

文心一言：技术与能力拆解

华泰研究

2023 年 2 月 28 日 | 中国内地

专题研究

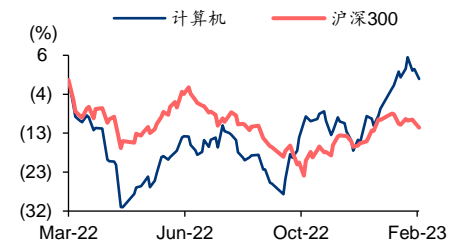
计算机

增持（维持）

研究员 谢春生
 SAC No. S0570519080006 xiechunsheng@htsc.com
 SFC No. BQZ938 +(86) 21 2987 2036

联系人 袁泽世, PhD
 SAC No. S0570122080053 yuanzeshi@htsc.com
 +(86) 21 2897 2228

行业走势图



资料来源：Wind，华泰研究

文心一言：基于文心大型模型的中文生成式对话产品

文心一言（ERNIE Bot）是基于文心大模型技术推出的生成式对话产品，由百度于 2023 年 2 月 7 日正式宣布推出，预计 3 月问世。文心一言的底层技术基础是文心大模型。文心一言的底层逻辑是，通过百度智能云提供服务，吸引企业和机构客户使用 API 和基础设施，共同搭建 AI 模型、开发应用，实现产业 AI 普惠。我们从技术和架构角度对文心大模型和百度智能云进行了拆解分析，类比已有的文心大模型应用展望了文心一言可能的能力，并探究了搜索引擎+API+生态融合的可能应用方式。截至 2023 年 2 月 25 日，已有超 450 家企业宣布加入“文心一言”生态圈。

文心大模型：文心一言的模型基础

飞桨是开源的产业级深度学习平台，而文心大模型是其中重要的基础模型库，也是“文心一言”的基础模型。更进一步划分，文心大模型包括 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，其中 NLP 大模型主要为 ERNIE 系列模型，是打造文心一言的关键。ERNIE 系列模型已迭代到 3.0 版本，将自回归和自编码网络融合进行预训练，并引入大规模知识图谱类数据，模型表现性能良好，为文心一言奠定了技术基础。

百度智能云：文心一言对外提供服务的窗口

文心一言通过百度智能云对外提供服务。百度智能云的业务架构包括百度大脑+平台+应用，文心大模型位于百度大脑基础层。此外，百度智能云推出了“云智一体 3.0”架构，包含行业应用、通用产品、AI 大底座和通用 cloud，形成了一套涵盖软硬件的“芯片-框架-大模型-行业应用”智能化闭环路径。其中，AI 大底座包括 IaaS 和 PaaS 两层，为文心一言提供了智算基础设施。2022 年上半年，百度智能云在中国 AI 公有云服务市场份额第一。

从文心一格/文心百中展望文心一言能力

文心一格和文心百中均是基于文心大模型推出的产品级应用，与文心一言定位相似。其中，文心一格基于文心大模型中的文图生成模型 ERNIE-ViLG，主要实现 AI 作画应用；文心百中是基于文心 ERNIE 大模型的端到端搜索引擎，用来替代传统搜索引擎复杂的特征及系统逻辑。我们认为，技术上来说，文心大模型已经具备了搜索、文图生成等功能，并成功得到应用，这些能力或将集成于文心一言，并增加预置作文生成、文案创作、情感分析等任务处理能力。

应用方式：搜索引擎+API+生态融合

我们认为，文心一言可能的应用方式包括搜索引擎、API、生态融合。搜索引擎方面，微软的新 Bing 与 ChatGPT 的融合已经证明了路线的可行性，文心一言或将集成于百度搜索，充分发挥国内市占率优势。大模型 API 方面，ERNIE-ViLG 文生图和 PLATO 的 API 已成功得到应用，文心一言或采用类似的 API 方式提供服务。生态融合方面，或可以接入 Apollo 自动驾驶平台、爱奇艺、小度等平台 and 终端，进一步扩展使用场景。

风险提示：宏观经济波动，下游需求不及预期。本报告内容均基于客观信息整理，不构成投资建议。

正文目录

文心一言：基于文心大型模型的中文生成式对话产品	3
文心大模型：“文心一言”基础模型库	3
ERNIE 3.0：打造文心一言的关键模型	4
文心大模型位于中国大模型市场第一梯队	6
ERNIE 3.0 vs GPT 系列：更适合中文 NLP 模型	7
百度智能云：文心一言对外提供服务的窗口	8
百度智能云：百度大脑+平台+应用三层架构	8
云智一体+AI 大底座打通文心一言基础设施与应用窗口	8
百度智能云：市场优势明显，云服务市场份额第一	10
从文心一格/文心百中展望文心一言能力	11
文心一格：ERNIE-ViLG 在产品层面的应用	11
文心百中：大模型驱动的产业级搜索系统	11
应用方式探索：搜索引擎+API+生态融合	13
可能的应用方式#1：文心一言+搜索引擎	13
可能的应用方式#2：大模型 API	13
可能的应用方式#3：产品级应用+生态融合	14
接入情况：超 450 家企业宣布加入“文心一言”生态圈	15
相关公司梳理	15
风险提示	16

