

计算机行业研究

买入（维持评级）

行业深度研究

证券研究报告

计算机组

分析师：王倩雯（执业 S1130522080001） 分析师：孟灿（执业 S1130522050001）

wangqianwen@gjzq.com.cn

mengcan@gjzq.com.cn

“文心一言”发布，国内应用预计加速

投资逻辑

百度于 2023 年 3 月 16 日召开关于“文心一言”的主题发布会，“文心一言”作为国内首款正式发布的生成式语言大模型，拥有文学创作、商业文案创作、数理逻辑推算、中文理解、多模态生成等功能。文心平台依托于百度自研的产

- **AI 大模型与行业大模型：**文心平台提供 NLP、CV、跨模态、生物计算四大主流 AI 研究领域的多个基础大模型，多个模型在技术层面实现突破创新，处于世界领先水平。其中，NLP 领域的 ERNIE 3.0 Zeus 采用自回归网络和自编码网络，能够兼顾自然语言理解和生成功能；CV 领域的 VIMER-UFO 2.0 使用超网络结构，推理时仅需激活总参数数量的 1/30；ERNIE-ViLG 2.0 引入基于语言和图像的知识增强算法，在文本图像生成任务中的表现已经超越了 OpenAI 旗下的 DALLIE-2。文心平台与各行业企业联手打造行业大模型，目前已覆盖能源、金融、航天、制造、传媒、城市、社科、电影等 8 个领域。目前已有 650 余家知名企业宣布接入“文心一言”，涵盖企业服务、金融 IT、汽车、传媒、教育、家电、金融等 10 余个行业。
- **一站式开发平台与开发工具：**开发平台方面，文心面向专业知识有限的 AI 应用开发者提供零门槛平台 EasyDL，面向专业开发者提供全功能平台 BML。目前这两个平台均仅支持使用百度文心的云端算力，可采用公有云、私有化离线、软硬一体等部署方式。开发工具方面，文心提供 ERNIEKit 自然语言处理开发工具及 PaddleFleetX 全流程开发套件。此外，现阶段文心平台还向开发者免费提供大模型 API。
- **产品与社区：**百度文心通过构建产品与社区经营，对外展示生成式模型的强大性能，以此吸引爱好者投入创作，进而构筑成熟的 AIGC 社区，可在实践中探索 AI 商业化模式。目前除“文心一言”外，文心平台提供产业级搜索系统“文心百中”、艺术创意辅助平台“文心一格”，以及采用众创模式的晒谷社区。

目前百度文心生成式语言大模型在参数规模、技术创新、应用落地方面均在国内同业中处于领先水平。3 月 16 日起，“文心一言”首批用户可通过邀请测试码接入体验，同时百度智能云将面向企业客户开放“文心一言”API 调用接口，有望进一步丰富百度 AI 生态，赋能更多行业伙伴。

投资建议

相较于 ChatGPT 等海外模型存在连接限制，百度文心平台可以为国内用户提供更便捷的技术入口。“文心一言”的发布将极大加速国内生成式 AI 模型的落地应用节奏，有望帮助生态伙伴实现用户体验提升和生产降本增效。建议关注接入“文心一言”生态的公司，如恒生电子、用友网络、汉得信息、凌志软件、宇信科技等。

风险提示

海外基础软硬件使用受限；应用落地不及预期；行业竞争加剧风险。

内容目录

1. 百度文心平台可提供一站式 AI 开发服务.....	3
2. 文心平台提供四大主流领域 AI 大模型及行业大模型.....	3
3. 文心平台提供一站式开发平台与工具.....	9
4. 文心平台提供搜索、创作等产品.....	12
5. 国内其他生成式语言模型.....	14
6. 投资建议.....	14
7. 风险提示.....	14

