撑各式应用，使其正在成为人工智能技术及应用的新基座。正如发电厂和高速公路一样，大模型将成为各行各业应用 AI 技术的底座和创新的源头。

2.2 模型公司以大模型为基础赋能 AI 产业链下游应用

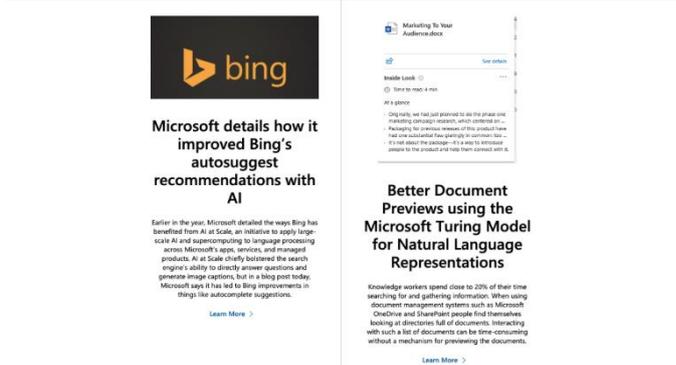
从当前阶段来看，国际上大模型研究开展的时间较早，在研发过程中积累了一些经验。目前有商业化应用案例的大模型主要包括 BERT、GPT-2、T-NLG、GPT-3 等。BERT 主要适用于搜索、营销、金融等依赖深层语义理解的领域，例如，在检索领域，谷歌目前已将 BERT 整合到搜索引擎中，为知识卡片和内容索引提供更为准确的结果。在营销领域，一些企业采用了开源的 BERT 模型，BERT 能够将大量的用户反馈内容进行整合，为下游的营销动作提供支持；GPT-2 更适合文本生成领域的任务，在营销、文案、对话、媒体等领域的应用较多；T-NLG 模型是自然语言生成模型，在文本生成、摘要等方面具有更强的性能；GPT-3 是目前商业化进展最快，应用案例最多的大模型。基于 GPT-3 的下游应用主要集中在内容创意生成、语言/风格互译、对话、搜索、游戏辅助开发等其它场景。

图10 GPT-2 根据关键词扩写营销文案



资料来源：Narrativa 官网，HTI

图11 T-NLG 应用案例



资料来源：微软官网，HTI

国内，商汤的业务模式能够大致体现 AI 大模型、AI 大装置承担基础设施功能，搭载各式平台从而赋能百业。商汤拥有强大的人工智能基础设施 SenseCore，它以低成本、高效率大规模生产高性能的人工智能模型。SenseCore 生产的人工智能模型具备感知智能、决策智能、智能内容生成及智能内容增强等功能，被集成到公司的产品及软件平台中以支持智慧城市、智慧商业、智慧生活以及智能汽车等垂直行业。业务模式上，公司使用 SenseCore 在模型库中组装现有模型，并针对增量需求训练新模型，从而以经济高效的方式配置产品方案。这类产品方案可通过软件平台向其他客户提供。

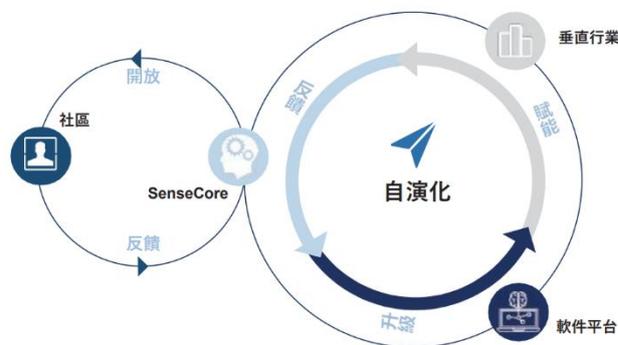
SenseCore 提供强大的人工智能模型，可扩展及适应广泛的应用场景。SenseCore 可以按需提供人工智能模型生产、训练、部署及迭代，以更大的灵活性服务于客户，并缩短商业化时间。每一次 SenseCore 随着技术突破而升级，公司针对不同领域垂直行业的软件平台都会同步升级，功能和性能方面得到明显改善。另一方面，众多垂直行业也为 SenseCore 提供了丰富的场景及反馈，以促进 SenseCore 的演进。这种飞轮效应使公司在与从事单一垂直行业人工智能公司竞争当中处于更有利的地位。

图12 商汤在 SenseCore 基础上搭载行业软件平台



资料来源：商汤集团招股说明书（聆讯后资料集），HTI

图13 SenseCore 带来的飞轮效应



资料来源：商汤集团招股说明书（聆讯后资料集），HTI

又比如云从科技，公司的研发团队高度认同“预训练大模型+下游任务迁移”的技术趋势，从 2020 年开始，已经陆续在 NLP、OCR、机器视觉、语音等多个领域开展预训练大模型的实践，不仅进一步提升了公司各项核心算法的性能效果，同时也大幅提升了公司的算法生产效率，已经在城市治理、金融、智能制造等行业应用中体现价值；其次，公司一直以来都在人机协同领域布局，打造了像人一样思考和工作的人机协同操作系统（CWOS），致力于整合打通视觉、语音、NLP 等多个领域的大模型，要彻底打通数字世界和物理世界，为“像人一样思考和工作”打下坚实的技术基础；公司也通过开放的人机协同操作系统实现了技术平台化，加上多年的行业深耕，能够通过“平台化的通用模型+带行业 knowhow 的专用模型”来帮助各行各业快速实现智能化升级。

图14 云从科技人机协同操作系统（CWOS）致力于整合多个领域的大模型



资料来源：云从科技官网，HTI

我们认为，模型公司能够发挥大模型作为 AI 基础设施与底座的能力，基于大模型“大规模预训练 + 微调”的范式形成产品方案，并满足 AI 产业链下游各行业客户的应用需求，模型公司将成为 AI 世界中的基础类公司。

3. 模型公司将对科技公司竞争格局带来颠覆

3.1 OpenAI 的前世今生

OpenAI 成立于 2015 年 12 月，是所处于美国旧金山的一个人工智能研究实验室，由非营利性的 OpenAI Inc. 及其营利性的子公司 OpenAI LP 组成。OpenAI 开展人工通用智能（AGI）研究，为了确保 AI 能够造福全人类，OpenAI 提供了一个基于 AI 的开发和研究框架，这也是其名字的来源（开放 AI 能力）。利用 OpenAI 的平台，客户可以快速提升开发技能并获取 AI 领域的专业知识，这有助于安全有效的通用人工智能（AGI）技术的发展。OpenAI 开发、维护和训练了一批可用于通用活动的 AI 模型，包括写作、阅读、编程和图像处理等，OpenAI 始终相信 AGI 对我们日常生活的影响将远远超过早期的 AI 技术。

马斯克、奥特曼和其他投资者于 2015 年 12 月宣布创建 OpenAI，并承诺向该项目投入超过 10 亿美元。通过公布其专利和研究成果，OpenAI 将与其他组织和研究人员“自由互动”。“OpenAI Gym”的公测版本于 2016 年 4 月 27 日发布，这是一个加强 OpenAI 研究领域的平台。2016 年 12 月 5 日，OpenAI 发布了“Universe”，这是一个用于开发和测试 AI 的平台，智能能力可以覆盖全球的网站、游戏和其他应用程序。

2019 年 3 月 11 日，OpenAI 宣布从“非盈利（non-profit）”性质过度到“有限盈利（capped for profit）”，利润上限为任何投资的 100 倍（创立了 OpenAI LP 公司）。也是在 2019 年，微软向 OpenAI 投资了 10 亿美金，并获得了 OpenAI 技术的商业化授权。从此，OpenAI 的一些技术开始出现在微软的产品和业务上。不过，OpenAI 与微软的合作其实从 2016 年就开始，2016 年，微软的云服务 Azure 为 OpenAI 提供了大规模实验的平台。Azure 彼时已经为他们提供了带有 InfiniBand 互连的 K80 GPU 的算力资源，以优化深度学习的训练。2020 年 9 月 22 日，OpenAI 开始授权微软使用他们的 GPT-3 模型，也是全球首个可以享受 GPT-3 能力的公司。

2020 年 6 月 11 日，OpenAI 发布了 OpenAI API，这也是 OpenAI 第一个商业化产品。官方解释了，他们认为开发商业产品是确保 OpenAI 有足够资金继续投入 AI 研究的有效手段。自此，OpenAI 也正是开始商业化运作。官方也解释了，使用 API 的方式提供模型而不是开源模型也将降低模型的使用门槛，毕竟对于中小企业来说，部署强大的 AI 模型所需要的成本可能更高。

2016 年至今，OpenAI 发布了很多人工智能相关的技术，从工具到算法到论文到模型，都有涉及。

2016 年

2016 年 4 月 27 日，OpenAI 发布了他们的第一个项目——OpenAI Gym Beta，这是一个用来开发和比较不同强化学习算法的工具。

2017 年

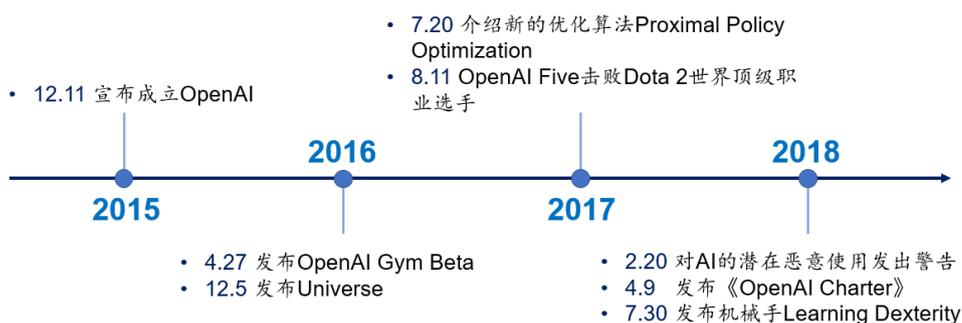
2017 年 5 月 24 日，OpenAI 开源了一个重现强化学习算法的工具——OpenAI Baselines。目标是提供用于正确的强化学习算法实现的一些最佳实践，以帮助大家提高强化学习的研究效率。

2018 年

2018 年 6 月 11 日，OpenAI 公布了一个在诸多语言处理任务上都取得了很好结果的算法，即著名的 GPT，这也是该算法的第一个版本。GPT 是第一个将 transformer 与无监督的预训练技术相结合，其取得的效果要好于当前的已知算法。这个算法算是 OpenAI 大语言模型的探索性的先驱，也使得后面出现了更强大的 GPT 系列。

同样是在 2018 年 6 月份，OpenAI 的 OpenAI Five 已经开始在 Dota2 游戏中击败业余人类团队，并表示在未来 2 个月将与世界顶级玩家进行对战。OpenAI Five 使用了 256 个 P100 GPUs 和 128000 个 CPU 核，每天玩 180 年时长的游戏来训练模型。在随后的几个月里 OpenAI Five 详情继续公布。在 8 月份的专业比赛中，OpenAI Five 输掉了 2 场与顶级选手的比赛，但是比赛的前 25-30 分钟内，OpenAI Five 的模型的有着十分良好的表现。OpenAI Five 继续发展并在 2019 年 4 月 15 日宣布打败了当时的 Dota2 世界冠军。

图15 OpenAI 发展时间线（2015-2018）



资料来源：公开资料整体，HTI

2019 年

2019 年 2 月 14 日，OpenAI 官宣 GPT-2 模型。GPT-2 模型有 15 亿参数，基于 800 万网页数据训练。2019 年 11 月 5 日，15 亿参数的完整版本的 GPT-2 预训练结果发布。

2019 年 3 月 4 日，OpenAI 发布了一个用于强化学习代理的大规模多代理游戏环境：Neural MMO。该平台支持在一个持久的、开放的任务中的存在大量的、可变的代理。

2019 年 4 月 25 日，OpenAI 公布了最新的研究成果：MuseNet，这是一个深度神经网络，可以用 10 种不同的乐器生成 4 分钟的音乐作品，并且可以结合从乡村到莫扎特到披头士的风格。这是 OpenAI 将生成模型从自然语言处理领域拓展到其它领域开始。

2020 年

2020 年 4 月 14 日，OpenAI 发布了 Microscope，这是一个用于分析神经网络内部特征形成过程的可视化工具，也是 OpenAI 为了理解神经网络模型所作出的努力。

2020 年 5 月 28 日，OpenAI 正式公布了 GPT-3 相关的研究结果，其参数高达 1750 亿，这也是当时全球最大的预训练模型，同年 9 月，GPT-3 的商业化授权给了微软。

2020年6月17日，OpenAI发布了Image GPT模型，将GPT的成功引入计算机视觉领域。

2021年

2021年1月5日，OpenAI发布CLIP，它能有效地从自然语言监督中学习视觉概念。CLIP可以应用于任何视觉分类基准，只需提供要识别的视觉类别的名称，类似于GPT-2和GPT-3的"zero-shot"能力。

2021年1月5日，OpenAI发布了DALL·E模型，其为120亿个参数的GPT 3版本，它被训练成使用文本-图像对的数据集，从文本描述中生成图像。

2021年8月10日，OpenAI发布了Codex。OpenAI Codex同样是GPT 3的后代；它的训练数据既包含自然语言，也包含数十亿行公开的源代码，包括GitHub公共存储库中的代码。OpenAI Codex就是Github Copilot背后的模型。

2022年

2022年1月27日，OpenAI发布了InstructGPT。这是比GPT 3更好的遵循用户意图的语言模型，同时也让它们更真实，且less toxic。

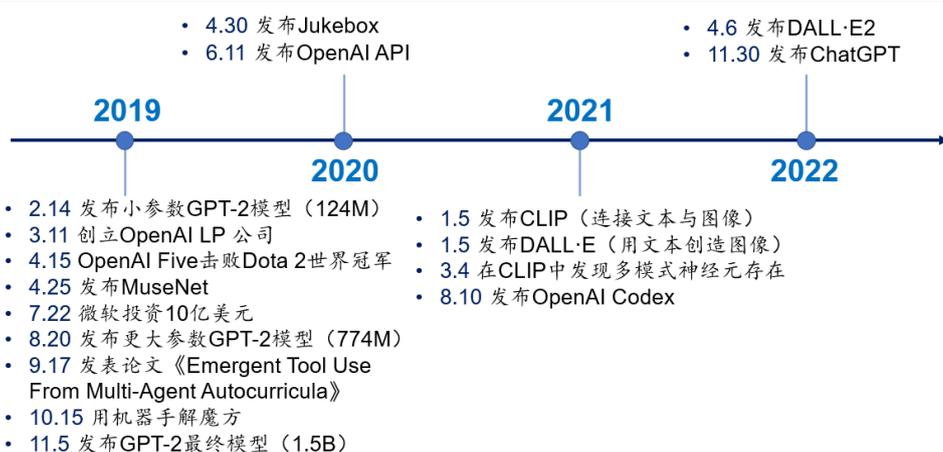
2022年3月15日，OpenAI新版本的GPT-3和Codex发布，新增了编辑和插入新内容的能力。

2022年4月6日，DALL·E2发布，其效果比第一个版本更加逼真，细节更加丰富且解析度更高。

2022年6月23日，OpenAI通过视频预训练（Video PreTraining, VPT）在人类玩Minecraft的大量无标签视频数据集上训练了一个神经网络来玩Minecraft，同时只使用了少量的标签数据。通过微调，该模型可以学习制作钻石工具，这项任务通常需要熟练的人类花费超过20分钟（24,000个动作）。它使用了人类原生的按键和鼠标运动界面，使其具有相当的通用性，并代表着向通用计算机使用代理迈出了一步。

2022年9月21日，OpenAI发布了Whisper，这是一个语音识别预训练模型，结果逼近人类水平，支持多种语言。

图16 OpenAI 发展时间线（2019-2022）



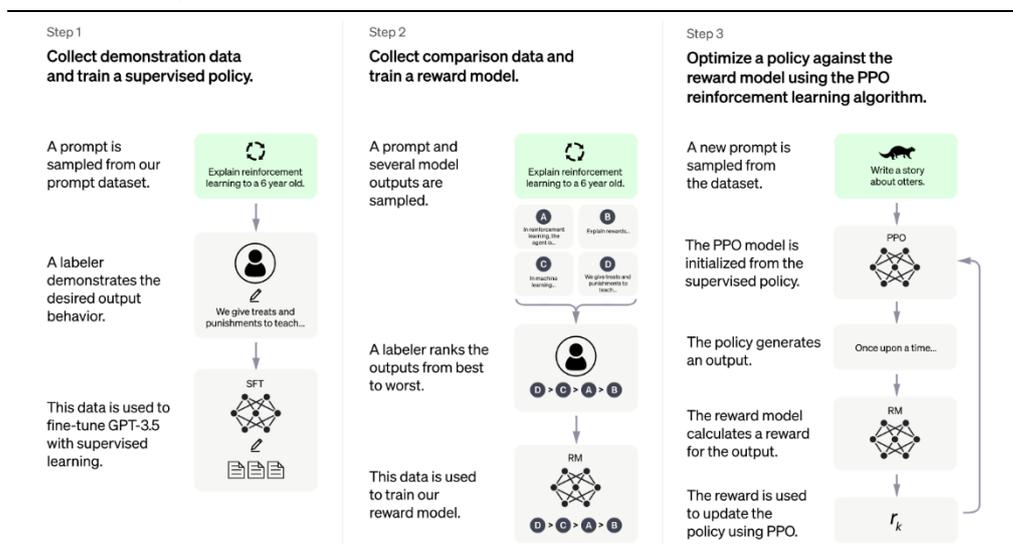
资料来源：公开资料整体，HTI

3.2 ChatGPT 的横空出世，对搜索领域或带来巨大冲击

2022年11月30日，OpenAI发布ChatGPT，这是一个AI对话系统，一款人工智能技术驱动的自然语言处理工具。它能够通过学习 and 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。