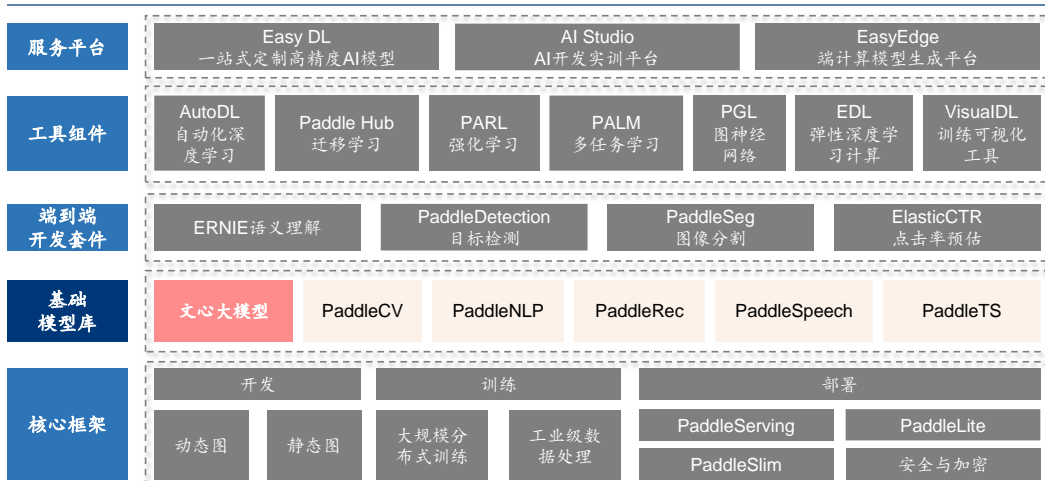
文心一言：基于文心大型模型的中文生成式对话产品

文心一言是基于文心大模型技术推出的生成式对话产品。文心一言英文名 ERNIE Bot，由百度于 2023 年 2 月 7 日正式宣布推出，预计 3 月问世。文心一言的底层技术基础是百度飞桨开源深度学习平台中的基础模型库—文心大模型，其中以 ERNIE 系列模型为核心。文心一言的底层逻辑是，通过百度智能云提供服务，吸引企业和机构客户使用百度智能云提供的 API 和基础设施，共同搭建 AI 模型、开发应用，实现产业 AI 普惠。

文心大模型：“文心一言”基础模型库

文心大模型是“文心一言”的基础模型。文心大模型是飞桨产业级开源深度学习平台中重要的基础模型库。飞桨是集深度学习核心框架、基础模型库、端到端开发套件、工具组件和服务平台于一体的产业级深度学习平台，于 2016 年正式开源。飞桨已广泛应用于工业、农业、服务业等，开发者达 535 万，创造模型 67 万个，服务企业事业单位 20 万家，产学研用紧密协同培养 AI 人才超 200 万。此外，飞桨模型库中的自然语言处理开发库 PaddleNLP，聚合了业界优质预训练模型并提供开箱即用的开发体验，覆盖 NLP 多场景，为文心一言提供了技术积累和补充。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。

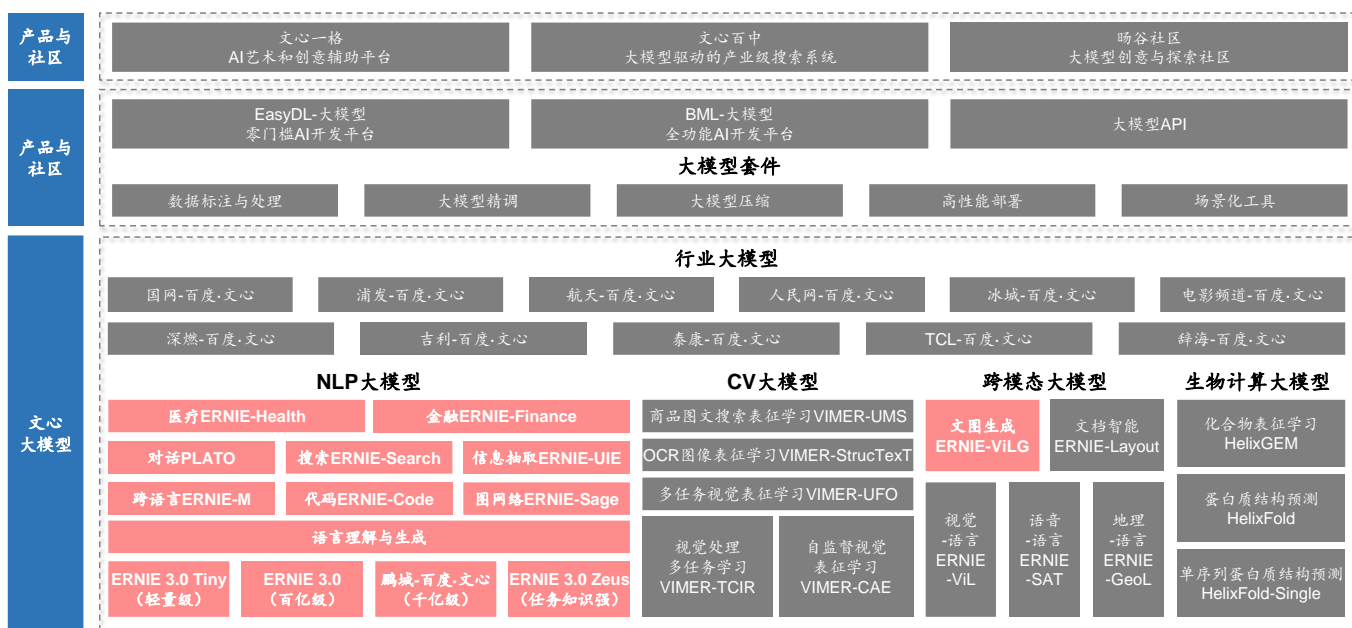
图表1：飞桨产品图谱



资料来源：飞桨官网、华泰研究

文心大模型底座+文心一格/百中，为“文心一言”打好技术与应用基础。2022 年 11 月 30 日，文心大模型新版发布，新增 11 个大模型，总量扩充至 36 个，涵盖电力、燃气、金融、航天、传媒、城市、影视、制造、社科等领域，构建了业界规模最大的产业大模型体系。基于文心大模型体系，一方面全面升级大模型开发套件、文心 API 和提供全流程开箱即用大模型能力的 EasyDL 和 BML 开发平台，另一方面推出 AI 艺术与辅助创作平台“文心一格”、产业级搜索系统“文心百中”和畅谷社区等基于大模型技术的产品应用。我们认为，文心大模型底座+文心一格/百中应用，为文心一言的问世奠定了技术和应用基础。