图表目录

图表 1: 百度文心平台全景图.....	3
图表 2: 文心 NLP 大模型可支持语言理解、语言生成等功能.....	3
图表 3: ERNIE 3.0 Zeus 创新性地提出知识增强方法.....	4
图表 4: 文心 NLP 大模型技术演进可划分为 3 条主线.....	5
图表 5: 文心 CV 大模型可支持一站式视觉任务定制和应用服务.....	5
图表 6: VIMER-UFO 2.0 采用超网络结构.....	6
图表 7: 文心跨模态大模型支持多场景应用.....	6
图表 8: ERNIE-ViLG 2.0 模型引入知识增强算法.....	7
图表 9: 文心生物大模型在多任务中已达到世界领先水平.....	7
图表 10: 文心行业大模型拥有丰富的合作案例.....	7
图表 11: 文心行业大模型融合知识增强技术.....	8
图表 12: 已有 650 余家公司接入百度“文心一言”.....	9
图表 13: EasyDL 提供一站式零代码 AI 模型开发.....	10
图表 14: EasyDL 可支持多场景任务.....	10
图表 15: BML 全功能 AI 开发平台架构概览.....	11
图表 16: EasyDL 与 BML 提供的模型部署方式.....	11
图表 17: EasyDL 与 BML 提供的模型部署方式.....	11
图表 18: 百度文心提供大模型 API.....	12
图表 19: 文心百中使用流程简单.....	13
图表 20: 文心一格收费价格 0.2-0.8 元/张.....	13
图表 21: 晒谷社区部分创意应用.....	13
图表 22: 国内主流生成式语言模型对比.....	14

1. 百度文心平台可提供一站式 AI 开发服务

百度于 2023 年 3 月 16 日召开关于“文心一言”的主题发布会，“文心一言”作为国内首款正式发布的生成式语言大模型，可支持文学创作、商业文案创作、数理逻辑推算、中文理解、多模态生成等功能。文心平台依托于百度自主研发的产业级深度学习平台飞桨（PaddlePaddle）打造，集先进的全领域 AI 大模型、全面的 AI 算法集、端到端开发套件和平台化服务于一体，致力为用户提供一站式 AI 开发服务。

图表1：百度文心平台全景图

产品与社区	文心一格 AI艺术和创意辅助平台		文心百中 大模型驱动的产业级搜索系统		畅谷社区 大模型创意与探索社区		
工具与平台	EasyDL-大模型 零门槛 AI 开发平台		BML-大模型 全功能 AI 开发平台		大模型 API		
	大模型套件						
	数据标注与处理	大模型精调	大模型压缩	高性能部署		场景化工具	
文心大模型	行业大模型						
	国网-百度文心	浦发-百度文心	航天-百度文心	人民网-百度文心	冰城-百度文心	电影频道-百度文心	
	深燃-百度文心	吉利-百度文心	泰康-百度文心	TCL-百度文心		辞海-百度文心	
	NLP 大模型		CV 大模型		跨模态大模型		生物计算大模型
	医疗 ERNIE-Health	金融 ERNIE-Finance	商品图文搜索表征学习 VIMER-UMS		文图生成 ERNIE-VILG	文档智能 ERNIE-Layout	化合物表征学习 HelixGEM
	对话 PLATO	搜索 ERNIE-Search	信息抽取 ERNIE-UIE	OCR图像表征学习 VIMER-StrucText			
	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		蛋白质结构预测 HelixFold	
	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	单序列蛋白质结构预测 HelixFold-Single	
ERNIE 3.0 Tiny (轻量级)	ERNIE 3.0 (百亿级)	鹏城-百度文心 (千亿级)	ERNIE 3.0 Zeus (任务知识增强)	视觉-语言 ERNIE-VIL	语音-语言 ERNIE-SAT	地理-语言 ERNIE-GeoL	

来源：百度文心，国金证券研究所

2. 文心平台提供四大主流领域 AI 大模型及行业大模型

文心大模型是百度文心平台的核心，包括自然语言处理（NLP）、计算机视觉（CV）、跨模态、生物计算四个主流 AI 研究领域下的多个基础大模型，此外百度文心还基于基础大模型与各行业企业深度合作、通过学习行业特色数据与知识而形成多个行业大模型。