ChatGPT经历了OpenAI开发的四代GPT模型的进化。此前的三代模型数据质量和数据规模不断提升，使得其生成能力不断精进，已经能够执行阅读理解、机器翻译、自动问答等任务，但本质上只是语言模型，不具备回答问题的能力。针对GPT-3，OpenAI引入了1750亿的训练参数，开启了超大模型时代，专家普遍认为，在封闭、静态和确定性环境中，该模型已经可以达到人类的决策水平。而ChatGPT模型基于GPT-4优化，引入了新的算法——从人类反馈中强化学习（RLHF），在训练中，训练师会对答案进行排序、打分或者给出高质量答案，令ChatGPT具备一定逻辑和常识，成为现阶段全球所发布的功能最全面的AI模型，远超同类产品的智能化水平。

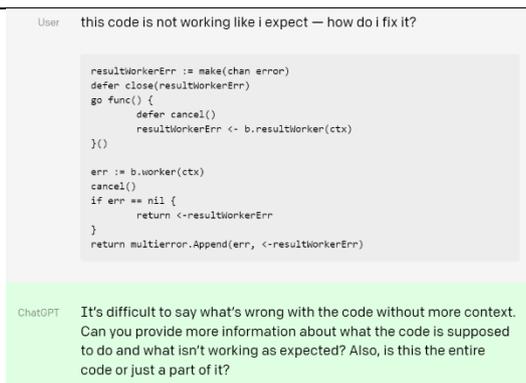
图17 ChatGPT 实现机理



资料来源：OpenAI 官网，HTI

正由于“通过文本与人对话”这一特点，ChatGPT 获得了得天独厚更为灵活的使用功能。以前也曾有其他 AI 聊天机器人，但没有一款能与 ChatGPT 相提并论。ChatGPT 能进行天马行空的长对话，可以回答问题，还能根据人们的要求撰写各种书面材料，例如商业计划书、广告宣传材料、诗歌、笑话、计算机代码和电影剧本等。

图18 ChatGPT 协助用户进行代码修改



资料来源：ChatGPT 官网，HTI

图19 ChatGPT 解答生活问题



资料来源：ChatGPT 官网，HTI

正是由于 ChatGPT 优秀的使用体验，使得其用户量获得了快速增长，发布短短 5 天，其用户数就超过 100 万，而 Facebook 获得 100 万用户，则用了整整 10 个月。**2023 年 1 月，ChatGPT 的月活跃用户数预计已达 1 亿，成为历史上用户增长最快的消费应用。**相比之下，达到 1 亿月活跃用户数，TikTok 花了约 9 个月的时间，Instagram 则花了两年半。在高技术力、优质体验、巨大客户群的背景下，ChatGPT 背后的模型公司 OpenAI 开始对科技公司产生巨大的冲击，而首当其冲的就是搜索领域的各大巨头。

在 ChatGPT 发布的第一天，就有生成式 AI 将取代搜索引擎的声音出现。ChatGPT 虽然常常被用于信息查找，但它本质上属于 AIGC（人工智能生产内容），是一种文本创造方式；而搜索引擎的思路是信息查找与整合，无法创造文本。也正由于这一特点，ChatGPT 能够给出更加“人性化”的交互问答服务。

ChatGPT 相对于传统搜索引擎具有难以替代的优势。输入端，ChatGPT 能够通过文本对话的方式与用户进行深入交流，从而更加明确用户需求，给出恰当的回复；相较之下，用户使用传统搜索引擎只能通过反复修改搜索关键词进行深入查找，过程繁复，且常常很难得到满意的回复。输出端，ChatGPT 更贴近“人性化”的输出，通过生成文本的方式模拟对话给出确定的答案；传统搜索引擎更贴近“搜索”，只能给出大量网站链接，需要用户自行甄别信息。

这就使得，ChatGPT 可能对搜索带来一个重大变化：用户将会转向聊天机器人寻求帮助，而不是通过谷歌提供的网站进行过滤。ChatGPT 有可能改变用户的搜索方式。与用户在搜索引擎（谷歌）中尝试许多不同的搜索词并查看结果列表（和广告）不同，大型语言模型可以做不同的事情。用户可以用正常的句子问语言模型一个问题，它可以用一段漂亮的文字进行总结，并提供使用过的源网页链接。这种方法的优点是用户不再需要阅读长长的结果列表，**可预计在搜索领域，ChatGPT 带来了一个破坏性创新，有望颠覆整个搜索领域的商业模式。**

根据 Statcounter，截止 2023.01，谷歌在搜索引擎市场的市占率达到了 92.09%，但是，一旦 ChatGPT 对“搜索”这一行为所带来的破坏性创新不断被用户接受，未来可能用户都不会再去使用“搜索引擎”去进行“搜索”这一行为了，**就类似马车被汽车替换一样，虽然初期汽车不如马车，但是随着不断完善，汽车将最终替代马车，即使某家公司在马车领域市占率再高，但是马车市场的空间将不断缩减直至最终消失。**对于谷歌来说，即使在搜索引擎领域，谷歌的市占率再高，但是如果 ChatGPT 使得整个搜索引擎市场的流量将不断下降，这对于 2022Q4 广告营收（主要来源于巨大流量）占比高达 78% 的 Alphabet 来说，将会带来巨大的冲击。

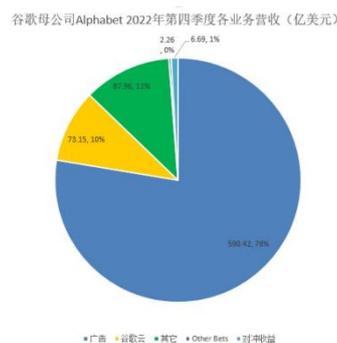
图20 主流搜索引擎市场占比（数据截止 2023.01）

主流搜索引擎市场占比	
名称	市场占比
谷歌	92.9%
必应	3.03%
Yahoo	1.22%
YANDEX	0.85%
Baidu	0.65%
DuckDuckGO	0.58%

数据来源 | Statcounter (截至2023年1月)

资料来源: Statcounter, HTI

图21 谷歌母公司 Alphabet 第四季度各业务营收占比



资料来源: Alphabet 2022 四季报, HTI

同时，从技术上来讲，ChatGPT 也可能会降低搜索引擎的门槛。谷歌搜索分很多层，从最开始的用户关键词输入到最后的信息提取，中间有一步叫做用户的关键词改写，因为用户的关键词可能是比较模糊，改写就能够把这个关键词变得更清楚，然后变成一个更加方便去提取信息的方式。如果中间的这一层让各个搜索公司自己管理，但是这些中间层最后都指向 ChatGPT，那么搜索引擎这一个业务的门槛就会降低，任何一个想做搜索的公司都可以基于 ChatGPT 开发自己的搜索引擎。

从 Alphabet 目前的反应来看，也能看出 ChatGPT 对于谷歌造成的冲击。在围绕谷歌 AI 战略的会议中，谷歌 CEO 桑达尔·皮查伊要求公司重新集中精力解决 ChatGPT 对其搜索引擎业务构成的威胁。近日，有消息称，谷歌正在测试一款类似 ChatGPT 的聊天机器人 Apprentice Bard，该产品基于谷歌对话模型 LaMDA，员工可以向其提问并获得类似 ChatGPT 的详细答案。据悉，即便近期刚发生的事，Apprentice Bard 也能回答出来。

2月2日，桑达尔·皮查伊更是表示，该公司很快将在其搜索引擎中添加高级人工智能功能。他预计，谷歌将在“未来几周或几个月”推出类似 ChatGPT 的基于人工智能的大型语言模型。皮查伊在财报电话会议上透露，用户很快就能以“搜索伴侣”的形式使用语言模型。他说，“很快，人们将能够直接与我们最新、最强大的语言模型互动，作为搜索的伴侣，以实验和创新的方式。”“人工智能是我们目前正在研究的最深刻的技术。我们的人工智能之旅才刚刚开始，最好的还在后面。”他补充说。

谷歌更是早在 2022 年 12 月就发布了“红色代码”，以紧急应对 ChatGPT 的崛起所带来的威胁。谷歌也正在要求员工测试 ChatGPT 的竞争对手，作为“红色代码”计划的一部分。

而中国的百度，同样也意识到了 ChatGPT 所带来的巨大挑战。根据报道，百度就计划在 3 月推出类似于 ChatGPT 的 AI 聊天机器人服务，百度文心（Ernie）大模型将成为其基础。百度计划在 3 月首次推出 ChatGPT 式的应用程序，最初将其嵌入百度主要的搜索服务中，该工具的名称尚未确定，将允许用户获得对话式的搜索结果，很像 ChatGPT。经过数年数据训练的大规模机器学习模型文心大模型，将成为百度即将推出的类似 ChatGPT 的工具的基础。

1 月初，百度搜索杰出架构师辜斯缪曾在一次小范围的媒体沟通会上说：“我们预计在 3 月会做一个比较大的升级。”辜斯缪提出搜索领域的三个重点技术趋势：第一是搜索从信息检索到检索+生成的混合系统。第二是搜索变成跨模态的理解和交互，搜索将以图片、视频、音频等各种媒介形式理解用户输入的信息。第三是在知识的理解和组织基础上完成搜索。据辜斯缪介绍，百度搜索今年最重要的工作是要把整个检索系统变成“检索+生成”双模系统。具体从终端用户体验而言，变化主要体现在

三个方面。第一，搜索结果的信息整合推理效果明显提升。第二，搜索有了创作能力。第三，个性化内容体验。

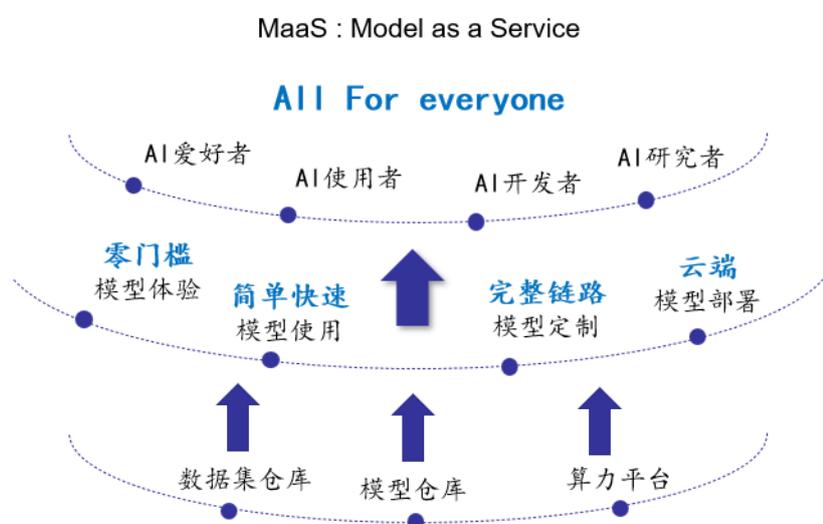
可以看出，百度也在尝试推出自己的“类 ChatGPT”产品，并对自身搜索引擎的功能进行大改，改动的方向同样也是再向“过去的百度引擎+类 ChatGPT”方向去推动。可以说，ChatGPT 已经真真切切地改变了搜索领域，对众多科技公司产生了巨大的挑战。

4. MaaS Model as a Service 模型即服务的产业结构

从 MaaS 整个的产业结构来看，核心应该是“模型→单点工具→应用场景”这样一个路径。

以 ChatGPT 为例，其所用的底层模型就是 GPT-4，而产生出的单点工具就是 ChatGPT，最终 ChatGPT 又能落地在对话、撰写书面材料、修改代码等多个细分的应用场景。

图22 MaaS 基本产业架构



资料来源：阿里云栖大会，HTI

4.1 大模型，MaaS 的最重要基座

以 OpenAI 为例，虽然目前其已经拥有类似于 ChatGPT 这类“单点工具”，但是其主要提供的还是各类大模型，包括：

- 访问执行各种自然语言任务的 GPT-3（GPT-3 以后的 GPT 系列，如 ChatGPT 的底座模型 GPT-3.5 目前不对除微软以外的客户开放使用）
- 将自然语言翻译成代码的 Codex
- 创建和编辑原始图像的 DALL-E

通过这三大模型系统，其他公司可以通过调用 API 实现自身 AI 产品的创造。以 DALL·E 为例，开发人员可以通过 OpenAI 的 API 将 DALL·E 直接集成到他们的应用程序和产品中。超过 300 万人已经在使用 DALL·E 来扩展他们的创造力并加快他们的工作流程，每天生成超过 400 万张图像。开发人员可以在几分钟内开始使用相同的技术进行构建自己的产品。

又例如商汤科技，旗下的 AIDC 就是商汤的大模型。在 WAIC 2022 上，商汤在 AI 大装置基础上衍生出四大平台体系：大装置 AI 云、车路协同、游戏、未来医院。SenseCore 商汤大装置 AI 云是基于 AI 大装置的一套开箱即用的工业级 AI 工具链，而另外三大平台可以理解为商汤“裁剪”出的行业大模型。大装置 AI 云无需前期投入，从工业 AI 流水线到大模型训练、验证、推理，它都能成倍提高研发效率，基础设施费用就大大降低。

图23 商汤大装置运行机理



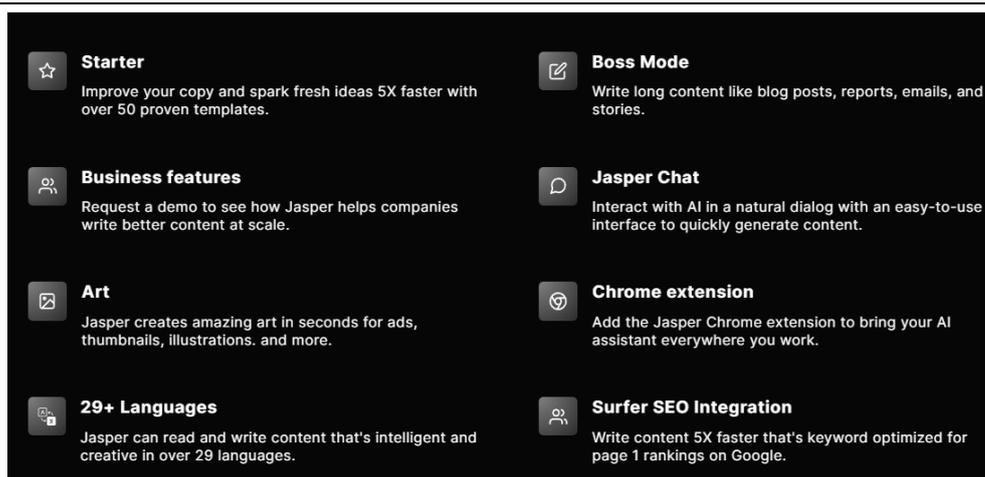
资料来源：SenseCore 官网，HTI

4.2 单点工具，大模型应用的补充

单点工具往往是基于大模型产生的能实际应用的产品，例如 ChatGPT，它就是在 GPT-3.5 模型的基础上，产生出的能“对话”的 AI 系统。

除了 ChatGPT 以外，也有许多类似的单点工具存在，例如 Jasper，其是一个内容写作工具，它使用人工智能为用户编写内容。它基于开源的 GPT-3 模型（同样为 OpenAI 出品），Jasper 声称可以创造 100% 的原创内容，还整合了著名的原创检测的工具 Copyscape。它也可以帮助重写用户不满意的段落，或纠正用户的语法。

图24 Jasper 功能介绍



资料来源: Jasper 官网, HTI

又例如商汤科技基于自身“大装置+大模型”推出的 **OpenXRlab** 扩展现实平台，也是一个**单点工具**。

该平台拥有领先算法，结合大量真人动作数据，商汤打造了一套基于深度学习的动作捕捉及生成解决方案。该方案适配性强、自由度高、易于部署、无需绿幕和专业动捕设备，即可高效率、低成本地实现 AIGC 内容生成。即便在日常直播等低算力设备应用场景，也可进行流畅、高精度的视频创作和实时互动。该方案还可配套提供用于虚拟 IP 生成的大规模、高质量、语义化的 3D 动作库，供企业进行数字人相关的自主高效创作，从而满足企业在虚拟 IP 内容定制、直播运营等活动中的广泛需求。结合企业品牌符号、标签等元素，商汤“虚拟 IP 解决方案”可智能生成符合品牌性格和调性的专属形象，让品牌形象更年轻、更鲜活、更具辨识度和记忆点，打破刻板印象，拉近与用户的距离。

图25 商汤科技为宁波银行创造的“数字人”