图2：文心大模型全景图



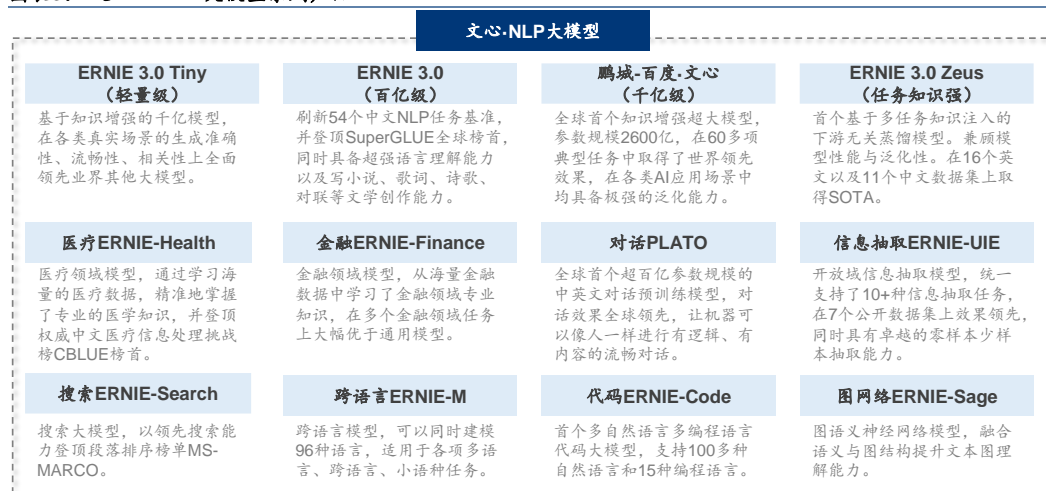
资料来源：文心大模型官网、华泰研究

文心大模型覆盖了 NLP 等多个 AI 应用场景。文心大模型包括 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。其中，NLP 大模型面向语言理解、语言生成等 NLP 场景，具备语言理解、对话生成、文学创作等能力。CV 大模型基于视觉技术，利用图像、视频等数据，提供视觉基础模型，以及视觉任务定制与应用能力。跨模态大模型可实现跨模态检索、图文生成、图片文档的信息抽取等应用的快速搭建。生物计算大模型融合生物领域研究对象的特性，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。行业大模型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。

ERNIE 3.0：打造文心一言的关键模型

文心一言是基于文心 NLP 大模型的生成式对话产品，ERNIE 模型是关键。文心 NLP 大模型打造了 ERNIE 模型系列，应用于金融、医疗、搜索、编程、图理解等领域。此外，文心 NLP 大模型还包含了全球首个知识增强超大模型鹏城-百度·文心（ERNIE 3.0 Titan），模型参数规模达到 2600 亿；全球首个超百亿参数规模的中英文对话预训练模型 PLATO，实现机器像人类一样进行有逻辑、有内容的流畅对话。文心一言英文名 ERNIE Bot，是基于 ERNIE 模型系列打造的生成式对话产品，ERNIE 是文心一言实现的关键。

图3：文心·NLP 大模型系列产品



资料来源：文心大模型官网、华泰研究

ERNIE 问世于 2019 年，目前已发展到 **ERNIE 3.0**。2019 年 4 月，ERNIE 1.0 问世（论文：ERNIE: Enhanced Representation through Knowledge Integration），是基于 BERT 模型的进一步优化，在中文的 NLP 任务上得到了当时较好的结果。2019 年 7 月，ERNIE 2.0 问世（论文：ERNIE 2.0: A Continual Pre-Training Framework for Language Understanding），该框架通过不断的多任务学习来逐步构建和学习预训练任务，最大限度地从训练语料库中提取词汇、句法和语义信息。2021 年 7 月，ERNIE 3.0 问世（论文：Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation）。目前，ERNIE 在 GLUE Benchmark（通用语言理解评估排名）上排名第五（截至 2023 年 2 月 22 日）。

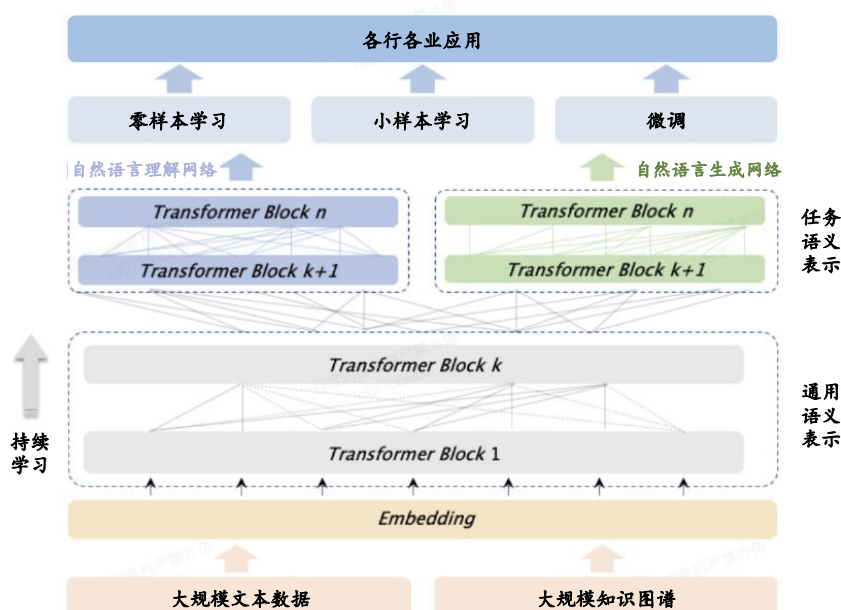
图表4：ERNIE 在 GLUE Benchmark 上排名第五（截至 2023 年 2 月 22 日）

Rank	Name	Model	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNL-m	MNL-mm	QNLI	RTE	WNLI	AX
1	Microsoft Alexander v-team	Turing ULR v6	91.3	73.3	97.5	94.2/92.3	93.5/93.1	76.4/90.9	92.5	92.1	96.7	93.6	97.9	55.4
2	JDExplore d-team	Vega v1	91.3	73.8	97.9	94.5/92.6	93.5/93.1	76.7/91.1	92.1	91.9	96.7	92.4	97.9	51.4
3	Microsoft Alexander v-team	Turing NLR v5	91.2	72.6	97.6	93.8/91.7	93.7/93.3	76.4/91.1	92.6	92.4	97.9	94.1	95.9	57.0
4	DTRL Team	DeBERTa + CLEVER	91.1	74.7	97.6	93.3/91.1	93.4/93.1	76.5/91.0	92.1	91.8	96.7	93.2	96.6	53.3
5	ERNIE Team - Baidu	ERNIE	91.1	75.5	97.8	93.9/91.8	93.0/92.6	75.2/90.9	92.3	91.7	97.3	92.6	95.9	51.7

资料来源：GLUE Benchmark 官网、华泰研究

ERNIE 3.0 是基于知识增强的多范式统一预训练框架。ERNIE 3.0 将自回归和自编码网络融合进行预训练，并在训练时引入大规模知识图谱类数据。其中，自回归网络基于 Transformer-XL 结构，支持长文本语言模型建模。自编码网络采用 ERNIE 2.0 的多任务学习增量式构建预训练任务，持续的进行语义理解学习，并增加了知识增强的预训练任务。多范式的统一预训练模式，不仅在 zero/few-shot（零样本/少样本学习）任务上展现了很强的能力，也能很好地处理传统的 fine-tune（微调）任务，使得 ERNIE 3.0 在理解任务、生成任务和零样本学习任务上取得了较好表现。