2.1 文心 NLP 大模型

文心 NLP 大模型是百度文心着重布局的领域，其主要面向语言理解、语言生成等 NLP 场景，包含一系列在功能和应用场景上各有侧重的语言大模型。

图表2：文心 NLP 大模型可支持语言理解、语言生成等功能

模型名称	应用场景	特点
ERNIE 3.0 Zeus	问答、智能创作、情感分析等各类自然语言理解和生成任务	<ul style="list-style-type: none"> 文心旗下最先进的自然语言大模型，参数规模达千亿级，创新地采用知识增强方法。在各类真实场景的生成准确性、流畅性、相关性上全面领先业界其他大模型。
ERNIE 3.0 Tiny	各类自然语言理解和生成任务	<ul style="list-style-type: none"> ERNIE 3.0 的轻量级版本，应用模型压缩技术大幅降低了推理算力需求，在效果和 ERNIE 3.0 持平的情况下获得数十倍至上百倍的速度提升。
ERNIE-Finance	金融场景相关的 NLP 任务	<ul style="list-style-type: none"> 在海量金融领域文本和通用文本上进行训练，具备丰富的金融领域知识，在金融问答，金融信息抽取等任务上表现较好
ERNIE-Health	医疗场景相关的 NLP 任务	<ul style="list-style-type: none"> 学习了海量的医疗专业术语和医疗专业问答数

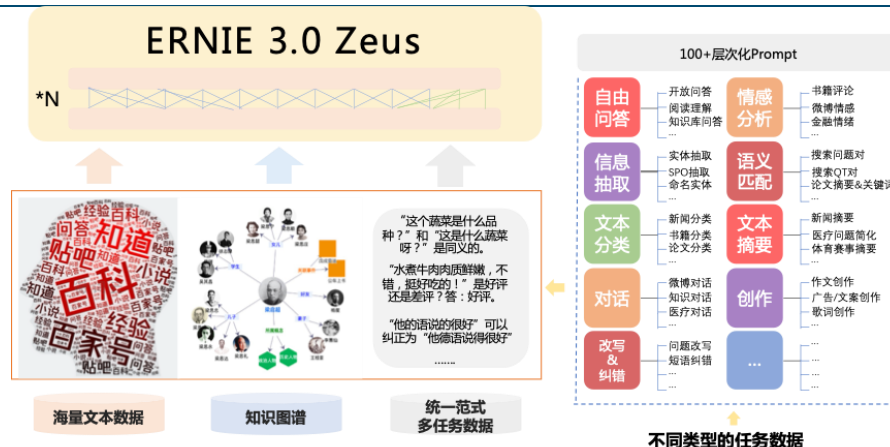
模型名称	应用场景	特点
ERNIE-Code	自然语言代码生成	<ul style="list-style-type: none"> 据，以超越人类医学专家水平的成绩登顶中文医疗信息处理权威榜单 CBLUE 采用多语言多任务联合训练的方式，支持 100 多种自然语言和 15 种编程语言。
ERNIE-UIE	产品反馈分析、服务质量评测、舆情监测等信息抽取场景	<ul style="list-style-type: none"> 用户可以使用自然语言自定义抽取目标，无需训练即可抽取输入文本中的对应信息。
ERNIE-M	跨语言各类任务	<ul style="list-style-type: none"> 突破了语料规模对多语言模型的学习效果限制，能够从单语语料中学习语言间的语义对齐关系
ERNIE-Search	检索场景的多种任务	<ul style="list-style-type: none"> 以领先搜索能力登顶段落排序榜单 MS-MARCO。
ERNIE-Sage	搜索、推荐、问答等任务	<ul style="list-style-type: none"> 能够同时对文本语义与图结构信息进行建模，显著提升了其在推荐场景的表现
PLATO	对话场景	<ul style="list-style-type: none"> 全球首个基于隐变量方式对话大模型，具备接近真人水平的多轮流畅对话能力，对话效果达到世界领先水平

来源：百度文心，国金证券研究所

ERNIE 3.0 Zeus 是目前百度文心旗下最先进的自然语言大模型，发布于 2022 年 5 月。在架构设计上，ERNIE 3.0 Zeus 采用 Transformer 作为骨干网络，并创新性地融合了自回归网络和自编码网络，使得模型同时具备了自然语言理解和自然语言生成能力。目前，以 ChatGPT 为首的大规模生成式预训练语言模型均仅基于自回归网络进行训练，导致这类模型在自然语言理解任务上表现较弱。

在模型训练上，ERNIE 3.0 Zeus 在 GPT-3 等传统大规模语言模型上做出了突破，创新性地提出了知识增强方法，在海量文本数据之外，将大规模知识图谱也引入模型训练过程，并持续对百余种不同形式的任务数据进行学习，实现了任务知识增强。相比于仅使用文本数据进行训练的传统大规模生成式预训练语言模型，ERNIE 3.0 Zeus 对常识知识掌握更好，具备更强的知识推理能力。