资料来源: 阿里云栖大会, HTI

4.3 应用场景，大模型应用的变现

虽说目前仍没有出现适合所有应用场景的“全能共通性”模型，但是单个大模型覆盖的应用场景仍在持续扩展。

以商汤科技的“书生”大模型为例，一个“书生”基模型即可全面覆盖分类、目标检测、语义分割、深度估计四大视觉核心任务。在 ImageNet 等 26 个最具代表性的下游场景中，书生模型广泛展现了很强的通用性，显著提升了这些视觉场景中长尾小样本设定下的性能。

图26 MaaS 基本产业架构

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		阳台社区 大模型探索与探索社区			
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				蛋白质结构预测 HelixFold
	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				单序列蛋白结构预测 HelixFold Single
	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	视觉-语言 ERNIE-VIL -VIL	语言-语言 ERNIE-SAI	地理-语言 ERNIE-Geol
	ERNIE 3.0 Tiny (亿参数)	ERNIE 3.0 (百亿级)	鹏城-百度文心 (1.42B)	ERNIE 3.0 Zeus (千亿级/推理)				

资料来源：百度文心大模型官网，HTI

又例如 GPT 模型系列，该模型本身是一种基于互联网可用数据训练的文本生成深度学习模型，但是该模型的产品，如 ChatGPT、Jasper 等，目前应用场景却持续扩展。例如，Jasper 虽然是文本书写工具，但是内置了 50 多个模板，适用于广告，博客，电商，邮件，SEO，视频，网站等多种使用场景。而 ChatGPT 作为对话式 AI 程序，其使用范围目前更是得到了很大的扩展，修改代码、撰写文字、搜索资源等等，已经成为了重要生产力工具之一。

可以预见，随着未来大模型的发展持续进行，大模型以及其产生的产品的应用场景将会持续扩展。

5. MaaS 将成为人工智能公司的核心商业模式

5.1 订阅制收费

OpenAI 在 2 月 1 日正式官宣了 ChatGPT 的「试点订阅计划」，这项付费服务被命名为 ChatGPT Plus，每月收费 20 美元，订阅者将获得许多好处，包括：

- 在高峰时段享有优先访问 ChatGPT 的权利
- 更快的应用响应时间
- 优先使用新功能和改进

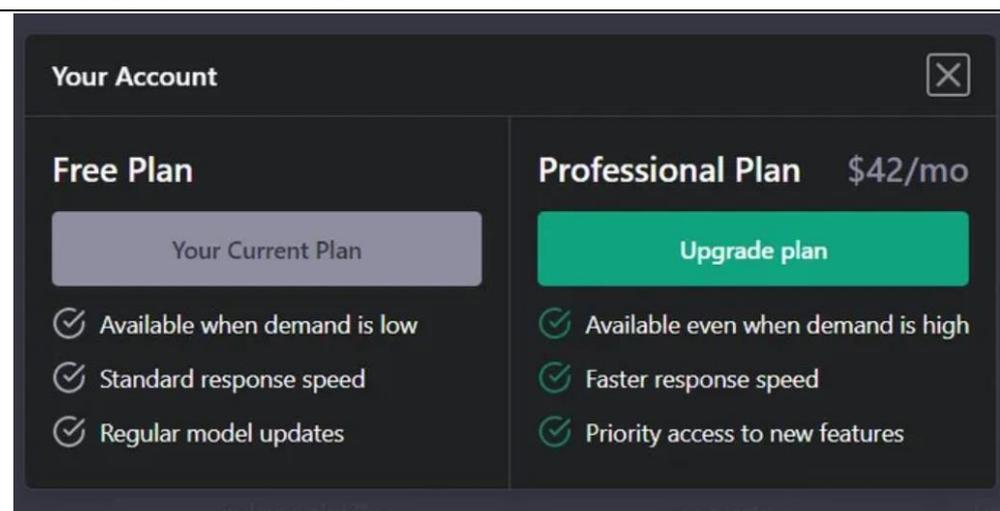
此前推出的 ChatGPT 版本被 OpenAI 称为 Research Preview Launch（研究预览版）。ChatGPT 推出一周，即收获 100 万注册用户，2023 年 1 月，ChatGPT 的活跃用户数或已达 1 亿，它可能是史上增长最快的消费者应用，但这也导致了许多问题，上线首周 ChatGPT 就多次因用户访问量过大导致服务器崩溃。此后为了应对访问量过大而导致的服务器压力，OpenAI 对 ChatGPT 进行了大规模限流，包括禁止来自云服务器的访问，限制每小时的提问数量，以及高峰时段排队等举措。可以看出，ChatGPT Plus 有效的改善了目前免费版面临的诸多痛点。

OpenAI 在声明中表示，ChatGPT Plus 将在未来几周率先在美国推出，之后尽快推行至其他国家。公司表示，有兴趣的用户可以注册 ChatGPT Plus 的等待名单，未来几周会发出邀请。

目前 ChatGPT 已经成为一款重要的生产力工具，被广大用户证明，可以写文案、写小说，写代码、改 bug、查资料，还能帮忙对资料进行归纳总结。所以，收费版的 ChatGPT Plus 的确拥有广泛的市场空间。我们可以做一个简单的计算，假设在目前的 1 亿用户中，有 30% 愿意付费，按照年付费 240 美元计算，年收费就能达到 72 亿美元，如果未来 ChatGPT 作为能够对标 office 的生产工具，付费用户数突破 10 亿人，市场将达到 2000 亿美元以上，而且这还仅仅是按照目前 20 美元一个月的收费来计算的，并没有考虑未来公司可能推出更高价格的订阅计划等，而如果加上未来可能存在的广告等盈利方法，整个市场空间将会更加广阔。

其实，就在之前，有网友就发现 OpenAI 似乎在内测另一付费版 ChatGPT——“ChatGPT Pro”，而这一计划的收费更是高达每月 42 美元。虽然这一计划最终未能正式发布，但是也不排除未来 OpenAI 针对 ChatGPT 推出更多价位的不同计划。

图27 传闻中的 ChatGPT Pro 和普通版 ChatGPT 对比

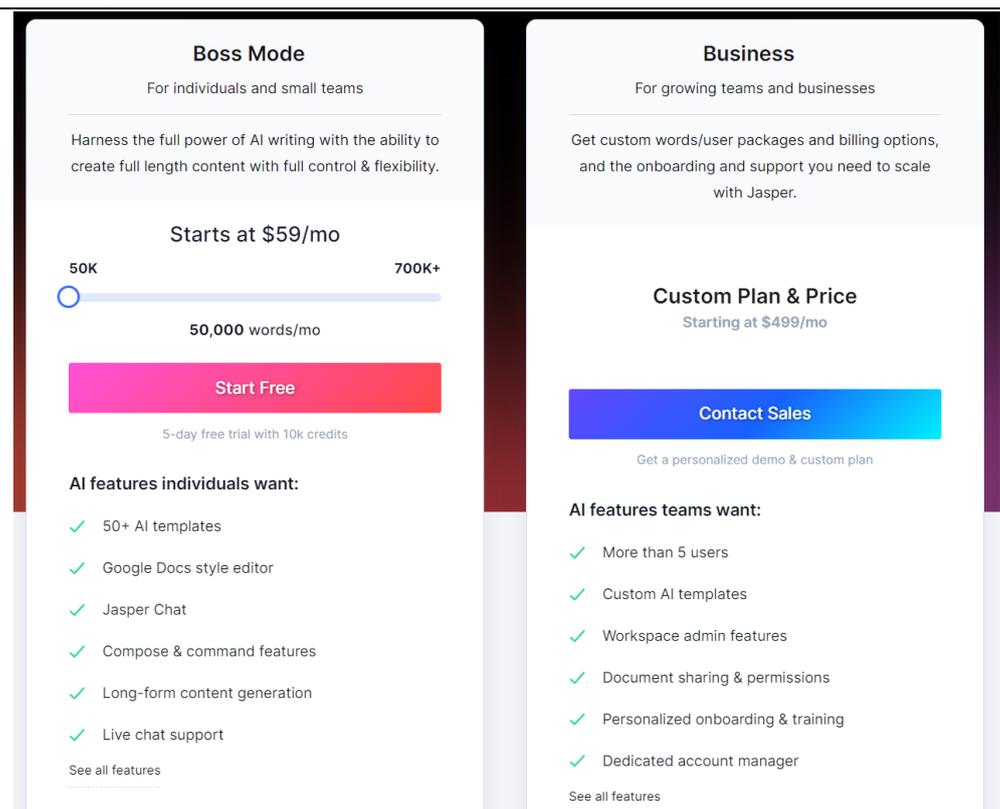


资料来源: Twitter, HTI

而除了 ChatGPT, 其他 MaaS 公司同样推出了许多商业订阅计划。以类似的 AI 聊天公司 Jasper 为例, 该公司产品 Jasper 以 GPT-3 (同样为 OpenAI 产品, 但已经开源) 为基座, 通过在底层调用 GPT-3 模型研发出属于自己的文本生成 AI 模型, 并推出了多项订阅服务。

其中, 订阅服务的收费主要分为三档: Starter 档的收费标准 29 美元月起, 更为常用的 Boss Mode 会员则是 59 美元/月起, 另外还有更高昂的企业级收费。

图28 Jasper 公司收费模式



资料来源: Twitter, HTI

即使是 59 美元/月的 Boss Mode 最基础档，其也只提供 5 万字/月的服务，对比 ChatGPT 不限量的生成字数，我们认为，ChatGPT 这类产品的订阅收费未来的天花板远不止 20 美元，这背后自然是更为广阔的蓝海市场。

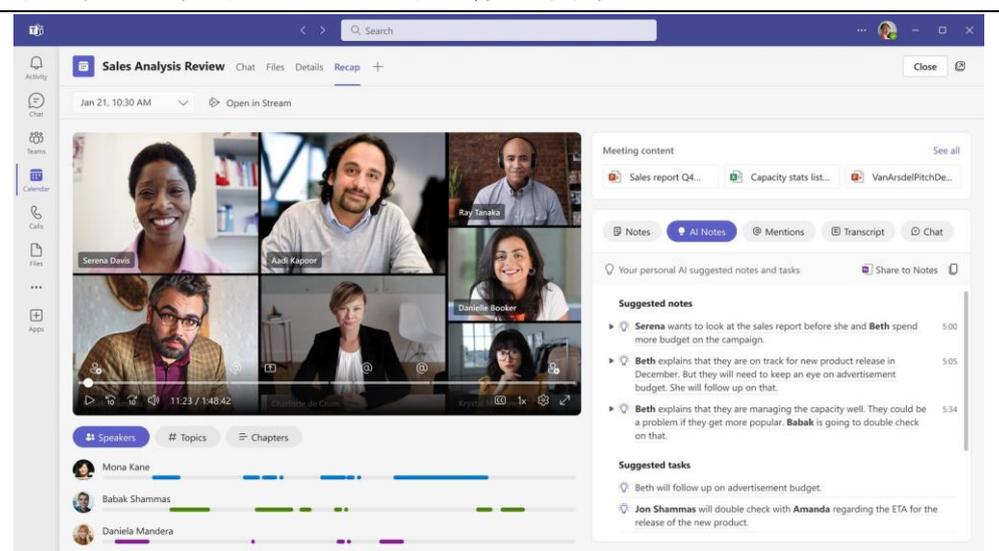
5.2 嵌入其他产品获得引流式收入

而除了模型本身进行订阅收费以外，MaaS 目前也在积极尝试其他各类收费模式。

例如，2 月 1 日，微软也宣布推出其视频会议及远程协作平台的高级付费版 **Microsoft Teams Premium**，这项高级服务将在 6 月份每月收费 7 美元，然后在 7 月份恢复到正常的 10 美元。

订阅者可享受由 **OpenAI GPT-3.5**（支撑 ChatGPT 的底层模型）提供支持的大型语言模型技术，即“智能回顾（Intelligent Recap）”这一重要功能，该功能将提供自动生成的会议记录、推荐任务和个性化标亮。

图29 智能回顾将会为 Teams Premium 提供许多重要的新功能



资料来源：微软官网，HTI

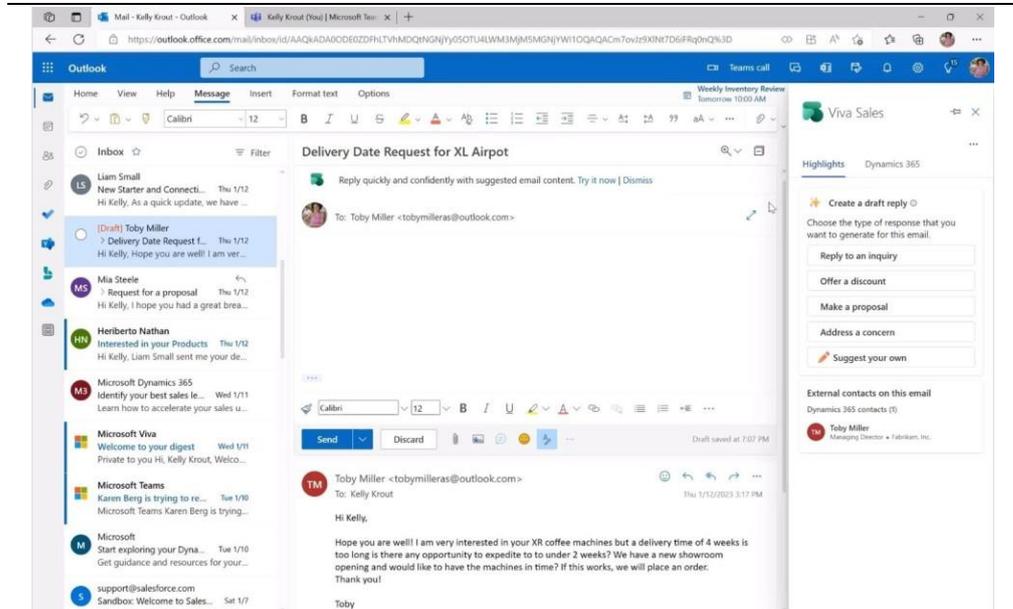
基于 OpenAI 的 GPT-3.5 模型，即使用户没有参加会议，「智能回顾」也能生成会议记录和要点。当会议在尴尬的时间举行时，或者当员工在不同的时区以及不在办公室时，这项功能使紧张的工作变得更容易了。笔记、提到的内容和完整的文字记录都是可用的，每个发言者的贡献都会在一个整齐的主题和章节的时间轴上突出显示。

如果有同事在会议中提到用户，用户还会得到个性化的时间线标记，以快速查看分享和讨论的内容。虽然大多数智能回顾功能目前已经可以使用，但微软表示有些功能要到 2023 年第二季度才会出现。

此外，2 月 2 日，微软旗下 Dynamics 365 产品线（ERP+CRM 程序）发布视频，宣布旗下客户关系管理软件 Viva Sales 也将集成 OpenAI 的技术，通过人工智能帮助销售人员完成许多繁杂且重复的文字工作。

利用 OpenAI 的 GPT-3.5 模型，Viva Sales 可以为电子邮件里客户的问题生成回复建议。该产品在 Outlook 电子邮件客户端中提供了创建特定回复的选项。例如，销售人员可以从“提供折扣”（Offer a discount）、“回复询问”（Reply to an inquiry）或“自行创建”（Suggest your own）等选项中进行选择，然后人工智能将创建可使用的内容。AI 程序从客户记录和 Office 电子邮件软件中提取数据，将它们用于生成个性化文本、定价细节和促销信息的电子邮件。

图30 GPT-3.5 将帮助 Viva Sales 用户自动编写邮件回复



资料来源：微软官网，HTI

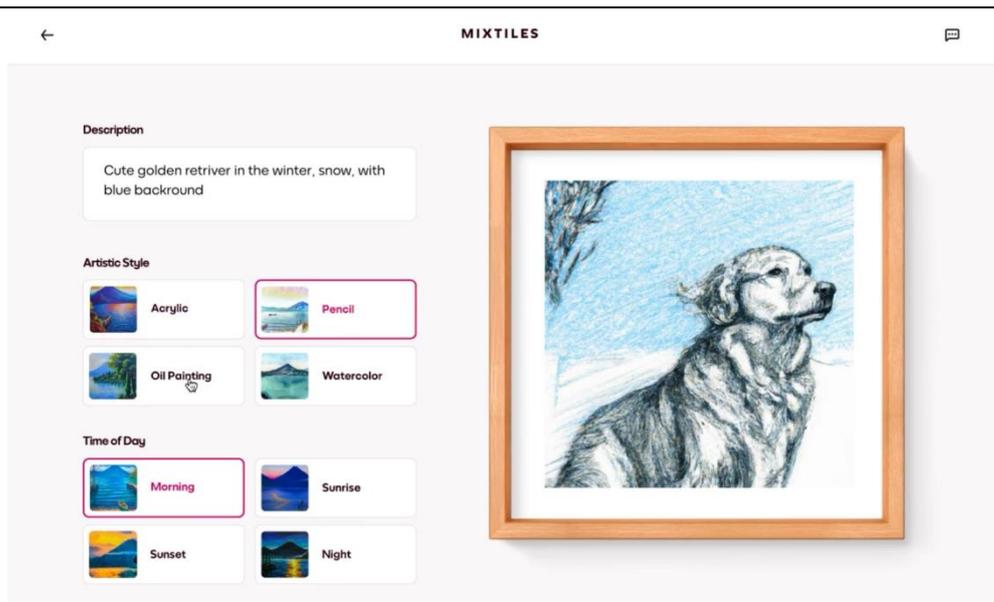
虽然，GPT-3.5 并未在 Viva Sales 中作为单独的收费项目存在，但是 Viva Sales 作为 CRM 产品本身每月就收费 40 美元，假设有 1000 万客户是因为 GPT-3.5 而选购 Viva Sales，则代表 GPT-3.5 每年能够为微软带来 48 亿美元的收入。