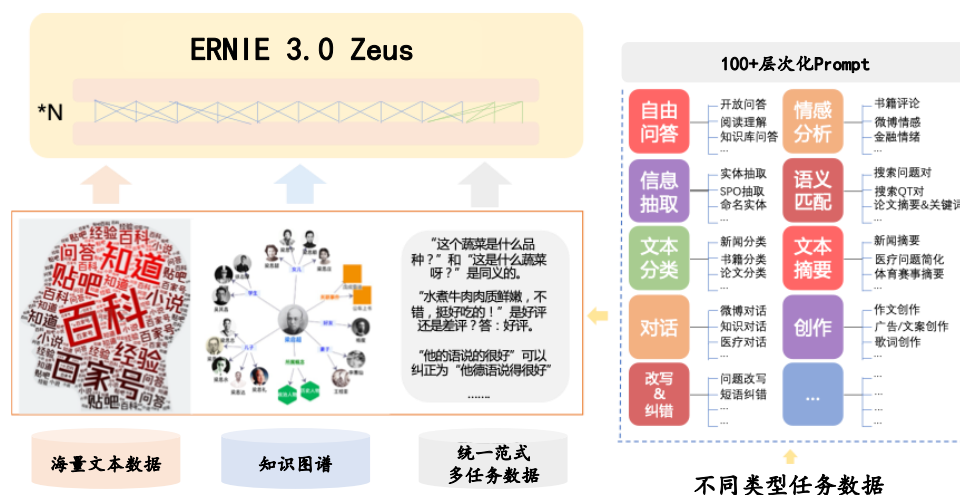
图表5：ERNIE 3.0 多范式统一预训练框架



资料来源：Large-Scale Knowledge Enhanced Pre-Training for Language Understanding And Generation、华泰研究

ERNIE 3.0 Zeus 进一步提升了模型的零样本/小样本学习能力。ERNIE 3.0 Zeus 是 ERNIE 3.0 的最新升级。ERNIE 3.0 Zeus 使用统一范式的多任务学习，建模数据中不同粒度的语义信息，并提出了层次化提示（Prompt）学习技术，显著提升了模型的零样本/小样本学习能力。在数据构造时，通过层次化的 Text Prompt 库将不同的任务统一组织成自然语言的形式，和海量无监督文本以及百度知识图谱联合学习。在训练时，引入了层次化 Soft Prompt，对不同任务之间的共性与特性进行建模，提升了模型对于不同下游任务的建模能力。

图表6：ERNIE 3.0 多范式统一预训练框架

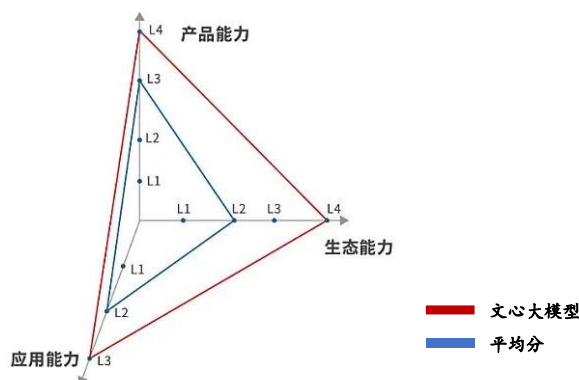


资料来源：文心大模型官网、华泰研究

文心大模型位于中国大模型市场第一梯队

文心大模型位于中国大模型市场第一梯队。IDC 在《2022 中国大模型发展白皮书》中提出了行业首个大模型评估框架，基于模型能力、工具平台能力等 6 个维度的 11 项指标，评估各模型厂商的水平。据 IDC 数据，百度文心大模型综合评估结果处于第一梯队，产品能力、生态能力、应用能力均处在领先地位，具有产业级、知识增强两大特色。文心大模型优异的能力是文心一言重要的技术保障。

图表7：中国大模型市场 2022 年评估结果



资料来源：IDC、华泰研究

ERNIE 3.0 vs GPT 系列：更适合中文 NLP 模型

ERNIE 3.0 最大特点在于多范式统一和知识图谱引入。OpenAI 的 GPT 系列模型问世于 2018 年 6 月，经过近 5 年的时间迭代了 GPT-1、GPT-2、GPT-3、Codex、InstructGPT、ChatGPT 等多个大版本。2022 年 11 月 ChatGPT 问世后，5 天之内用户突破百万。2023 年 2 月，微软宣布旗下浏览器 EDGE 和搜索引擎 Bing 接入 ChatGPT，用对话的方式取代传统的搜索。将 ERNIE 3.0 与 GPT 系列相比，其最大特点在于采用多范式统一的大规模预训练框架，融合了自回归网络+自编码网络，并在训练时引入大规模知识图谱类数据。经过 ERNIE 3.0 训练后的模型可以通过零镜头学习、少镜头学习或微调轻松地地为自然语言理解和生成任务定制，其英文版在 2021 年 7 月 SuperGLUE 基准测试(SuperGLUE 是 Facebook 人工智能研究中心、Google DeepMind、华盛顿大学以及纽约大学共同推出的用于衡量高性能语义理解 AI 的基准测试)中取得第一名，超过人类绩效 0.8% (90.6% vs 89.8%)。

图表8：ERNIE 3.0 与 GPT 系列对比

比较内容	ERNIE 3.0	GPT 系列
模型网络	自回归网络+自编码网络	自回归网络
模型框架	多范式统一的大规模预训练框架。自回归网络基于 Transformer-XL 结构。多范式统一预训练指的是基于同一个网络进行多种自然语言处理学习范式的统一学习。	全系列均以 Transformer 解码器架构为基础，每代可能会进行略微调整。
训练数据	大规模文本数据+大规模知识图谱	以大规模文本数据为主
训练参数量	参数量最大的 ERNIE 3.0 Titan 达 2600 亿	目前参数量最大为 GPT-3 的 1750 亿
训练数据集大小	4TB	目前数据集最大的 GPT-3 超 570GB
应用情况	已有 AI 作画应用文心一格和产业级搜索文心百中	已有 ChatGPT