5.3 不同的 Model 应对不同的客户需求和客户定价

仍以 OpenAI 为例，在推出 ChatGPT 以前，其主要收入来源于对外部客户开源自身大模型，通过提供 API 接口调用来收费。

例如用于创建和编辑原始图像的 DALL·E 模型，就有大量对 AI 生成图像这一功能存在需求的公司选择调用该模型制造自身的产品，如 Mixtiles，其是一家发展迅速的照片创业公司，其利用 DALL·E API 来创建和构建能引起情感共鸣的艺术品（图片），通过引导用户完成捕捉童年记忆、梦想目的地等的创作过程。

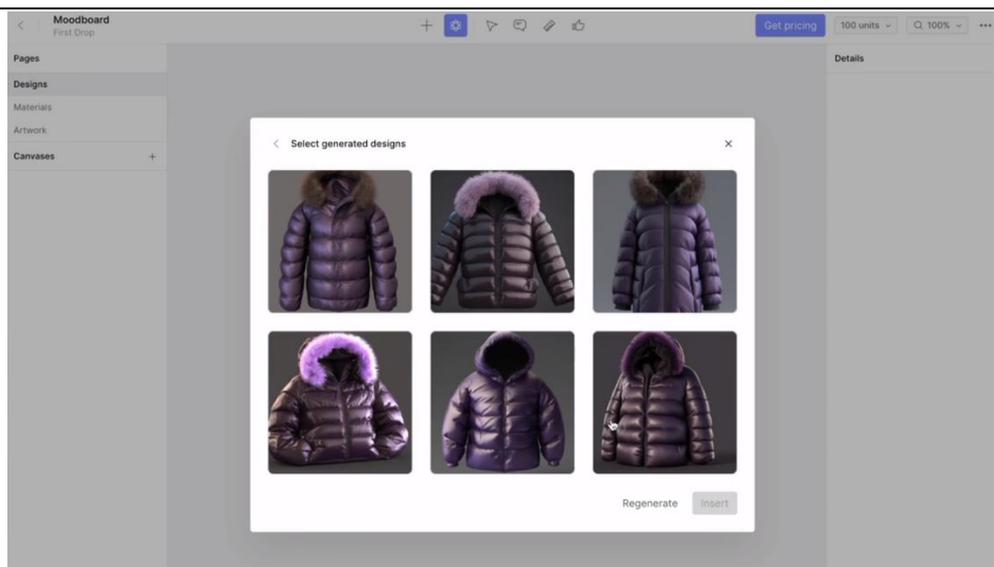
图31 GPT-3.5 将帮助 Viva Sales 用户自动编写邮件回复



资料来源：OpenAI 官网，HTI

而 CALA 同样调用了 DALL.E 模型，但是 CALA 作为数字平台，将整个设计流程——从产品构思一直到电子商务支持和订单履行统一到了一起，DALL.E 仅在这一流程中起到了支持作用，使得 CALA 的智能工具允许用户从自然文本描述或上传的参考图像中生成新的设计理念图。

图32 GPT-3.5 将帮助 Viva Sales 用户自动编写邮件回复



资料来源：OpenAI 官网，HTI

可以看出，相较于 Mixtiles，CALA 对于 DALL.E 模型的应用更偏商业，对于细节的要求也更高，那在这种情况下，CALA 和 Mixtiles 两者调用 DALL.E 模型的收费肯定是有所不同的。

而这都是针对图像生成模型 DALL·E，而文本生成模型的 GPT-3，其收费肯定跟 DALL·E 又有许多的不同。又比如，前文提及的 Jasper 对于 GPT-3 的调用是需要收取费用的，如果未来的 GPT-4 恢复对外商用，那么作为比 GPT-3 更强大的模型，相信其的收费也会高于调用 GPT-3。

总结来说，不同的 Model 应对不同的客户需求和客户定价肯定也会有所不同，客户的要求越高，使用的模型越好，那么大模型的收费也将会更高。

6. 全球重要人工智能 MaaS 公司

6.1 商汤科技：A2.0 时代的全球领导者

商汤科技拥有深厚的学术积累，并长期投入于原创技术研究，不断增强行业领先的全栈式人工智能能力，涵盖感知智能、决策智能、智能内容生成和智能内容增强等关键技术领域，同时包含 AI 芯片、AI 传感器及 AI 算力基础设施在内的关键能力。此外，商汤前瞻性打造新型人工智能基础设施——SenseCore 商汤 AI 大装置，打通算力、算法和平台，大幅降低人工智能生产要素价格，实现高效率、低成本、规模化的 AI 创新和落地，进而打通商业价值闭环，解决长尾应用问题，推动人工智能进入工业化发展阶段。商汤科技业务涵盖智慧商业、智慧城市、智慧生活、智能汽车四大板块，相关产品与解决方案深受客户与合作伙伴好评。

近年来人工智能已经迈入到一个新的阶段，即现在正在进入一个超大模型的时代，**超大模型的训练 pipeline，需要相当大的算力支撑，而这恰恰是商汤的优势所在。**2022 年 1 月 24 日，商汤人工智能计算中心 AIDC 正式启动运营。AIDC 是 SenseCore 商汤 AI 大装置的重要算力基座，其设计峰值算力超过 3740 Petaflops，是亚洲最大的人工智能计算中心之一。商汤人工智能计算中心 AIDC 配合商汤自研的资源管理和调度框架，能够实现资源支持按需调用、动态伸缩，从而在模型训练中，支持不同优先级训练任务的资源调度和资源抢占。

但在商汤 AI 大装置中，AIDC 不到 1/9。**SenseCore 商汤 AI 大装置**是商汤打造的高效率、低成本、规模化的新型人工智能基础设施。SenseCore 商汤 AI 大装置由计算基础设施、深度学习平台、模型层三个架构组成，分别对应算力层、平台层、算法层。SenseCore 商汤 AI 大装置能够大幅降低人工智能生产要素的成本，提高人工智能的生产效率，实现人工智能以自动化、自适应的方式进行生产和落地，全面实现企业服务智能化、城市管理智能化和个人生活智能化，推动人工智能进入工业化发展阶段。商汤 AI 大装置 SenseCore 的全貌，包含了三层：

- 算力层（AI 芯片及处理卡+AIDC+AI 传感器）
- 平台层（模型生产+训练平台+数据平台）
- 算法层（算法工具箱+开源框架）

图33 SenseCore 商汤 AI 大装置组成结构



资料来源：商汤科技官网，商汤集团招股说明书（聆讯后资料集），HTI

区别于“AI 模型”小作坊式的打造，AI 大装置更像流水线工厂，可以实现不同场景的算法模型的底层抽象，以模块化平台套件打造通用型服务平台。特别是针对 AI 落地中更长尾的客户和场景，能够在组合不同算法套件的基础上完成新场景的定制，以低边际成本实现对新场景的规模化覆盖。并且商汤基于 AI 平台基础设施的系统协同性，可以以半自动化、自适应的方式，就能实现批量的算法模型生产和迭代升级。

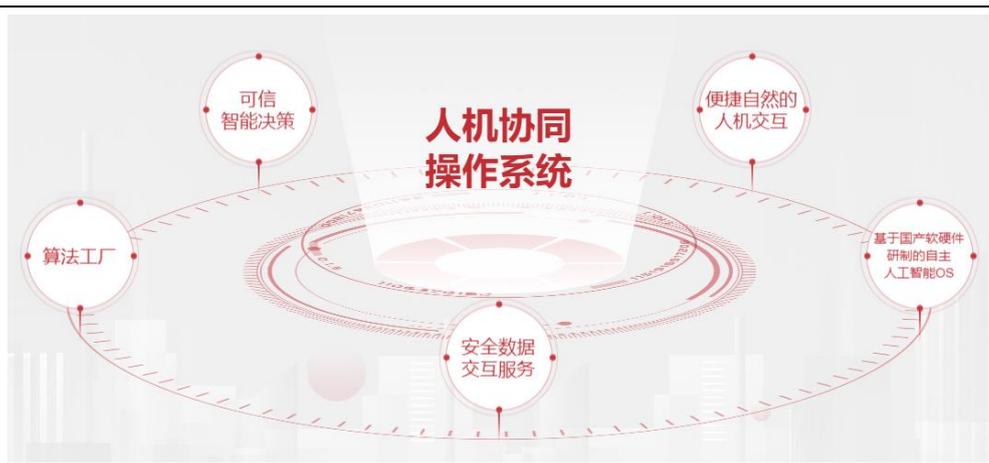
对于商汤来讲，一方面商汤的超算中心有大量的算力支撑，能够更有效做数据挖掘。另一方面是有超大模型，能够快速生成高精度的伪标签，以及赋能业务端的模型。这样在超大模型跟超算中心的加持下，可以更好地解决这些数据闭环里特定场景的问题。截至 2021 年中，商汤已建成世界上最大的计算模型，该模型拥有超过 300 亿个参数。这能够帮助公司实现快速模型迭代，从而为公司大模型在各行各业的持续落地奠定坚实的基础。

6.2 云从科技：AI 人工智能基础模型赋能应用

云从科技成立于 2015 年，是一家提供高效人机协同操作系统和行业解决方案的人工智能企业。公司一方面凭借着自主研发的人工智能核心技术打造了人机协同操作系统，通过对业务数据、硬件设备和软件应用的全面连接，把握人工智能生态的核心入口，为客户提供信息化、数字化和智能化的人工智能服务；另一方面，公司基于人机协同操作系统，赋能智慧金融、智慧治理、智慧出行、智慧商业等应用场景，为更广泛的客户群体提供以人工智能技术为核心的行业解决方案。公司秉持开放的人机协同发展理念，与上下游独立硬件供应商、独立软件供应商、平台及行业应用提供方共建共享。公司的人机协同操作系统一方面包括根据不同应用场景客户需求提供的多种人机协同应用产品和整体操作系统，服务于客户单点业务效能提升和整体业务场景智能化升级；另一方面也包括轻量化且功能全面的“轻舟平台”（即通用服务平台），通过开放式地引入生态伙伴共同开发 AI 应用及配套 SaaS 服务，使人工智能服务惠及更广泛的各行业客户。根据财经网，全世界 AI 头部公司基本都在 10 亿以上收入水平，中国 AI 领域平台型企业只有 2-3 家。公司同时满足了“10 亿收入”与“平台型企业”，作为业内稀缺的 AI 平台型龙头企业，未来发展前景广阔。

公司的研究团队高度认同“预训练大模型+下游任务迁移”的技术趋势，从 2020 年开始，已经陆续在 NLP、OCR、机器视觉、语音等多个领域开展预训练大模型的实践，不仅进一步提升了公司各项核心算法的性能效果，同时也大幅提升了公司的算法生产效率，已经在城市治理、金融、智能制造等行业应用中体现价值；其次，公司一直以来都在人机协同领域布局，打造了像人一样思考和工作的人机协同操作系统（CWOS），致力于整合打通视觉、语音、NLP 等多个领域的大模型，不止于像 ChatGPT 那样在文本世界实现超级智能，还要彻底打通数字世界和物理世界，为“像人一样思考和工作”打下坚实的技术基础；公司也通过开放的人机协同操作系统实现了技术平台化，加上多年的行业深耕，能够通过“平台化的通用模型+带行业 knowhow 的专用模型”来帮助各行各业快速实现智能化升级。公司认为，大模型给 AI 行业带来了巨大的想象空间和市场机会，也是公司的巨大想象空间和市场机会。

图34 云从科技人机协同操作系统五大特色



资料来源：云从科技官网，HTI

此外，根据公司招股说明书，公司的股权结构全由内资组成，而纯内资的背景和技术实力的领先，也使得公司承担了多项国家级项目。2022 年 10 月，科技部正式批复了最新一批国家人工智能开放创新平台名单，支持云从科技建设视听交互国家新一代人工智能开放创新平台，这是人工智能行业第一个融合多种技术的智能交互平台。此次云从科技承建该平台也是继国家发改委“人工智能基础资源公共服务平台”和“高准确度人脸识别系统产业化及应用项目”之后的又一国家级重大项目建设任务。这是云从科技全力打造的人机协同操作系统重要组成部分，根据规划，平台总体建设目标是要打造国际领先的视听交互支撑平台，形成视听交互技术创新领域的策源地、成果转化的助推器和开放服务体制机制的试验田。在自主可控受重视度不断提升的当下，公司这类纯内资的 AI 领军企业有望成为重要的 AI 基础设施提供商。

6.3 科大讯飞：有望在预训练认知智能大模型上代表中国实现突破

科大讯飞股份有限公司成立于 1999 年，是亚太地区知名的智能语音和人工智能上市企业。自成立以来，一直从事智能语音、自然语言理解、计算机视觉等核心技术研究并保持了国际前沿技术水平；积极推动人工智能产品和行业应用落地，致力让机器“能听会说，能理解会思考”，用人工智能建设美好世界。作为技术创新型企业，科大讯飞坚持源头核心技术创新，多次在语音识别、语音合成、机器翻译、图文识别、图像理解、阅读理解、机器推理等各项国际评测中取得佳绩。两次荣获“国家科技进步奖”及中国信息产业自主创新荣誉“信息产业重大技术发明奖”，被任命为中文语音交互技术标准工作组组长单位，牵头制定中文语音技术标准。

科大讯飞坚持“平台+赛道”的发展战略。基于拥有自主知识产权的核心技术，2010年，科大讯飞在业界发布以智能语音和人机交互为核心的人工智能开放平台——讯飞开放平台，为开发者提供一站式人工智能解决方案。截至2022年5月31日，讯飞开放平台已开放493项AI产品及能力，聚集超过337.3万开发者团队，总应用数超过150.1万，累计覆盖终端设备数35.1亿+，AI大学堂学员总量达到69.9万，链接超过420万生态伙伴，以科大讯飞为中心的人工智能产业生态持续构建。科大讯飞在智能语音和人工智能核心研究和产业化方面的突出成绩得到了社会各界和国内外的广泛认可，作为“中国人工智能国家队”一员已形成共识。

2018年起，预训练模型范式就逐步开始在自然语言处理领域应用，科大讯飞也正是在那一时期推出了预训练模型。科大讯飞人工智能高级阶段——认知智能领域的首个国家重点实验室。2022年在国家重点实验室重组后的首批20个国家标杆实验室中，讯飞是唯一一家承建认知智能国家重点实验室。同时，讯飞面向认知智能领域陆续开源了6大类、超过40个通用领域的系列中文预训练语言模型，开源3年模型库月均调用量超1000万，成为业界最广泛流行的中文预训练模型之一，在Github平台的中文预训练模型的星标数达13346位列第一（第二名为6351）。2022年，讯飞进一步发布了面向多模态领域的两个轻量级预训练模型，实现多尺度层级预训练、有监督强化聚类表征、多预训练任务联合优化框架等创新点，在参数量远小于业界公开模型（参数量小20倍以上）的情况下识别效果提升了20~30%，效果上实现了业界领先，同时也获得全球多模态阅读理解评测冠军等优异成绩。

公司有信心在预训练认知智能大模型上代表中国实现突破。首先，科大讯飞承建了中国唯一的认知智能国家重点实验室，具备算法实现和持续创新的能力，已经在众多国际比赛中夺得冠军；其次，公司的训练服务器、推理服务器在国产自主平台运行至关重要+；第三，预训练模型在专业领域，诸如在教育领域，医疗领域，如果能够和行业场景深度结合，潜在商业价值的机会空间巨大。

图35 讯飞工业级预训练模型已在多个技术方向实现了行业落地应用



资料来源：科大讯飞 2022 全球 1024 开发者节官网，HTI

在人工智能的特定领域如智能语音、机器翻译、OCR 等方面，科大讯飞一直保持国际领先，并相继在知识推理阅读理解比赛 OpenBookQA、QASC 中夺冠，超过人类平均水平，同时在逻辑推理阅读理解比赛 ReClor 中刷新全球最好成绩。第二，在教育、医疗等关乎到行业知识和专业理解能力的专业领域，科大讯飞持续形成“从 0 到 1”的突破，形成综合优势。例如，在医疗领域，从基层医生看病的“从 0 到 1”，到三甲医院医生 AI 助手的“从 0 到 1”，不断突破。与其他大厂相比，科大讯飞在典型行业赛道上的相关核心技术始终保持领先，而且不但保持领先，同时能形成商业变现，形成了从核心技术，到合作资源，到数据迭代的综合优势。第三，认知智能方面，科大讯飞将是这一轮技术突破的受益者。科大讯飞在国家平台上有唯一的认知智能国家重点实验室。在教育医疗这些民生领域，国际大厂难以进入中国，科大讯飞一直持续在该领域深耕，已经具备这样的技术、人才，特别是具备品牌影响力和公信力。在未来的技术红利兑现中，科大讯飞将是显著的受益者。

公司在 2022 年启动了讯飞超脑 2030 计划，是一个规划并指引公司中长期的技术与产品发展路线的重大项目，构建基于认知智能的人机协作自我进化的复杂智能系统，让机器人的感知能力能够适应各种复杂场景，深度理解领域专业知识，具备持续自主学习的能力。即让机器全方面感知人和环境，懂各行各业知识，有通识和情感，能灵活运动，会多维表达，打造可持续自主进化的复杂智能系统，助力家庭陪伴机器人走进千家万户。

6.4 百度：文心大模型持续赋能各产业智能化升级

2021 年 12 月，百度发布了全球首个知识增强千亿级大模型——鹏城-百度·文心大模型，产业级知识增强“文心大模型”系列大模型也正式对外。百度文心大模型正从技术创新和加速产业应用两方面，推动中国 AI 发展更进一步。

2022 年 11 月，百度发布了文心大模型的最新升级，包括新增 11 个大模型，大模型总量增至 36 个，构建起业界规模最大的产业大模型体系，并通过大模型工具与平台的升级和文心一格、文心百中等基于大模型技术的产品应用，进一步降低大模型产业化门槛，让更多企业和开发者步入 AI 应用的新阶段。

目前，文心大模型已经拥有了：

文心·NLP 大模型：面向语言理解、语言生成等 NLP 场景，具备超强语言理解能力以及对话生成、文学创作等能力。

文心·CV 大模型：基于领先的视觉技术，利用海量的图像、视频等数据，为企业和开发者提供强大的视觉基础模型，以及一整套视觉任务定制与应用能力。

文心·跨模态大模型：基于知识增强的跨模态语义理解关键技术，可实现跨模态检索、图文生成、图片文档的信息抽取等应用的快速搭建，落实产业智能化转型的 AI 助力。

文心·生物计算大模型：融合自监督和多任务学习，并将生物领域研究对象的特性融入模型。构建面向化合物分子、蛋白分子的生物计算领域预训练模型，赋能生物医药行业。

行业大模型：文心大模型与各行业企业联手，在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36 文心大模型全景图

产品与社区	文心一格 AI艺术创作辅助平台		文心百中 大模型驱动的产业级搜索系统		赐谷社区 大模型探索与探索社区	
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	
	深蓝-百度文心		吉利-百度文心		泰康-百度文心	
	TCL-百度文心		薛海-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	
	对话 PLATO		搜索 ERNIE-Search		信息抽取 ERNIE-UIE	
	跨语言 ERNIE-M		代码 ERNIE-Code		图网络 ERNIE-Sage	
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		
ERNIE 3.0 Tiny (亿参数)		ERNIE 3.0 (百亿级)		鹏城-百度文心 (千亿级)		
ERNIE 3.0 Zeus (万亿级)		视觉-语言 ERNIE -VIL		语言-语言 ERNIE -SAI		
		地理-语言 ERNIE -Geol		文档智能 ERNIE-Layout		
				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		

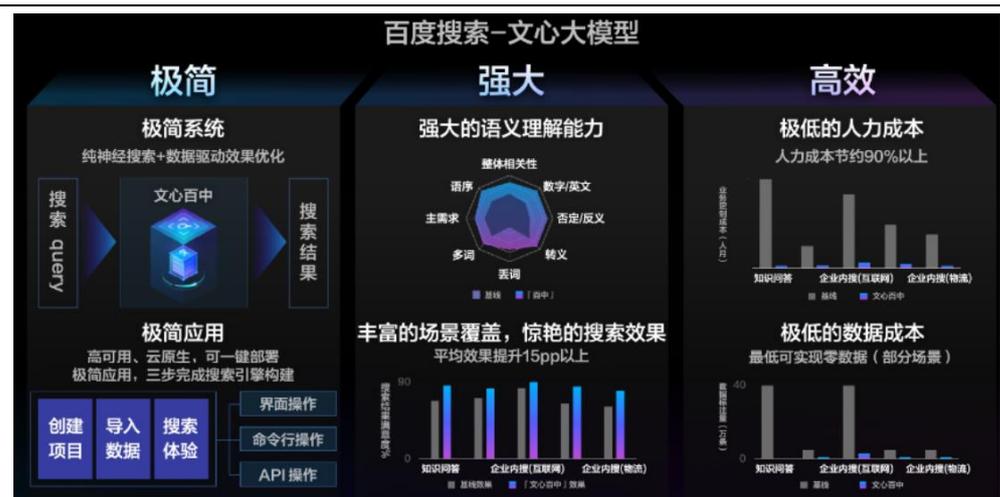
资料来源：百度文心大模型官网，HTI

为了让开发者更加方便、快速地使用大模型的最新能力，文心大模型工具与平台进行了全面更新。在工具方面，文心大模型开发套件全面升级，发布行业大模型定制工具，同时提供更加完备的场景化建模工具、更为丰富的模型精调工具，以及大模型可信学习工具，让大模型开发训练更加灵活高效；文心 API 服务支持多种模型尺寸的精调，体验专区提供更多可体验能力，同时基于大规模集群算力分布式推理加速能力，可满足大规模高性能产业级调用。

飞桨（百度深度学习平台）企业版 AI 开发平台 EasyDL 和 BML 也围绕大模型的技术应用进行了全新升级，发布了提供全流程开箱即用的大模型平台能力，加速大模型的产业化落地。EasyDL 零门槛 AI 开发平台预置了各种通用任务场景大模型，可以让 AI 应用开发者获得更好的模型训练效果。在 BML 全功能 AI 开发平台上，端到端的并行训练优化方案大幅提升了大模型训练性能；预置的基于大模型的场景模型生产线，让开发者在主流的场景开箱即用，快速构建大模型场景应用；新增的自动化推理服务监控能力，及时感知变化和问题，驱动反馈和迭代，让 AI 应用的高效迭代闭环。

目前，百度文心大模型已经产生了多款面向 C 端的单点产品，例如产业级搜索系统“文心百中”。文心百中具有三大特性：简单、强大、高效。文心百中采用纯神经搜索架构加数据驱动搜索效果优化策略设计而成，应用简便，仅需三步即可在线完成搜索引擎的构建；文心百中拥有强大的语义理解能力，相比传统模型实现了质的提升，已在多个场景展现惊艳的搜索效果；文心百中具备搜索系统从 0 到 1 的高效建设能力，相比传统搜索系统可减少人力成本 90%以上；同时，基于强大的小样本学习能力的文心大模型，文心百中仅需少量的数据，即可实现不同行业的搜索效果优化，十分高效。

图37 文心百中：大模型驱动的产业级搜索系统



资料来源：百度文心大模型官网，HTI

文心大模型与飞桨共享生态，围绕“产业级”和“知识增强”两大特色，持续构建适合产业应用的模型体系。未来，将有越来越多的企业和开发者通过大模型加深度学习平台的模式，步入AI应用的新阶段，进一步加速产业智能化升级。

6.5 华为：以昇腾 AI 芯片为基联合产业界开创大模型产业化新模式

华为的大模型之路与其腾 AI 芯片有着密不可分的关系，腾 AI 芯片总共有两款，一个是 2018 年发布的腾 310，它主要针对的是推理应用；另一个是 2019 年发布的针对训练应用的腾 910，其也被称为算力最强的 AI 处理器。

而正是在昇腾 AI 芯片的支持下，华为与合作伙伴共同开启了大模型之路，华为通过前期与业界伙伴的共同探索，开创了一条大模型产业化落地的新模式，即围绕某个领域的大模型成立产学研用的产业联合体，打通科研创新到产业落地整个流程。这样一来，大模型的创新既可以更准确地契合行业场景需求，又能够促进产业合作伙伴直接基于大模型创新孵化行业应用。

这种大模型产业化新模式已经取得丰硕成果。2021 年底，华为与合作伙伴基于全球首个三模态大模型“紫东·太初”和全球首个智能遥感框架“武汉.LuoJia”，分别成立了多模态人工智能产业联合体和智能遥感开源生态联合体。2022 年 9 月，华为围绕业界首个工业流体仿真大模型「东方·御风」以及西工大的 AI 湍流大模型成立了智能流体力学产业联合体。

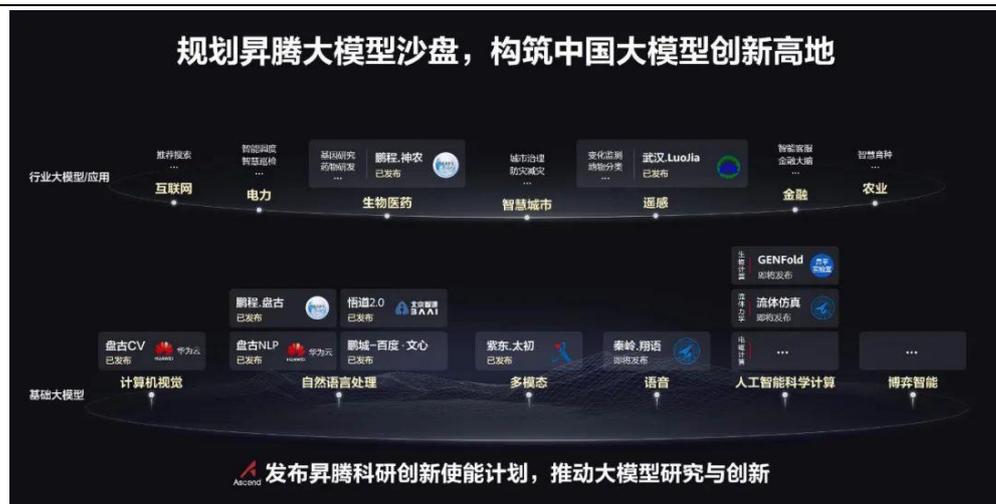
2022 年 6 月，在华为伙伴暨开发者大会 2022 上，华为更是发布人工智能大模型全流程使能体系，该体系包含从规划、开发到产业化全流程，使能大模型发展，与业界共筑中国大模型生态，开创大模型产业化新模式。

这一全流程体系包括：

(1) 规划昇腾大模型沙盘，与产业界共筑中国大模型创新高地。过去的一年，产业界基于腾 AI 先后推出鹏程.盘古、鹏程.神农、紫东.太初、武汉.LuoJia、华为云盘古系列有影响力的大模型。今年，为进一步鼓励大模型的研究与创新，华为推出腾科研创新使能计划，通过资金、算力和技术的扶持，鼓励高校及科研院所，基于昇

腾大模型沙盘，开展大模型的研究和创新，在前沿领域和热点行业打造出领先的大模型。

图38 华为联合产业界规划昇腾大模型沙盘



资料来源：中青在线，HTI

(2) 打造大模型开发使能平台，让大模型易开发、易适配、易部署。针对基础模型开发，华为推出基于昇思 MindSpore 和 ModelArts 结合的大模型开发套件，通过算法开发、并行计算、存储优化、断点续训，实现大模型的高效开发；为快速适配行业应用，推出基于 MindX 的大模型微调组件，实现一键式微调和低参数调优功能；在模型推理部署方面，推出基于 MindStudio 大模型部署套件，实现分布式推理服务化、模型轻量化和动态加密部署功能。

(3) 从科研创新到行业落地，开创产业聚集新模式。去年，基于全球首个智能遥感框架及数据集武汉·LuoJia 和全球首个三模态大模型紫东·太初，成立智能遥感开源生态联盟和多模态人工智能产业联盟，60 余家伙伴已陆续孵化出多个行业解决方案。今年，华为将与伙伴一起，共同成立 AI 流体力学、AI 生物医药、以及智慧育种等产业联盟，助力相关领域的大模型创新和产业化发展。

可以看出，华为自身在整个大模型领域中处于基础建设者的位置，它致力于联合学界和业界伙伴形成完整的大模型发展路径，从基础的计算设备，到大模型的训练、微调、推理部署全流程，再到创建大模型产业联合体，构建了一套融合了易用性与强性能的大模型完整解决方案，这在有望持续推动我国大模型在融合科技创新、赋能产业化落地中重要作用的发挥。

6.6 OpenAI: AIGC C 端商业化落地领军者，与微软加大合作

ChatGPT 发布短短 5 天，其用户数就超过 100 万，而 Facebook 获得 100 万用户，则用了整整 10 个月。2023 年 1 月，ChatGPT 的月活跃用户数预计已达 1 亿，成为历史上用户增长最快的消费应用。相比之下，达到 1 亿月活跃用户数，TikTok 花了约 9 个月的时间，Instagram 则花了两年半。OpenAI 在很短的时间内，就获得了广泛的 C 端用户群，而随着 ChatGPT Plus 这一收费项目的推出，可以说，OpenAI 已经成为了 AIGC 在 C 端商业化落地的领军者。

1月23日，OpenAI 正式在官方博客宣布，将扩展与微软的合作关系，微软将会对 OpenAI 进行一项为期多年、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。

根据声明：“过去，OpenAI 与微软合作构建了多个由 Azure 提供支持的超级计算系统，用来训练 OpenAI 所有的模型。Azure 独特的架构设计对于为 OpenAI 的 AI 训练和推理工作负载提供一流的性能和规模至关重要。微软将增加对这些系统的投资，以加速 OpenAI 的独立研究，而 Azure 仍将是 OpenAI 研究、API 和产品中所有工作负载的独家云提供商。”

据报道，此次微软计划向 OpenAI 投资高达 100 亿美元。OpenAI 此次融资对象还包括其他风险投资公司在内。这次最新融资中，OpenAI 的估值将达到 290 亿美元，而在上轮融资时，公司的估值仅 140 亿美元。OpenAI 这笔交易的条款较为复杂。报道称，若这笔交易达成，微软将获得 OpenAI 的 75% 利润，直到它收回投资，而在此期间 OpenAI 需要明确如何通过 ChatGPT 和图像创建工具 Dall-E 等其他产品获取利润。一旦 OpenAI 的利润达到了微软的“门槛”，微软将收回投资。此后，微软将持有 OpenAI 49% 的股份，其他投资者将持有另外的 49%，OpenAI Inc 将获得 2%。

OpenAI 目前商业价值的飙升，也佐证了市场对于其商业模式的认可，作为 AIGC 端商业化落地领军者，在微软加大投资的背景下，其发展速度或进入快车道，据报道，日前，GPT 系列最新版模型 GPT-4 或已经面试，日前，集成了 GPT-4 的微软旗下搜索引擎 Bing 新版本悄然上线，但在被少数用户体验之后又迅速下线，在资金充裕的背景下，OpenAI 未来产品更新速度有望持续加快。

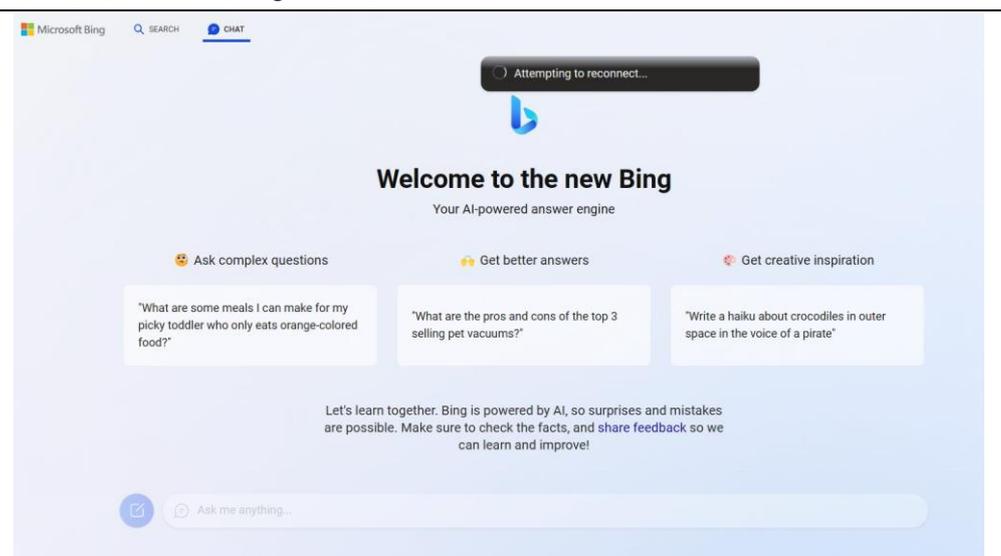
6.7 微软：高度重视 AI 大模型，将大模型融入全产品线

除了上文介绍的 Microsoft Teams Premium、Viva Sales 以外，微软也在加速将大模型融入其其他产品线。1月，瑞士达沃斯论坛期间，微软 CEO 纳德拉就曾表示，AIGC 堪比工业革命，微软也将全线接入 ChatGPT。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。这些产品包括且不限于 Bing 搜索引擎、包含 Word、PPT、Excel 的 Office 全家桶、Azure 云服务、Teams 聊天程序等等。

“我们将把它（ChatGPT）整合到我们自己的应用程序中，所以你完全可以期待。”纳德拉在采访中提到，“微软的每个产品都将具有一些相同的 AI 功能，以彻底改变产品。”纳德拉表示，在微软产品中使用人工智能是该公司计划将其对 OpenAI 的早期投资商业化的第三部分。第一部分是通过其 Azure 云计算平台访问 OpenAI 的工具，这家科技巨头正在全面推出 Azure OpenAI 服务，此前在 2021 年 11 月首次亮相后，首次允许有限数量的企业客户访问该服务；第二部分是将 OpenAI 的基础模型作为平台提供，以便任何行业的任何实体都可以在其上构建组织。纳德拉强调，ChatGPT 和 GPT 模型是多年来和 OpenAI 深入合作下的产物。