资料来源：文心大模型官网、OpenAI 官网、华泰研究

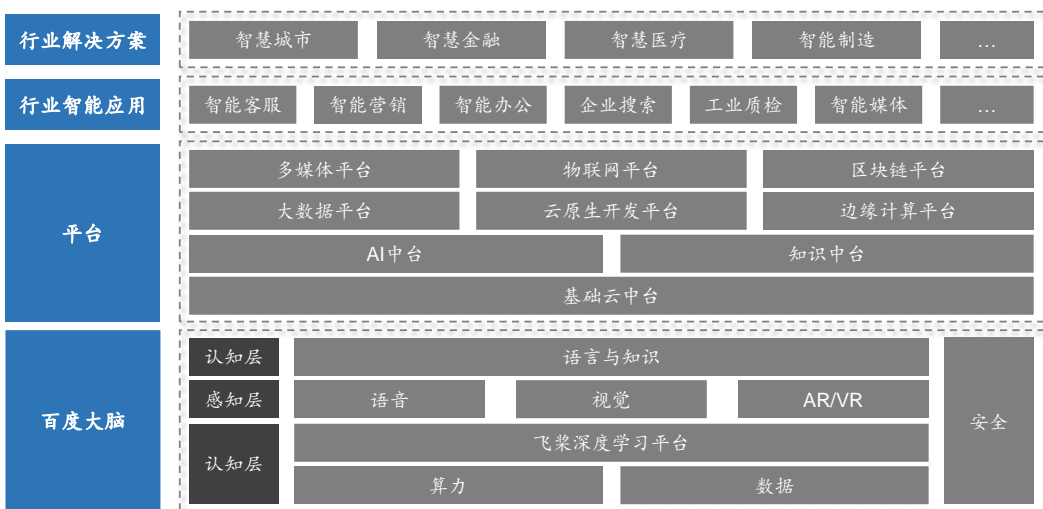
我们认为，ERNIE 3.0 作为文心一言的核心模型引擎，在英文版的测试中本身已经取得了优秀结果。在中文 NLP 方面，ERNIE 3.0 比 ChatGPT 更具有天然的语言优势，或将更好的支持中文搜索问答、内容创作生成、虚拟人物、智能客服、智能写作等应用。

百度智能云：文心一言对外提供服务的窗口

百度智能云：百度大脑+平台+应用三层架构

百度智能云的业务架构包括百度大脑+平台+应用，文心大模型位于百度大脑基础层。架构底层是百度大脑，包括基础层、感知层、认知层和安全，是百度核心技术引擎。其中，基础层的飞桨深度学习平台提供的文心大模型，是文心一言的技术基础。中间是平台层，包括通用的基础云平台、AI 中台、知识中台，以及针对场景的平台和其他关键组件。上层为智能应用和解决方案层，在基础层和平台的支持下为各行各业赋能。同时，百度智能云打造了一体化的安全体系，全面覆盖 AI 模型安全到行业生态安全。

图表9：百度智能云业务新架构



资料来源：百度智能云官方公众号、华泰研究

云智一体+AI 大底座打通文心一言基础设施与应用窗口

文心一言通过百度智能云对外提供服务。2023 年 2 月，在 AI+工业互联网高峰论坛上，百度智能云宣布“文心一言”将通过百度智能云对外提供服务。百度智能云采用云智一体架构，以云计算为基础，以 AI 为抓手，借助百度大脑、飞桨平台等对传统生态输出 AI 能力，为企业和开发者提供全球领先的人工智能、大数据和云计算服务。2022 年，百度智能云推出“云智一体 3.0”架构，形成了一套涵盖软硬件的“芯片-框架-大模型-行业应用”智能化闭环路径，环与环之间通过关键自研技术形成信息反馈，实现端到端优化。

“云智一体 3.0”架构包含四层：

- 1) 行业层**：切入重点行业里的核心场景，例如制造业中的质量监控、安全生产和工厂节能等，发挥 AI 对行业的赋能作用。
- 2) 通用产品层**：将不同行业的通用需求，沉淀为标准化的通用 AI 产品，再进行垂直领域的行业拓展，例如智能客服、数字人等。
- 3) AI 大底座**：由 AI IaaS 和 AI PaaS 组成，整合百度自研的 AI 芯片“昆仑芯”、飞桨深度学习框架、文心大模型，面向企业 AI 开发和应用提供完整解决方案。
- 4) 通用 cloud**：满足海量的计算需求。

图表10：“云智一体 3.0”架构



资料来源：百度智能云官方公众号、华泰研究

AI 大底座为文心一言提供了智算基础设施。2022 年，百度智能云发布国内首个全栈自研的 AI 基础设施“AI 大底座”。AI 大底座在 AI IaaS 层（百舸 AI 异构计算平台）整合自研 AI 芯片“昆仑芯”，在 AI 计算、存储、加速、容器方面进行系统优化，提供高性价比的算力，承载数据的处理、超大模型的训练和推理；在 AI PaaS 层（AI 中台），整合飞桨深度学习框架及百度文心大模型，打通百度的样本中心、模型中心、AI 开发平台、AI 服务运行平台，实现从数据存储到模型训练、生产、部署、测试的全链路、批量化生产，面向企业模型开发的全生命周期提供完整解决方案。AI 大底座是各项底层 AI 技术的集大成者，通过 AI 底层技术的通用化、模块化，实现 AI 服务的规模化。

图表11：百度 AI 大底座

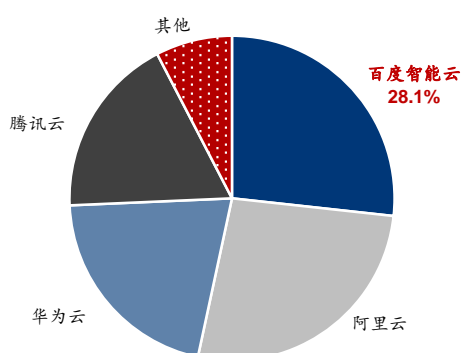


资料来源：百度智能云官方公众号、华泰研究

百度智能云：市场优势明显，云服务市场份额第一

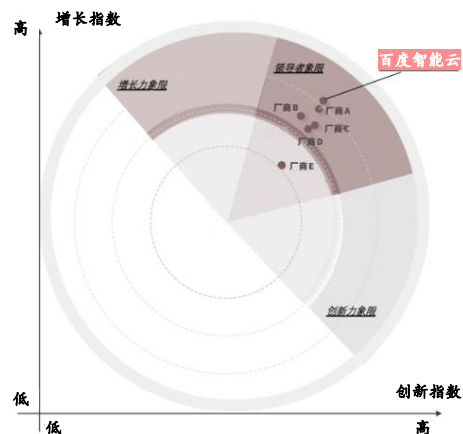
2022 年上半年，百度智能云在中国 AI 公有云服务市场份额第一。据 IDC《2022H1 中国 AI 云服务市场研究报告》数据，2022 年上半年中国 AI 公有云服务市场规模将达 74.6 亿元，百度智能云稳居第一（连续四年），整体市场份额占比 28.1%，并且在人脸人体、图像视频两个规模最大的子市场继续保持第一。同时，在增速较快的对话式 AI 市场中，据沙利文联合头豹研究院共同发布的《2021 年中国对话式 AI 市场报告》数据，百度智能云增长指数、创新指数均排名第一，稳居市场领导者梯队。百度智能云的市场地位，为文心一言的广泛推广打下了基础。

图表12：1H22 中国人工智能公有云服务市场份额



资料来源：IDC、华泰研究

图表13：中国对话式 AI 市场综合竞争表现



资料来源：沙利文研究院、华泰研究

从文心一格/文心百中展望文心一言能力

文心一格和文心百中均是基于文心大模型的产品级应用，与文心一言定位相似。文心一格和文心百中是目前文心大模型成功应用的范例。其中，文心一格基于文心大模型中的文图生成模型 ERNIE-ViLG，主要实现 AI 作画应用。文心百中是基于文心 ERNIE 大模型的端到端搜索引擎，用来替代传统搜索引擎复杂的特征及系统逻辑。未来推出的文心一言，与文心一格和文心百中具有相似的定位，或将共同补全文心大模型在对话生成、图像生成和搜索等领域的应用图谱。

我们认为，技术上来说，文心大模型已经具备了搜索、文图生成等功能，并成功得到应用，这些能力或将集成于文心一言。此外，据百度官方信息，将提供基于 ERNIE 3.0 的文本理解与创作 API，预置作文生成、文案创作、情感分析等任务提示（prompt），支持用户自定义 prompt，模型根据零样本或少样本的输入提示生成结果。这些功能或将同样与文心一言息息相关。

文心一格：ERNIE-ViLG 在产品层面的应用

文心一格是文心 AI 作画大模型 ERNIE-ViLG 在产品层面的应用。2022 年 8 月 19 日，在中国图象图形大会 CCIG 2022 上，AI 艺术和创意辅助平台文心一格正式发布。用户将创想以自然语言的形式输入到模型中，并设定化作风格，模型即可根据需求生成相关画作。在 WAVE SUMMIT+2022 深度学习开发者峰会上发布的新版本文心一格，在此前的基础上，增加支持以图生图、可控的增删改等图片二次编辑和一键生成视频。目前，用户通过充值的方式实现文心一格的画作创作。

图表 14：文心一格创作过程



资料来源：文心一格、华泰研究

图表 15：文心一格部分作品

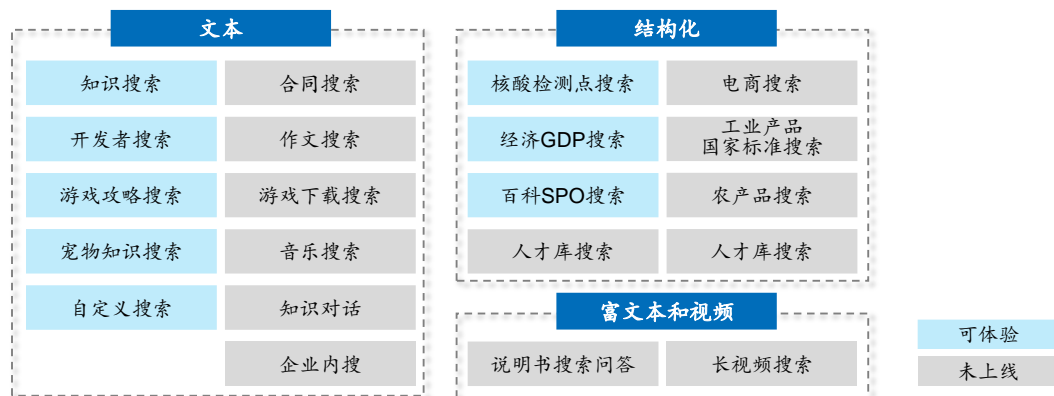


资料来源：文心一格、华泰研究

文心百中：大模型驱动的产业级搜索系统

文心百中是基于大模型的端到端产业级搜索引擎。文心百中以文心 ERNIE 大模型为基础，替代了传统搜索引擎复杂的特征及系统逻辑，实现低成本接入各类企业和开发者应用，以数据驱动的优化模式提高行业效率和应用效果。目前，文心百中覆盖的搜索场景包括文本、结构化、富文本、视频等领域，并且以文本场景为主。

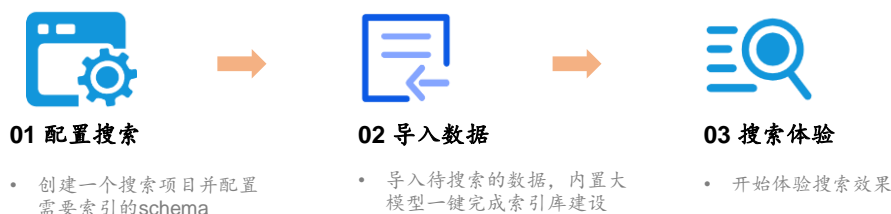
图16：产业级搜索系统文心百中的搜索场景图谱



资料来源：文心大模型官网、华泰研究

文心百中在技术和应用上均有优势。技术上，1) 架构：技术架构简单，采用以大模型为核心的端到端搜索架构，一套架构支持各类异构数据搜索；2) 优化：采用数据驱动的低门槛效果优化，以及点击日志自动采集的自我优化；3) 语义理解：单模型刷新 60+ 中文 NLP 任务基准，并登顶 SuperGLUE（一种特征匹配网络）全球榜首。应用上：1) 部署：支持容器化一键部署，支持可视化控制台、命令行、API 操作；2) 迁移：大模型具备强大的小样本能力，行业迁移所需的标注数据量少，部分场景无需优化即可上线；3) 对接：不同类型的异构数据可复用一套模型、一套接口，直接建库直接搜索；4) 系统建设：以 AI 大模型替代传统复杂的人工策略构建，导入待搜索的数据即可完成索引库的建设。