根据媒体 Semafor 的一份新报告，微软计划在未来几周内将 GPT 的下一代版本 GPT-4 整合到旗下搜索引擎必应（Bing）中。而在 2 月 5 日，集成了 GPT-4 的新版 Bing 就曾短暂上线，部分幸运用户提前访问了新版 Bing。新版 Bing 在页面中不再仅仅只是一个搜索框，而是描述为“你身边的研究助理、个人规划师和创意伙伴”。新版 Bing 和传统网络搜索引擎第一个主要区别是搜索框，它不是一个搜索栏，而是一个聊天框。它的尺寸要大得多，并且鼓励使用自然语言而不是关键字驱动搜索词。用户可以让 Bing 查找特定的主题或想法，甚至征求它的意见，它会在聊天气泡中回复你的问题。据悉，新的 Bing 还能够根据您的想法调整其搜索查询。用户可以用自然语言告诉它自身的计划或要求，例如饮食需求或日程安排冲突，它会尽力考虑这些要求来调整搜索结果。

图39 集成了 GPT-4 的 Bing 全新版本



资料来源: Twitter, HTI

除了对 OpenAI 的投资，微软自身也从未放松对大模型的重视，2021 年，微软就和英伟达的共同造出了超大语言模型：高达 5300 亿参数的「Megatron-Turing」。可以看出，微软始终高度重视 AI 大模型领域，与 OpenAI 的独家合作，更是使得微软能直接享受到 GPT 大模型家族最新版本的强大能力，例如目前的 GPT-3.5 和即将推出的 GPT-4，这对于微软未来的发展将起到重要的助力。

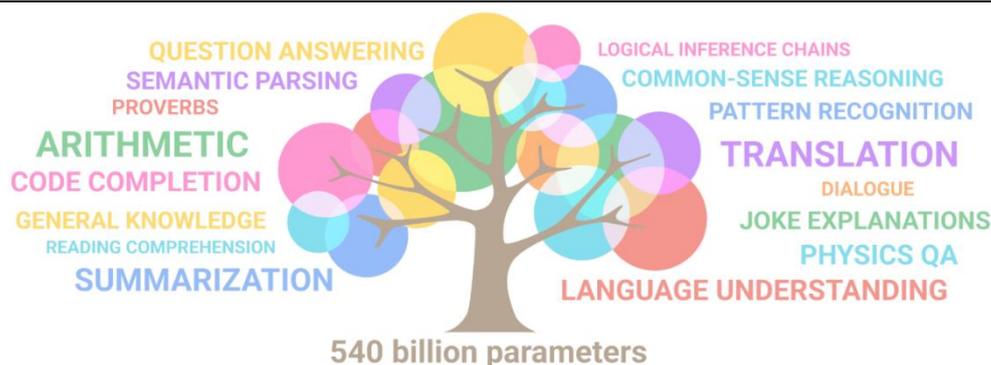
6.8 谷歌：AI 大模型领域的领军人之一，积累深厚

谷歌是大模型领域的领军人之一，谷歌早在就 2018 年 10 月发布了 BERT 大模型，它利用 BooksCorpus 和英文维基百科里纯文字的部分，无须标注数据，用设计的两个自监督任务来做训练，训练好的模型通过微调在 11 个下游任务上实现最佳性能。

2021 年，在一篇学术论文中，Google AI 介绍了 Pathways 语言模型 (PaLM)，这是一个使用 Pathways 系统训练的 5400 亿参数、密集解码器的 Transformer 模型，它使我们能够有效地训练一个模型并跨越多个 TPU。Google AI 在数百个语言理解和生成任务上对 PaLM 进行了评估，发现它在大多数任务中实现了最先进的小样本性能，在许多情况下都有显著的优势。Pathways 系统通过 PaLM 训练的 5400 亿参数语言模型进行了首次大规模使用演示，训练任务成功扩展到 6144 个芯片上，这是迄今为止用于训练的最大基于 TPU 的系统配置。

PaLM 实现了 57.8% 的 FLOP 硬件训练效率，这是该规模的 LLM 所达到的最高水平。由于采用并行策略和 Transformer 相结合的新构架，允许并行计算注意力和前馈层，从而实现 TPU 编译器优化加速。PaLM 使用英语和多语言数据集进行训练，这些数据集包括高质量的网络文档、书籍、维基百科、对话和 GitHub 代码。为此，Google AI 还创建了一个“无损”词汇表，保留所有空格（对代码尤其重要），将词汇表外的 Unicode 字符拆分为字节，并将数字拆分为单独的标记，每个数字一个。

图40 5400 亿参数使得 PaLM 跨任务性能显著提高



资料来源：谷歌研究院官网，HTI

在 PaLM 大模型的基础上，谷歌于 2022 年 12 月发布了 Med-PaLM 这一产品，首先，谷歌研究院和 DeepMind 团队以医疗问答为研究对象，提出了一个医学问答基准 MultiMedQA，包括医学考试、医学研究和消费者医学问题；在 MultiMedQA 上评估了 PaLM 及微调变体 Flan-PaLM；提出了指令提示 x 调整，让 Flan-PaLM 进一步与医学接轨，最终产生了 Med-PaLM。

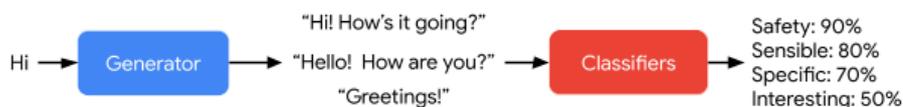
在经历一系列考核后，该模型被证实“几乎达到”了人类医生的水平。Med-PaLM 在科学常识方面的正确率在 92% 以上，在理解、检索和推理能力方面，也几乎达到了人类医生的水平，并在克服隐性偏见方面略胜一筹。不过，该研究负责人也表示，Med-PaLM 给出的答案在整体上仍然不如临床医生，该模型在实际应用前有待进一步完善。2022 年 8 月份，谷歌就将大型语言模型首次集成到机器人中，有了 AI 模型的加持，机器人能像人类一样响应完整的命令。谷歌研究人员就此做了演示。当对机器人说：“我饿了，你能给我点零食吗？”机器人在自助餐厅中搜索一番后，打开了一个抽屉，找到一袋薯片并将它拿给研究人员。

除此以外，谷歌在 2021 年开发者大会上还公布了 LaMDA（Language Model for Dialogue Applications），这是一个大型自然语言对话模型。LaMDA 也基于 Transformer 架构。后者是谷歌公司于 2017 年发布并开源的神经网络架构。利用该架构生成的模型，可以被训练阅读一组单词（比如一句话或一个段落），并且关注这些单词之间的联系，然后预测接下来会是什么单词。与其他模型不同的是，LaMDA 在对话方面接受了更多训练。

在预训练阶段，谷歌从公共数据中创建了一个 1.56T 的数据集投喂给 LaMDA，让它对自然语言有一个初步的认识。到这一步，LaMDA 就可以根据给到的句子预测上下文。但这种预测通常还是不够精准，所以需要进一步微调。在对话期间，LaMDA 会根据对话生成几个备选回答，并通过 LaMDA 分类器预测每个回答的有趣程度（SSI）和安全程度（Safety）的得分。安全分数低的答案就会首先被过滤掉。其余的答案会根据有趣程度排序，并将分数最高的备选答案作为回答。从 LaMDA 的训练过程可以看到，它并不理解自己的回答。但机器学习优势就在于它可以从大量的数据中寻找规律，通过自己的量化评价体系，找出最接近正确的那个回答。

LaMDA 最为核心的就是进行“开放域”（Open Domain）对话的能力。而这项能力的重要依托，就是相比于现有的对话模型，LaMDA 更能理解对话的语境。它可以通过阅读句子或段落来“破译”对话意图，发现单词之间的关联，并能预测接下来可能出现的单词，从而做出合乎语境的回答。在这样的能力支撑下，LaMDA 能够和人在无穷无尽的话题转换中聊下去，进行长时间的开放性对话。用谷歌官方的话来形容，就是“能够用自由流动的方式，谈论无穷无尽的主题”。

图41 LaMDA 训练模式



资料来源：谷歌研究院官网，HTI

面对 ChatGPT 的崛起，谷歌内部也在积极开发与 ChatGPT 类似的「Apprentice Bard」AI 聊天系统。它便是基于 LaMDA 大模型，用户可在文本框中输入问题或提示，之后就能得到文字答复。

据称谷歌已要求 LaMDA 团队把「Apprentice Bard」作为首要任务，并告诉他们不要去参加无关的会议。在目前的示范样本中，「Apprentice Bard」给出的回答内容包含了最近发生事件的信息，**而这是认知基本停留在 2021 年之前的 ChatGPT 暂时还无法做到的。**

可以看出，虽然 ChatGPT 的发布的确领先了谷歌的步伐，但是谷歌过去在大模型领域拥有深厚的积累，这些能力将帮助谷歌在未来较短的时间迎头赶上。

7. 风险提示

行业应用不及预期，AIGC 技术发展不及预期的风险。

APPENDIX 1

Summary

- Big models are the trend and future of AI.
- Model companies will be the basic class of companies in the AI world.
- Model companies will bring disruption to the competitive landscape of technology companies.
- MaaS will become the core business model of AI companies.
- Top picks: Sense Time Group (20 HK), Cloudwalk Group (688327 CH), Iflytek (002230 CH), Baidu (BIDU US), Huawei (unlisted).
- Risk: industry application is not as expected, AIGC technology development is not as expected.

附录 APPENDIX

重要信息披露

本研究报告由海通国际分销，海通国际是由海通国际研究有限公司(HTIRL)，Haitong Securities India Private Limited (HSIPL)，Haitong International Japan K.K. (HTIJKK)和海通国际证券有限公司(HTISCL)的证券研究团队所组成的全球品牌，海通国际证券集团(HTISG)各成员分别在其许可的司法管辖区内从事证券活动。

IMPORTANT DISCLOSURES

This research report is distributed by Haitong International, a global brand name for the equity research teams of Haitong International Research Limited ("HTIRL"), Haitong Securities India Private Limited ("HSIPL"), Haitong International Japan K.K. ("HTIJKK"), Haitong International Securities Company Limited ("HTISCL"), and any other members within the Haitong International Securities Group of Companies ("HTISG"), each authorized to engage in securities activities in its respective jurisdiction.

HTIRL 分析师认证 Analyst Certification:

我，郑宏达，在此保证 (i) 本研究报告中的意见准确反映了我们对本研究中提及的任何或所有目标公司或上市公司的个人观点，并且 (ii) 我的报酬中没有任何部分与本研究报告中表达的具体建议或观点直接或间接相关；及就此报告中所讨论目标公司的证券，我们（包括我们的家属）在其中均不持有任何财务利益。I, Nathan Zheng, certify that (i) the views expressed in this research report accurately reflect my personal views about any or all of the subject companies or issuers referred to in this research and (ii) no part of my compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this research report; and that I (including members of my household) have no financial interest in the security or securities of the subject companies discussed.

我，杨林，在此保证 (i) 本研究报告中的意见准确反映了我们对本研究中提及的任何或所有目标公司或上市公司的个人观点，并且 (ii) 我的报酬中没有任何部分与本研究报告中表达的具体建议或观点直接或间接相关；及就此报告中所讨论目标公司的证券，我们（包括我们的家属）在其中均不持有任何财务利益。I, Lin Yang, certify that (i) the views expressed in this research report accurately reflect my personal views about any or all of the subject companies or issuers referred to in this research and (ii) no part of my compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this research report; and that I (including members of my household) have no financial interest in the security or securities of the subject companies discussed.

我，洪琳，在此保证 (i) 本研究报告中的意见准确反映了我们对本研究中提及的任何或所有目标公司或上市公司的个人观点，并且 (ii) 我的报酬中没有任何部分与本研究报告中表达的具体建议或观点直接或间接相关；及就此报告中所讨论目标公司的证券，我们（包括我们的家属）在其中均不持有任何财务利益。I, Lin Hong, certify that (i) the views expressed in this research report accurately reflect my personal views about any or all of the subject companies or issuers referred to in this research and (ii) no part of my compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed in this research report; and that I (including members of my household) have no financial interest in the security or securities of the subject companies discussed.

利益冲突披露 Conflict of Interest Disclosures

海通国际及其某些关联公司可从事投资银行业务和/或对本研究中的特定股票或公司进行做市或持有自营头寸。就本研究报告而言，以下是有关该等关系的披露事项（以下披露不能保证及时无遗漏，如需了解及时全面信息，请发邮件至 ERD-Disclosure@htisec.com）

HTI and some of its affiliates may engage in investment banking and / or serve as a market maker or hold proprietary trading positions of certain stocks or companies in this research report. As far as this research report is concerned, the following are the disclosure matters related to such relationship (As the following disclosure does not ensure timeliness and completeness, please send an email to ERD-Disclosure@htisec.com if timely and comprehensive information is needed).

海通证券股份有限公司和/或其子公司（统称“海通”）在过去 12 个月内参与了 BABA.US 的投资银行项目。投资银行项目包括：1、海通担任上市前辅导机构、保荐人或主承销商的首次公开发行项目；2、海通作为保荐人、主承销商或财务顾问的股权或债务再融资项目；3、海通作为主经纪商的新三板上市、目标配售和并购项目。

Haitong Securities Co., Ltd. and/or its subsidiaries (collectively, the "Haitong") have a role in investment banking projects of BABA.US within the past 12 months. The investment banking projects include 1. IPO projects in which Haitong acted as pre-listing tutor, sponsor, or lead-underwriter; 2. equity or debt refinancing projects of BABA.US for which Haitong acted as sponsor, lead-underwriter or financial advisor; 3. listing by introduction in the new three board, target placement, M&A projects in which Haitong acted as lead-brokerage firm.

0020.HK 目前或过去 12 个月内是海通的投资银行业务客户。

0020.HK is/was an investment bank clients of Haitong currently or within the past 12 months.

阿里巴巴（北京）软件服务有限公司, 阿里巴巴（成都）软件技术有限公司, 阿里巴巴（中国）网络技术有限公司, 杭州阿里巴巴创业投资管理有限公司 及杭州阿里巴巴泽泰信息技术有限公司 目前或过去 12 个月内是海通的客户。海通向客户提供非投资银行业务的证券相关业务服务。

阿里巴巴（北京）软件服务有限公司, 阿里巴巴（成都）软件技术有限公司, 阿里巴巴（中国）网络技术有限公司, 杭州阿里巴巴创业投资管理有限公司 and 杭州阿里巴巴泽泰信息技术有限公司 is/was a client of Haitong currently or within the past 12 months. The client has been provided for non-investment-banking securities-related services.

海通在过去 12 个月中获得对 0020.HK 提供投资银行服务的报酬。

Haitong received in the past 12 months compensation for investment banking services provided to 0020.HK.

海通预计将（或者有意向）在未来三个月内从 0020.HK 获得投资银行服务报酬。

Haitong expects to receive, or intends to seek, compensation for investment banking services in the next three months from 0020.HK.

海通在过去的 12 个月中从阿里巴巴（北京）软件服务有限公司,阿里巴巴（成都）软件技术有限公司,阿里巴巴（中国）网络技术有限公司,杭州阿里巴巴创业投资管理有限公司 及杭州阿里巴巴泽泰信息技术有限公司获得除投资银行服务以外之产品或服务的报酬。

Haitong has received compensation in the past 12 months for products or services other than investment banking from 阿里巴巴（北京）软件服务有限公司,阿里巴巴（成都）软件技术有限公司,阿里巴巴（中国）网络技术有限公司,杭州阿里巴巴创业投资管理有限公司 and 杭州阿里巴巴泽泰信息技术有限公司。

海通担任 0700.HK 有关证券的做市商或流动性提供者。

Haitong acts as a market maker or liquidity provider in the securities of 0700.HK.

评级定义 (从 2020 年 7 月 1 日开始执行):

海通国际（以下简称“HTI”）采用相对评级系统来为投资者推荐我们覆盖的公司：优于大市、中性或弱于大市。投资者应仔细阅读 HTI 的评级定义。并且 HTI 发布分析师观点的完整信息，投资者应仔细阅读全文而非仅看评级。在任何情况下，分析师的评级和研究都不能作为投资建议。投资者的买卖股票的决策应基于各自情况（比如投资者的现有持仓）以及其他因素。

分析师股票评级

优于大市，未来 12-18 个月内预期相对基准指数涨幅在 10%以上，基准定义如下

中性，未来 12-18 个月内预期相对基准指数变化不大，基准定义如下。根据 FINRA/NYSE 的评级分布规则，我们会将中性评级划入持有这一类别。

弱于大市，未来 12-18 个月内预期相对基准指数跌幅在 10%以上，基准定义如下

各地股票基准指数：日本 – TOPIX, 韩国 – KOSPI, 台湾 – TAIEX, 印度 – Nifty100, 美国 – SP500; 其他所有中国概念股 – MSCI China.

Ratings Definitions (from 1 Jul 2020):

Haitong International uses a relative rating system using Outperform, Neutral, or Underperform for recommending the stocks we cover to investors. Investors should carefully read the definitions of all ratings used in Haitong International Research. In addition, since Haitong International Research contains more complete information concerning the analyst's views, investors should carefully read Haitong International Research, in its entirety, and not infer the contents from the rating alone. In any case, ratings (or research) should not be used or relied upon as investment advice. An investor's decision to buy or sell a stock should depend on individual circumstances (such as the investor's existing holdings) and other considerations.

Analyst Stock Ratings

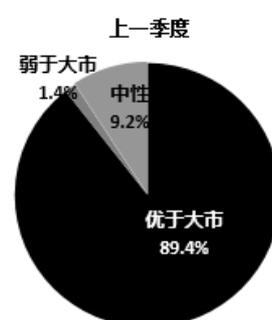
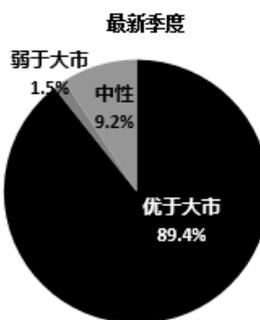
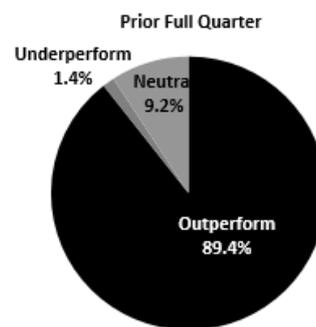
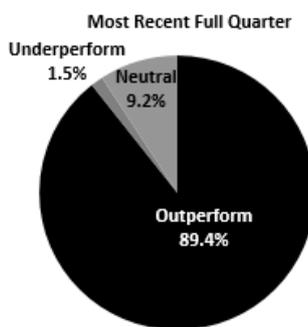
Outperform: The stock's total return over the next 12-18 months is expected to exceed the return of its relevant broad market benchmark, as indicated below.

Neutral: The stock's total return over the next 12-18 months is expected to be in line with the return of its relevant broad market benchmark, as indicated below. For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category.

Underperform: The stock's total return over the next 12-18 months is expected to be below the return of its relevant broad market benchmark, as indicated below.

Benchmarks for each stock's listed region are as follows: Japan – TOPIX, Korea – KOSPI, Taiwan – TAIEX, India – Nifty100, US – SP500; for all other China-concept stocks – MSCI China.

评级分布 Rating Distribution



截至 2022 年 12 月 31 日海通国际股票研究评级分布

	优于大市	中性 (持有)	弱于大市
海通国际股票研究覆盖率	89.4%	9.2%	1.5%
投资银行客户*	5.2%	7.3%	8.3%

*在每个评级类别里投资银行客户所占的百分比。

上述分布中的买入，中性和卖出分别对应我们当前优于大市，中性和落后大市评级。

只有根据 FINRA/NYSE 的评级分布规则，我们才将中性评级划入持有这一类别。请注意在上表中不包含非评级的股票。

此前的评级系统定义（直至 2020 年 6 月 30 日）：

买入，未来 12-18 个月内预期相对基准指数涨幅在 10%以上，基准定义如下

中性，未来 12-18 个月内预期相对基准指数变化不大，基准定义如下。根据 FINRA/NYSE 的评级分布规则，我们会将中性评级划入持有这一类别。

卖出，未来 12-18 个月内预期相对基准指数跌幅在 10%以上，基准定义如下

各地股票基准指数：日本 – TOPIX, 韩国 – KOSPI, 台湾 – TAIEX, 印度 – Nifty100; 其他所有中国概念股 – MSCI China.

Haitong International Equity Research Ratings Distribution, as of Dec 31, 2022

	Outperform	Neutral (hold)	Underperform
HTI Equity Research Coverage	89.4%	9.2%	1.5%
IB clients*	5.2%	7.3%	8.3%

*Percentage of investment banking clients in each rating category.

BUY, Neutral, and SELL in the above distribution correspond to our current ratings of Outperform, Neutral, and Underperform.

For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category. Please note that stocks with an NR designation are not included in the table above.

Previous rating system definitions (until 30 Jun 2020):

BUY: The stock's total return over the next 12-18 months is expected to exceed the return of its relevant broad market benchmark, as indicated below.

NEUTRAL: The stock's total return over the next 12-18 months is expected to be in line with the return of its relevant broad market benchmark, as indicated below. For purposes only of FINRA/NYSE ratings distribution rules, our Neutral rating falls into a hold rating category.

SELL: The stock's total return over the next 12-18 months is expected to be below the return of its relevant broad market benchmark, as indicated below.

Benchmarks for each stock's listed region are as follows: Japan – TOPIX, Korea – KOSPI, Taiwan – TAIEX, India – Nifty100; for all other China-concept stocks – MSCI China.

海通国际非评级研究：海通国际发布计量、筛选或短篇报告，并在报告中根据估值和其他指标对股票进行排名，或者基于可能的估值倍数提出建议价格。这种排名或建议价格并非为了进行股票评级、提出目标价格或进行基本面估值，而仅供参考使用。

Haitong International Non-Rated Research: Haitong International publishes quantitative, screening or short reports which may rank stocks according to valuation and other metrics or may suggest prices based on possible valuation multiples. Such rankings or suggested prices do not purport to be stock ratings or target prices or fundamental values and are for information only.

海通国际 A 股覆盖：海通国际可能会就沪港通及深港通的中国 A 股进行覆盖及评级。海通证券（600837.CH），海通国际于上海的母公司，也会于中国发布中国 A 股的研究报告。但是，海通国际使用与海通证券不同的评级系统，所以海通国际与海通证券的中国 A 股评级可能有所不同。

Haitong International Coverage of A-Shares: Haitong International may cover and rate A-Shares that are subject to the Hong Kong Stock Connect scheme with Shanghai and Shenzhen. Haitong Securities (HS; 600837 CH), the ultimate parent company of HTISG based in Shanghai, covers and publishes research on these same A-Shares for distribution in mainland China. However, the rating system employed by HS differs from that used by HTI and as a result there may be a difference in the HTI and HS ratings for the same A-share stocks.

海通国际优质 100 A 股 (Q100) 指数：海通国际 Q100 指数是一个包括 100 支由海通证券覆盖的优质中国 A 股的计量产品。这些股票是通过基于质量的筛选过程，并结合对海通证券 A 股团队自下而上的研究。海通国际每季对 Q100 指数成分作出复审。

Haitong International Quality 100 A-share (Q100) Index: HTI's Q100 Index is a quant product that consists of 100 of the highest-quality A-shares under coverage at HS in Shanghai. These stocks are carefully selected through a quality-based screening process in combination with a review of the HS A-share team's bottom-up research. The Q100 constituent companies are reviewed quarterly.

MSCI ESG 评级免责声明条款：尽管海通国际的信息供货商（包括但不限于 MSCI ESG Research LLC 及其联属公司（「ESG 方」）从其认为可靠的来源获取信息（「信息」），ESG 方均不承担或保证此处任何数据的原创性、准确性和/或完整性，并明确表示不作出任何明示或默示的担保，包括可商售性和针对特定目的的适用性。该信息只能供阁下内部使用，不得以任何形式复制或重新传播，并不得用作任何金融工具、产品或指数的基础或组成部分。此外，信息本质上不能用于判断购买或出售何种证券，或何时购买或出售该证券。即使已被告知可能造成的损害，ESG 方均不承担与此处任何资料有关的任何错误或遗漏所引起的任何责任，也不对任何直接、间接、特殊、惩罚性、附带性或任何其他损害赔偿（包括利润损失）承担任何责任。

MSCI ESG Disclaimer: Although Haitong International's information providers, including without limitation, MSCI ESG Research LLC and its affiliates (the "ESG Parties"), obtain information (the "Information") from sources they consider reliable, none of the ESG Parties warrants or guarantees the originality, accuracy and/or completeness, of any data herein and expressly disclaim all express or implied warranties, including those of merchantability and fitness for a particular purpose. The Information may only be used for your internal use, may not be reproduced or disseminated in any form and may not be used as a basis for, or a component of, any financial instruments or products or indices. Further, none of the Information can in and of itself be used to determine which securities to buy or sell or when to buy or sell them. None of the ESG Parties shall have any liability for any errors or omissions in connection with any data herein, or any liability for any direct,

indirect, special, punitive, consequential or any other damages (including lost profits) even if notified of the possibility of such damages.

盟浪义利 (FIN-ESG) 数据通免责声明条款: 在使用盟浪义利 (FIN-ESG) 数据之前, 请务必仔细阅读本条款并同意本声明:

第一条 义利 (FIN-ESG) 数据系由盟浪可持续数字科技有限责任公司 (以下简称“本公司”) 基于合法取得的公开信息评估而成, 本公司对信息的准确性及完整性不作任何保证。对公司述的评估结果造成的任何直接或间接损失负责。

第二条 盟浪并不因收到此评估数据而将收件人视为客户, 收件人使用此数据时应根据自身实际情况作出自我独立判断。本数据所载内容反映的是盟浪在最初发布本数据日期当日的判断, 盟浪有权在不发出通知的情况下更新、修订与发出其他与本数据所载内容不一致或有不同结论的数据。除非另行说明, 本数据 (如财务业绩数据等) 仅代表过往表现, 过往的业绩表现不作为日后回报的预测。

第三条 本数据版权归本公司所有, 本公司依法保留各项权利。未经本公司事先书面许可授权, 任何个人或机构不得将本数据中的评估结果用于任何营利性目的, 不得对本数据进行修改、复制、编译、汇编、再次编辑、改编、删减、缩写、节选、发行、出租、展览、表演、放映、广播、信息网络传播、摄制、增加图标及说明等, 否则因此给盟浪或其他第三方造成损失的, 由用户承担相应的赔偿责任, 盟浪不承担责任。

第四条 如本免责声明未约定, 而盟浪网站平台载明的其他协议内容 (如《盟浪网站用户注册协议》《盟浪网用户服务 (含认证) 协议》《盟浪网隐私政策》等) 有约定的, 则按其他协议的约定执行; 若本免责声明与其他协议约定存在冲突或不一致的, 则以本免责声明约定为准。

SusallWave FIN-ESG Data Service Disclaimer: Please read these terms and conditions below carefully and confirm your agreement and acceptance with these terms before using SusallWave FIN-ESG Data Service.

1. FIN-ESG Data is produced by SusallWave Digital Technology Co., Ltd. (In short, SusallWave)'s assessment based on legal publicly accessible information. SusallWave shall not be responsible for any accuracy and completeness of the information. The assessment result is for reference only. It is not for any investment advice for any individual or institution and not for basis of purchasing, selling or holding any relative financial products. We will not be liable for any direct or indirect loss of any individual or institution as a result of using SusallWave FIN-ESG Data.
2. SusallWave do not consider recipients as customers for receiving these data. When using the data, recipients shall make your own independent judgment according to your practical individual status. The contents of the data reflect the judgment of us only on the release day. We have right to update and amend the data and release other data that contains inconsistent contents or different conclusions without notification. Unless expressly stated, the data (e.g., financial performance data) represents past performance only and the past performance cannot be viewed as the prediction of future return.
3. The copyright of this data belongs to SusallWave, and we reserve all rights in accordance with the law. Without the prior written permission of our company, none of individual or institution can use these data for any profitable purpose. Besides, none of individual or institution can take actions such as amendment, replication, translation, compilation, re-editing, adaption, deletion, abbreviation, excerpts, issuance, rent, exhibition, performance, projection, broadcast, information network transmission, shooting, adding icons and instructions. If any loss of SusallWave or any third-party is caused by those actions, users shall bear the corresponding compensation liability. SusallWave shall not be responsible for any loss.
4. If any term is not contained in this disclaimer but written in other agreements on our website (e.g. *User Registration Protocol of SusallWave Website, User Service (including authentication) Agreement of SusallWave Website, Privacy Policy of Susallwave Website*), it should be executed according to other agreements. If there is any difference between this disclaimer and other agreements, this disclaimer shall be applied.

重要免责声明:

非印度证券的研究报告: 本报告由海通国际证券集团有限公司 (“HTISGL”) 的全资附属公司海通国际研究有限公司 (“HTIRL”) 发行, 该公司是根据香港证券及期货条例 (第 571 章) 持有第 4 类受规管活动 (就证券提供意见) 的持牌法团。该研究报告在 HTISGL 的全资附属公司 Haitong International (Japan) K.K. (“HTIJK”) 的协助下发行, HTIJK 是由日本关东财务局监管为投资顾问。

印度证券的研究报告: 本报告由从事证券交易、投资银行及证券分析及受 Securities and Exchange Board of India (“SEBI”) 监管的 Haitong Securities India Private Limited (“HTSIPL”) 所发行, 包括制作及发布涵盖 BSE Limited (“BSE”) 和 National Stock Exchange of India Limited (“NSE”) 上市公司 (统称为「印度交易所」) 的研究报告。HTSIPL 于 2016 年 12 月 22 日被收购并成为海通国际证券集团有限公司 (“HTISG”) 的一部分。

所有研究报告均以海通国际为名作为全球品牌, 经许可由海通国际证券股份有限公司及/或海通国际证券集团的其他成员在其司法管辖区发布。

本文件所载信息和观点已被编译或源自可靠来源, 但 HTIRL、HTISCL 或任何其他属于海通国际证券集团有限公司 (“HTISG”) 的成员对其准确性、完整性和正确性不做任何明示或暗示的声明或保证。本文件中所有观点均截至本报告日期, 如有更改, 恕不另行通知。本文件仅供参考使用。文件中提及的任何公司或其股票的说明并非意图展示完整的内容, 本文件并非/不应被解释为对证券买卖的明示或暗示地出价或征价。在某些司法管辖区, 本文件中提及的证券可能无法进行买卖。如果投资产品以投资者本国货币以外的币种进行计价, 则汇率变化可能会对投资产生不利影响。过去的表现并不一定代表将来的结果。某些特定交易, 包括设计金融衍生工具的, 有产生重大风险的可能性, 因此并不适合所有的投资者。您还应认识到本文件中的建议并非为您量身定制。分析师并未考虑到您自身的财务情况, 如您的财务状况和风险偏好。因此您必须自行分析并在适用的情况下咨询自己的法律、税收、会计、金融和其他方面的专业顾问, 以期在投资之前评估该项建议是否适合于您。若由于使用本文件所载的材料而产生任何直接或间接的损失, HTISG 及其董事、雇员或代理人对此均不承担任何责任。

除对本文内容承担责任的分析师外, HTISG 及我们的关联公司、高级管理人员、董事和雇员, 均可不时作为主事人就本文件所述的任何证券或衍生品持有长仓或短仓以及进行买卖。HTISG 的销售员、交易员和其他专业人士均可向 HTISG 的相关客户和公司提供与本文件所述意见相反的口头或书面市场评论意见或交易策略。HTISG 可做出与本文件所述建议或意见不一致的投资决策。但 HTIRL 没有义务来确保本文件的收件人了解到该等交易决定、思路或建议。

请访问海通国际网站 www.equities.htisec.com, 查阅更多有关海通国际为预防和避免利益冲突设立的组织和行政安排的内容信息。

非美国分析师披露信息: 本项研究首页上列明的海通国际分析师并未在 FINRA 进行注册或者取得相应的资格, 并且不受美国 FINRA 有关与本项研究目标公司进行沟通、公开露面和自营

证券交易的第 2241 条规则之限制。

IMPORTANT DISCLAIMER

For research reports on non-Indian securities: The research report is issued by Haitong International Research Limited (“HTIRL”), a wholly owned subsidiary of Haitong International Securities Group Limited (“HTISGL”) and a licensed corporation to carry on Type 4 regulated activity (advising on securities) for the purpose of the Securities and Futures Ordinance (Cap. 571) of Hong Kong, with the assistance of Haitong International (Japan) K.K. (“HTIJKK”), a wholly owned subsidiary of HTISGL and which is regulated as an Investment Adviser by the Kanto Finance Bureau of Japan.

For research reports on Indian securities: The research report is issued by Haitong Securities India Private Limited (“HSIPL”), an Indian company and a Securities and Exchange Board of India (“SEBI”) registered Stock Broker, Merchant Banker and Research Analyst that, inter alia, produces and distributes research reports covering listed entities on the BSE Limited (“BSE”) and the National Stock Exchange of India Limited (“NSE”) (collectively referred to as “Indian Exchanges”). HSIPL was acquired and became part of the Haitong International Securities Group of Companies (“HTISG”) on 22 December 2016.

All the research reports are globally branded under the name Haitong International and approved for distribution by Haitong International Securities Company Limited (“HTISCL”) and/or any other members within HTISG in their respective jurisdictions.

The information and opinions contained in this research report have been compiled or arrived at from sources believed to be reliable and in good faith but no representation or warranty, express or implied, is made by HTIRL, HTISCL, HSIPL, HTIJKK or any other members within HTISG from which this research report may be received, as to their accuracy, completeness or correctness. All opinions expressed herein are as of the date of this research report and are subject to change without notice. This research report is for information purpose only. Descriptions of any companies or their securities mentioned herein are not intended to be complete and this research report is not, and should not be construed expressly or impliedly as, an offer to buy or sell securities. The securities referred to in this research report may not be eligible for purchase or sale in some jurisdictions. If an investment product is denominated in a currency other than an investor's home currency, a change in exchange rates may adversely affect the investment. Past performance is not necessarily indicative of future results. Certain transactions, including those involving derivatives, give rise to substantial risk and are not suitable for all investors. You should also bear in mind that recommendations in this research report are not tailor-made for you. The analyst has not taken into account your unique financial circumstances, such as your financial situation and risk appetite. You must, therefore, analyze and should, where applicable, consult your own legal, tax, accounting, financial and other professional advisers to evaluate whether the recommendations suits you before investment. Neither HTISG nor any of its directors, employees or agents accepts any liability whatsoever for any direct or consequential loss arising from any use of the materials contained in this research report.