图17：三步搭建搜索系统



资料来源：文心大模型官网、华泰研究

文心百中提供体验版和正式版两个版本。两个版本的区分体现在模型搜索效果、部署、性能和功能上。体验版用户可以申请本地部署，正式版目前提供试用通道。文心百中支持知识搜索、开发者搜索、经济 GDP 搜索、游戏攻略搜索、宠物知识搜索、百科 SPO 搜索等，搜索结果相比传统搜索引擎的关键词搜索更为精确。此外，正式版的自定义搜索功能，可以支持用户上传本地的数据后，获取基于大模型的语义搜索能力。

图18：文心百中提供体验版和正式版两个版本

	体验版	正式版
模型	基础搜索效果	极致搜索效果
部署	快速部署方案	企业级部署方案
性能	基础性能	极致性能
功能	语义搜索功能	更多搜索能力、自主模型优化

资料来源：文心大模型官网、华泰研究

图19：文心百中搜索结果示例



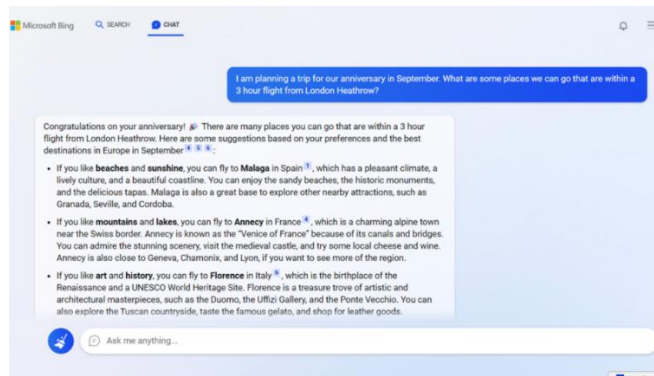
资料来源：文心大模型官网、华泰研究

应用方式探索：搜索引擎+API+生态融合

可能的应用方式#1：文心一言+搜索引擎

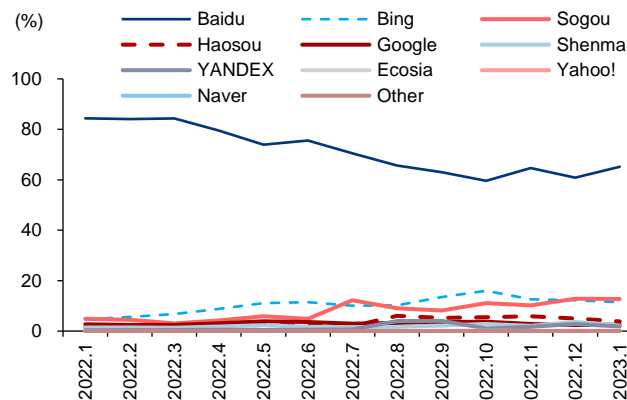
微软证明了搜索引擎+类 GPT 产品是可行路线。2023 年 2 月 8 日，微软发布基于 ChatGPT 的新版 EDGE 浏览器和 Bing 搜索引擎，在提供传统搜索内容的同时，用户可以与搜索引擎进行对话交流，获得更全面的答案。反观国内，百度搜索在国内市场份额处于绝对的领先地位，据 statcounter 数据，截至 2023 年 1 月，百度搜索在国内搜索引擎市场份额为 65.2%。我们认为，文心一言+百度搜索在国内具有较大优势，且微软与 ChatGPT 的融合已经证明了该模式的有效性。

图表20：集成 ChatGPT 的 Bing



资料来源：微软官网、华泰研究

图表21：百度搜索引擎在国内的市场份额情况



资料来源：statcounter、华泰研究

可能的应用方式#2：大模型 API

文心一言或将提供大模型 API 相关功能。据文心大模型官网信息，目前，文心大模型提供的大模型 API 包括 ERNIE-ViLG 文生图和 PLATO，以及正在开发的 ERNIE 3.0 文本理解与创作。ERNIE 3.0 文本理解与创作与文心一言官网相关联，我们认为，文心一言等生成式对话产品或将同样提供大模型 API 相关功能。

图表22：文心大模型提供的大模型 API



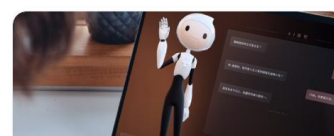
ERNIE 3.0 文本理解与创作

- 预置作文生成、文案创作、情感分析等任务提示（prompt），支持用户自定义 prompt，模型根据零样本或少样本的输入提示生成结果。



ERNIE-ViLG 文生图

- 文生图领域的大模型服务，支持输入一段文本描述，并选择生成风格和分辨率，模型会根据输入的内容自动创作出符合要求的图像。



PLATO

- 提供基于 PLATO 大模型的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。

资料来源：文心大模型官网、华泰研究

ERNIE-ViLG 文生图是文心大模型中跨模态文图生成大模型。ERNIE-ViLG 可通过自然语言实现图像生成与编辑，并支持 API 集成服务能力。最新版本 ERNIE-ViLG 2.0 采用基于知识增强算法的混合降噪专家建模，是全球首个知识增强、参数规模最大的 AI 作画大模型，在文本生成图像公开权威评测集 MS-COCO 和人工盲评上超越了 OpenAI 的 DALL-E 2 等模型，在语义可控性、图像清晰度、中国文化理解等方面优势明显。

图表23: ERNIE-ViLG AI 作画大模型套餐类型

50张	600张	私有化部署 更高QPS使用
9.9元	109元	
10000张	100000张	
1500元	13000元	合作咨询

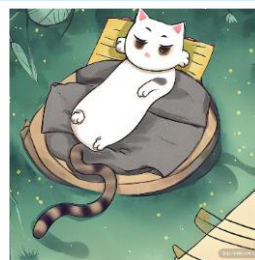
购买后可享受：专业技术支持、7*12小时人工答疑（9:00-21:00）、API接口调用时水印去除、专属服务器加速生成

资料来源：文心大模型官网、华泰研究

图表24: 基于 Prompt “一只猫在晒太阳，卡通” 生成的图片

请用中文输入Prompt，参考形式：画面主体，细节描述，修饰词

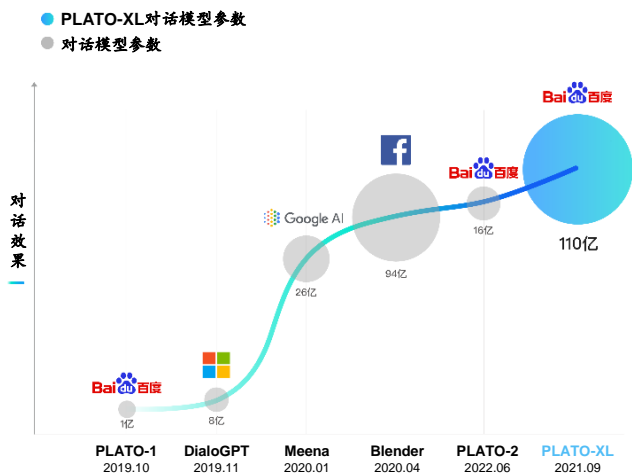
——有一只猫在晒太阳，卡通



资料来源：文心大模型官网、华泰研究

文心 PLATO 是大规模开放域对话模型，支持 API 调用。PLATO 与 ERNIE 同样属于 NLP 大模型。PLATO 是世界首个基于隐变量的生成式开放域对话大模型，利用隐变量建模开放域对话中的一对多关系（一个输入对应多个正确输出）。模型采用 Unified Transformer 框架共享生成模型中的编码器和解码器参数，通过课程学习方式提升模型训练效率，在大规模高质量对话语料上基于飞桨深度学习框架进行训练。PLATO 具备接近真人水平的多轮流畅对话能力，其中英文模型在开放域下的对话合理性、丰富度、吸引度等各项指标上达到领先水平。PLATO 同样支持 API 调用，但据官网消息，目前内测已满额，暂不开放体验申请。

图表25: PLATO 的对话效果达到了世界领先水平



资料来源：文心大模型官网、华泰研究

图表26: PLATO 的 API 调用服务暂不可用



资料来源：文心大模型官网、华泰研究

可能的应用方式#3: 产品级应用+生态融合

打造产品级应用，实现生态圈融合。产品上，文心一言可能打造与文心一格、文心百中类似的产品级应用，通过直接的方式提供服务，类似 ChatGPT 通过网页即可实现访问。生态上，结合百度自身的自动驾驶、视频等已有生态，融合文心一言的新搜索形态，或可以接入 Apollo 自动驾驶平台、爱奇艺、小度等平台终端，扩展使用场景。此外，虚拟人也是可能的应用方向之一，例如官宣文心一言的新闻编辑人“希加加、度晓晓、叶悠悠、林开开”，或是基于文心大模型的虚拟人。

接入情况：超 450 家企业宣布加入“文心一言”生态圈

超 450 家企业已宣布加入“文心一言”生态圈。截至 2023 年 2 月 25 日，涵盖互联网、媒体、金融、汽车、企业软件等行业的超 450 家企业已宣布加入“文心一言”生态圈。据 2023 年数智金融峰会信息，“对于金融行业来说，文心一言将率先在智能检索、投研助手、金融数字人、智能客服、智能创作等场景落地”。其他落地场景或将包括智能客服、智能对话、智能创作、知识管理等内容、信息相关的场景。

图表 27：首批加入文心一言生态圈的部分伙伴

序号	行业	公司名称
1	银行	邮储银行
2		百信银行
3		重庆银行
4		新网银行
5	汽车	吉利汽车
6		集度汽车
7		零跑汽车
8	能源	能链
9	互联网	36 氪
10		新浪
11		携程
12	IT 软件	中软国际
13		软通动力
14		东软集团
15		致远互联
16	媒体	人民政协报
17		工信部新闻宣传中心
18		新民晚报
19		澎湃新闻
20		环球时报

资料来源：百度智能云官方公众号、华泰研究

相关公司梳理

文心一言等类 ChatGPT 产品产业链涉及公司众多，主要包括前台应用开发商，以及后台算力和基础设施提供商。我们从相关应用、算力芯片、服务器、IDC 服务等角度，对产业链相关公司进行如下梳理：

- 1、相关应用厂商：科大讯飞、金山办公、同花顺、宇信科技、汉得信息；
- 2、算力芯片厂商：景嘉微、寒武纪、海光信息、龙芯中科、中国长城等；
- 3、服务器厂商：浪潮信息、中科曙光等、拓维信息；
- 4、IDC 服务商：宝信软件等。

图表28：提及公司列表

公司代码	公司简称
BIDU US	百度
未上市	OpenAI
002230 CH	科大讯飞
688111 CH	金山办公
300033 CH	同花顺
300674 CH	宇信科技
300170 CH	汉得信息
300474 CH	景嘉微
688256 CH	寒武纪
688041 CH	海光信息
688047 CH	龙芯中科
000066 CH	中国长城
000977 CH	浪潮信息
603019 CH	中科曙光
002261 CH	拓维信息
600845 CH	宝信软件

资料来源：Bloomberg、华泰研究

风险提示

宏观经济波动。若宏观经济波动，产业变革及新技术的落地节奏或将受到影响，宏观经济波动还可能对 IT 投资产生负面影响，从而导致整体行业增长不及预期。

下游需求不及预期。若下游数字化需求不及预期，相关的数字化投入增长或慢于预期，致使行业增长不及预期。

本报告内容均基于客观信息整理，不构成投资建议。

免责声明

分析师声明

本人，谢春生，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。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。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、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的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师谢春生本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期

（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15% 以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15% 以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格, 经营许可证编号为: 91320000704041011J

香港: 华泰金融控股(香港)有限公司具有香港证监会核准的“就证券提供意见”业务资格, 经营许可证编号为: AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员, 具有在美国开展经纪交易商业业务的资格, 经营业务许可编号为: CRD#:298809/SEC#:8-70231

华泰证券股份有限公司**南京**

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码: 210019

电话: 86 25 83389999/传真: 86 25 83387521

电子邮件: ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码: 518017

电话: 86 755 82493932/传真: 86 755 82492062

电子邮件: ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码: 100032

电话: 86 10 63211166/传真: 86 10 63211275

电子邮件: ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码: 200120

电话: 86 21 28972098/传真: 86 21 28972068

电子邮件: ht-rd@htsc.com

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心58楼5808-12室

电话: +852-3658-6000/传真: +852-2169-0770

电子邮件: research@htsc.com

<http://www.htsc.com.hk>

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话: +212-763-8160/传真: +917-725-9702

电子邮件: Huatai@htsc-us.com

<http://www.htsc-us.com>

©版权所有2023年华泰证券股份有限公司