HTISG and our affiliates, officers, directors, and employees, excluding the analysts responsible for the content of this document, will from time to time have long or short positions in, act as principal in, and buy or sell, the securities or derivatives, if any, referred to in this research report. Sales, traders, and other professionals of HTISG may provide oral or written market commentary or trading strategies to the relevant clients and the companies within HTISG that reflect opinions that are contrary to the opinions expressed in this research report. HTISG may make investment decisions that are inconsistent with the recommendations or views expressed in this research report. HTI is under no obligation to ensure that such other trading decisions, ideas or recommendations are brought to the attention of any recipient of this research report.

Please refer to HTI's website www.equities.htisec.com for further information on HTI's organizational and administrative arrangements set up for the prevention and avoidance of conflicts of interest with respect to Research.

Non U.S. Analyst Disclosure: The HTI analyst(s) listed on the cover of this Research is (are) not registered or qualified as a research analyst with FINRA and are not subject to U.S. FINRA Rule 2241 restrictions on communications with companies that are the subject of the Research; public appearances; and trading securities by a research analyst.

分发和地区通知:

除非下文另有规定，否则任何希望讨论本报告或者就本项研究中讨论的任何证券进行任何交易的收件人均应联系其所在国家或地区的海通国际销售人员。

香港投资者的通知事项: 海通国际证券股份有限公司 (“HTISCL”) 负责分发该研究报告，HTISCL 是在香港有权实施第 1 类受规管活动 (从事证券交易) 的持牌公司。该研究报告并不构成《证券及期货条例》(香港法例第 571 章) (以下简称“SFO”) 所界定的要约邀请，证券要约或公众要约。本研究报告仅提供给 SFO 所界定的“专业投资者”。本研究报告未经过证券及期货事务监察委员会的审查。您不应仅根据本研究报告中所载的信息做出投资决定。本研究报告的收件人就研究报告中产生或与之相关的任何事宜请联系 HTISCL 销售人员。

美国投资者的通知事项: 本研究报告由 HTIRL, HSIPL 或 HTIJKK 编写。HTIRL, HSIPL, HTIJKK 以及任何非 HTISG 美国联营公司，均未在美国注册，因此不受美国关于研究报告编制和研究分析人员独立性规定的约束。本研究报告提供给依照 1934 年“美国证券交易法”第 15a-6 条规定的豁免注册的「美国主要机构投资者」 (“Major U.S. Institutional Investor”) 和「机构投资者」 (“U.S. Institutional Investors”)。在向美国机构投资者分发研究报告时，Haitong International Securities (USA) Inc. (“HTI USA”) 将对报告的内容负责。任何收到本研究报告的美国投资者，希望根据本研究报告提供的信息进行任何证券或相关金融工具买卖的交易，只能通过 HTI USA。HTI USA 位于 340 Madison Avenue, 12th Floor, New York, NY 10173, 电话 (212) 351-6050。HTI USA 是在美国于 U.S. Securities and Exchange Commission (“SEC”) 注册的经纪商，也是 Financial Industry Regulatory Authority, Inc. (“FINRA”) 的成员。HTIUSA 不负责编写本研究报告，也不负责其中包含的分析。在任何情况下，收到本研究报告的任何美国投资者，不得直接与分析师直接联系，也不得通过 HSIPL, HTIRL 或 HTIJKK 直接进行买卖证券或相关金融工具的交易。本研究报告中出现的 HSIPL, HTIRL 或 HTIJKK 分析师没有注册或具备 FINRA 的研究分析师资格，因此可能不受 FINRA 第 2241 条规定的与目标公司的交流，公开露面和分析师账户持有的交易证券等限制。投资本研究报告中讨论的任何非美国证券或相关金融工具 (包括 ADR) 可能存在一定风险。非美国发行的证券可能没有注册，或不受美国法规的约束。有关非美国证券或相关金融工具的信息可能有限制。外国公司可能不受审计和汇报的标准以及与美国境内生效相符的监管要求。本研究报告中以美元以外的其他货币计价的任何证券或相关金融工具的投资或收益的价值受汇率波动的影响，可能对该等证券或相关金融工具的价值或收入产生正面或负面影响。美国收件人的所有问询请联系：

Haitong International Securities (USA) Inc.
340 Madison Avenue, 12th Floor
New York, NY 10173
联系人电话: (212) 351 6050

DISTRIBUTION AND REGIONAL NOTICES

Except as otherwise indicated below, any Recipient wishing to discuss this research report or effect any transaction in any security discussed in HTI's research should contact the Haitong International salesperson in their own country or region.

Notice to Hong Kong investors: The research report is distributed by Haitong International Securities Company Limited (“HTISCL”), which is a licensed corporation to carry on Type 1 regulated activity (dealing in securities) in Hong Kong. This research report does not constitute a solicitation or an offer of securities or an invitation to the public within the meaning of the SFO. This research report is only to be circulated to “Professional Investors” as defined in the SFO. This research report has not been reviewed by the Securities and Futures Commission. You should not make investment decisions solely on the basis of the information contained in this research report. Recipients of this research report are to contact HTISCL salespersons in respect of any matters arising from, or in connection with, the research report.

Notice to U.S. investors: As described above, this research report was prepared by HTIRL, HSIPL or HTIJKK. Neither HTIRL, HSIPL, HTIJKK, nor any of the non U.S. HTISG affiliates is registered in the United States and, therefore, is not subject to U.S. rules regarding the preparation of research reports and the independence of research analysts. This research report is provided for distribution to “major U.S. institutional investors” and “U.S. institutional investors” in reliance on the exemption from registration provided by Rule 15a-6 of the U.S. Securities Exchange Act of 1934, as amended. When distributing research reports to “U.S. institutional investors,” HTI USA will accept the responsibilities for the content of the reports. Any U.S. recipient of this research report wishing to effect any transaction to buy or sell securities or related financial instruments based on the information provided in this research report should do so only through Haitong International Securities (USA) Inc. (“HTI USA”), located at 340 Madison Avenue, 12th Floor, New York, NY 10173, USA; telephone (212) 351 6050. HTI USA is a broker-dealer registered in the U.S. with the U.S. Securities and Exchange Commission (the “SEC”) and a member of the Financial Industry Regulatory Authority, Inc. (“FINRA”). HTI USA is not responsible for the preparation of this research report nor for the analysis contained therein. Under no circumstances should any U.S. recipient of this research report contact the analyst directly or effect any transaction to buy or sell securities or related financial instruments directly through HSIPL, HTIRL or HTIJKK. The HSIPL, HTIRL or HTIJKK analyst(s) whose name appears in this research report is not registered or qualified as a research analyst with FINRA and, therefore, may not be subject to FINRA Rule 2241 restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account. Investing in any non-U.S. securities or related financial instruments (including ADRs) discussed in this research report may present certain risks. The securities of non-U.S. issuers may not be registered with, or be subject to U.S. regulations. Information on such non-U.S. securities or related financial instruments may be limited. Foreign companies may not be subject to audit and reporting standards and regulatory requirements comparable to those in effect within the U.S. The value of any investment or income from any securities or related financial instruments discussed in this research report denominated in a currency other than U.S. dollars is subject to exchange rate fluctuations that may have a positive or adverse effect on the value of or income from such securities or related financial instruments. All inquiries by U.S. recipients should be directed to:

Haitong International Securities (USA) Inc.
340 Madison Avenue, 12th Floor
New York, NY 10173
Attn: Sales Desk at (212) 351 6050

中华人民共和国的通知事项: 在中华人民共和国（下称“中国”，就本报告目的而言，不包括香港特别行政区、澳门特别行政区和台湾）只有根据适用的中国法律法规而收到该材料的人员方可使用该材料。并且根据相关法律法规，该材料中的信息并不构成“在中国从事生产、经营活动”。本文件在中国并不构成相关证券的公开发售或认购。无论根据法律规定或其他任何规定，在取得中国政府所有的批准或许可之前，任何法人或自然人均不得直接或间接地购买本材料中的任何证券或任何实权益。接收本文件的人员须遵守上述限制性规定。

加拿大投资者的通知事项: 在任何情况下该等材料均不得被解释为在任何加拿大的司法管辖区内出售证券的要约或认购证券的要约邀请。本材料中所述证券在加拿大的任何要约或出售行为均只能在豁免向有关加拿大证券监管机构提交招股说明书的前提下由 Haitong International Securities (USA) Inc. (“HTI USA”) 予以实施，该公司是一家根据 National Instrument 31-103 Registration Requirements, Exemptions and Ongoing Registrant Obligations (“NI 31-103”) 的规定得到「国际交易商豁免」 (“International Dealer Exemption”) 的交易商，位于艾伯塔省、不列颠哥伦比亚省、安大略省和魁北克省。在加拿大，该等材料在任何情况下均不得被解释为任何证券的招股说明书、发行备忘录、广告或公开发行。加拿大的任何证券委员会或类似的监管机构均未审查或以任何方式批准该等材料、其中所载的信息或所述证券的优点，任何与此相反的声明即属违法。在收到该等材料时，每个加拿大的收件人均将被视为属于 National Instrument 45-106 Prospectus Exemptions 第 1.1 节或者 Securities Act (Ontario) 第 73.3(1) 节所规定的「认可投资者」 (“Accredited Investor”)，或者在适用情况下 National Instrument 31-103 第 1.1 节所规定的「许可投资者」 (“Permitted Investor”)。

新加坡投资者的通知事项: 本研究报告由 Haitong International Securities (Singapore) Pte Ltd (“HTISSPL”) [公司注册编号 201311400G] 于新加坡提供。HTISSPL 是符合《财务顾问法》 (第 110 章) (“FAA”) 定义的豁免财务顾问，可 (a) 提供关于证券、集体投资计划的部分、交易所衍生品合约和场外衍生品合约的建议 (b) 发行或公布有关证券、交易所衍生品合约和场外衍生品合约的研究分析或研究报告。本研究报告仅提供给符合《证券及期货法》 (第 289 章) 第 4A 条项下规定的机构投资者。对于因本研究报告而产生的或与之相关的任何问题，本研究报告的收件人应通过以下信息与 HTISSPL 联系：

Haitong International Securities (Singapore) Pte. Ltd
50 Raffles Place, #33-03 Singapore Land Tower, Singapore 048623
电话: (65) 6536 1920

日本投资者的通知事项: 本研究报告由海通国际证券有限公司所发布，旨在分发给从事投资管理的金融服务提供商或注册金融机构（根据日本金融机构和交易法 (“FIEL”) 第 61 (1) 条，第 17-11 (1) 条的执行及相关条款）。

英国及欧盟投资者的通知事项: 本报告由从事投资顾问的 Haitong International Securities Company Limited 所发布，本报告只面向有投资相关经验的专业客户发布。任何投资或与本报告相关的投资行为只面对此类专业客户。没有投资经验或相关投资经验的客户不得依赖本报告。Haitong International Securities Company Limited 的分支机构的净长期或短期金融权益可能超过本研究报告中提及的实体已发行股本总额的 0.5%。特别提醒有些英文报告有可能此前已经通过中文或其它语言完成发布。

澳大利亚投资者的通知事项: Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited 和 Haitong International Securities (UK) Limited 分别根据澳大利亚证券和投资委员会（以下简称“ASIC”）公司（废除及过度性）文书第 2016/396 号规章在澳大利亚分发本项研究，该等规章免除了根据 2001 年《公司法》在澳大利亚为批发客户提供金融服务时海通国际需持有澳大利亚金融服务许可的要求。ASIC 的规章副本可在以下网站获取：www.legislation.gov.au。海通国际提供的金融服务受外国法律法规规定的管制，该等法律与在澳大利亚所适用的法律存在差异。

印度投资者的通知事项: 本报告由从事证券交易、投资银行及证券分析及受 Securities and Exchange Board of India (“SEBI”) 监管的 Haitong Securities India Private Limited (“HTSIPL”) 所发布，包括制作及发布涵盖 BSE Limited (“BSE”) 和 National Stock Exchange of India Limited (“NSE”) (统称为「印度交易所」) 研究报告。

版权所有：海通国际证券集团有限公司 2019 年。保留所有权利。

People's Republic of China (PRC): In the PRC, the research report is directed for the sole use of those who receive the research report in accordance with the applicable PRC laws and regulations. Further, the information on the research report does not constitute "production and business activities in the PRC" under relevant PRC laws. This research report does not constitute a public offer of the security, whether by sale or subscription, in the PRC. Further, no legal or natural persons of the PRC may directly or indirectly purchase any of the security or any beneficial interest therein without obtaining all prior PRC government approvals or licenses that are required, whether statutorily or otherwise. Persons who come into possession of this research are required to observe these restrictions.

Notice to Canadian Investors: Under no circumstances is this research report to be construed as an offer to sell securities or as a solicitation of an offer to buy securities in any jurisdiction of Canada. Any offer or sale of the securities described herein in Canada will be made only under an exemption from the requirements to file a prospectus with the relevant Canadian securities regulators and only by Haitong International Securities (USA) Inc., a dealer relying on the "international dealer exemption" under National Instrument 31-103 Registration Requirements, Exemptions and Ongoing Registrant Obligations ("NI 31-103") in Alberta, British Columbia, Ontario and Quebec. This research report is not, and under no circumstances should be construed as, a prospectus, an offering memorandum, an advertisement or a public offering of any securities in Canada. No securities commission or similar regulatory authority in Canada has reviewed or in any way passed upon this research report, the information contained herein or the merits of the securities described herein and any representation to the contrary is an offence. Upon receipt of this research report, each Canadian recipient will be deemed to have represented that the investor is an "accredited investor" as such term is defined in section 1.1 of National Instrument 45-106 Prospectus Exemptions or, in Ontario, in section 73.3(1) of the Securities Act (Ontario), as applicable, and a "permitted client" as such term is defined in section 1.1 of NI 31-103, respectively.

Notice to Singapore investors: This research report is provided in Singapore by or through Haitong International Securities (Singapore) Pte Ltd ("HTISSPL") [Co Reg No 201311400G. HTISSPL is an Exempt Financial Adviser under the Financial Advisers Act (Cap. 110) ("FAA") to (a) advise on securities, units in a collective investment scheme, exchange-traded derivatives contracts and over-the-counter derivatives contracts and (b) issue or promulgate research analyses or research reports on securities, exchange-traded derivatives contracts and over-the-counter derivatives contracts. This research report is only provided to institutional investors, within the meaning of Section 4A of the Securities and Futures Act (Cap. 289). Recipients of this research report are to contact HTISSPL via the details below in respect of any matters arising from, or in connection with, the research report:

Haitong International Securities (Singapore) Pte. Ltd.

10 Collyer Quay, #19-01 - #19-05 Ocean Financial Centre, Singapore 049315

Telephone: (65) 6536 1920

Notice to Japanese investors: This research report is distributed by Haitong International Securities Company Limited and intended to be distributed to Financial Services Providers or Registered Financial Institutions engaged in investment management (as defined in the Japan Financial Instruments and Exchange Act ("FIEL") Art. 61(1), Order for Enforcement of FIEL Art. 17-11(1), and related articles).

Notice to UK and European Union investors: This research report is distributed by Haitong International Securities Company Limited. This research is directed at persons having professional experience in matters relating to investments. Any investment or investment activity to which this research relates is available only to such persons or will be engaged in only with such persons. Persons who do not have professional experience in matters relating to investments should not rely on this research. Haitong International Securities Company Limited's affiliates may have a net long or short financial interest in excess of 0.5% of the total issued share capital of the entities mentioned in this research report. Please be aware that any report in English may have been published previously in Chinese or another language.

Notice to Australian investors: The research report is distributed in Australia by Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited, and Haitong International Securities (UK) Limited in reliance on ASIC Corporations (Repeal and Transitional) Instrument 2016/396, which exempts those HTISG entities from the requirement to hold an Australian financial services license under the Corporations Act 2001 in respect of the financial services it provides to wholesale clients in Australia. A copy of the ASIC Class Orders may be obtained at the following website, www.legislation.gov.au. Financial services provided by Haitong International Securities (Singapore) Pte Ltd, Haitong International Securities Company Limited, and Haitong International Securities (UK) Limited are regulated under foreign laws and regulatory requirements, which are different from the laws applying in Australia.

Notice to Indian investors: The research report is distributed by Haitong Securities India Private Limited ("HSIPL"), an Indian company and a Securities and Exchange Board of India ("SEBI") registered Stock Broker, Merchant Banker and Research Analyst that, inter alia, produces and distributes research reports covering listed entities on the BSE Limited ("BSE") and the National Stock Exchange of India Limited ("NSE") (collectively referred to as "Indian Exchanges").

This research report is intended for the recipients only and may not be reproduced or redistributed without the written consent of an authorized signatory of HTISG.

Copyright: Haitong International Securities Group Limited 2019. All rights reserved.

<http://equities.htisec.com/x/legal.html>