图表3: ERNIE 3.0 Zeus 创新性地提出知识增强方法



来源：百度文心，国金证券研究所

文心 NLP 大模型在技术演进上可以划分为 ERNIE 模型迭代、在多模态领域的进展、PLATO 模型迭代三条主线。

- 主线一：文心 ERNIE 的版本迭代。一方面，从 2019 年的文心 ERNIE1.0 到 2022 年的 ERNIE 3.0 Zeus 的版本迭代中，文心 NLP 大模型在参数规模上实现了指数级

增长，最新的 ERNIE 3.0 Zeus 在参数规模上已与国际先进水平到达同一量级。

另一方面，百度文心在 NLP 大模型研发的四年中，创新性地提出了知识增强方法，这使文心 NLP 大模型具备比 ChatGPT 等同类语言模型更强的常识掌握和知识推理能力，并在 SuperGLUE 和 GLUE 等复杂语言理解榜单中多次登顶。

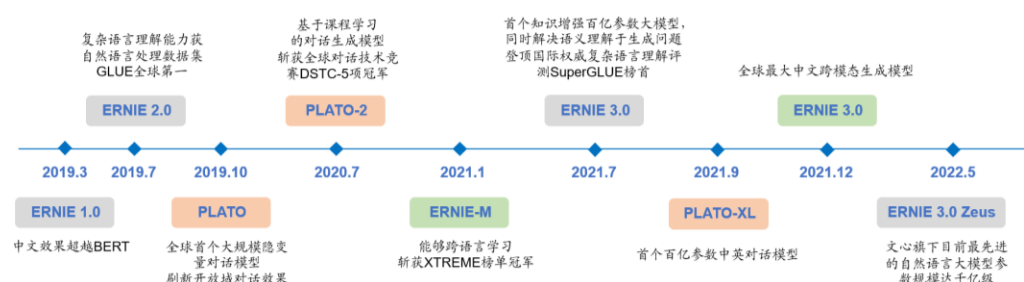
- 主线二：文心 ERNIE 在跨模态、跨语言任务方面的进步。一方面，跨语言学习任务的突破打破了语料规模对多语言模型的学习效果限制。长久以来，互联网中存在的高质量中文文本数据显著低于英文文本，跨语言学习使得模型可以从英文数据中学习适用于中文文本的语言联系，大大降低了模型训练对海量中文训练数据的要求。

另一方面，跨模态学习任务的突破为百度文心 NLP 大模型的发展奠定了良好的技术基础。百度文心在 ERNIE-Sage 首次对图结构进行了学习，使 AI 模型获得了同时学习文本信息和图结构的能力。知识增强技术是百度文心 NLP 大模型的核心技术路线之一，而知识图谱作为知识增强技术的实现方式同样采用图结构。跨模态学习任务的突破将推动百度文心知识增强方法的演进，未来有望实现图结构和文本信息对齐，使得 AI 具备常识掌握能力。

- 主线三：对话生成大模型文心 PLATO 迭代演进。一方面，从 2019 年的 PLATO 到 2021 年的 PLATO-XL，文心语言大模型实现了参数规模的迅猛增长，这使得文心语言大模型的智能水平出现了较为显著的提高。

另一方面，PLATO 在技术路线上创新性的采用了基于隐变量的模型构建方式，使得模型在不降低鲁棒性的前提下，开放域对话生成能力进一步增强。

图表4：文心 NLP 大模型技术演进可划分为 3 条主线



来源：百度文心，AI 大模型公众号，国金证券研究所

2.2 文心 CV 大模型

文心 CV 大模型基于领先的视觉技术，利用海量的图像、视频等数据训练，致力于为企业和开发者提供强大的计算机视觉基础模型，以及一站式视觉任务定制与应用服务。

图表5：文心 CV 大模型可支持一站式视觉任务定制和应用服务

模型名称	应用场景	特点
VIMER-UFO 2.0	智慧城市	视觉多任务统一大模型，可针对特定任务抽取轻量级小模型，兼顾大模型效果和小模型推理性能，能够有效减少推理算力消耗，易于落地部署
VIMER-CAE	图像分类、图像检测、图像分割	先进的视觉自监督预训练模型，在图像分类、目标检测、语义分割等达到世界领先水平
VIMER-StrucTexT 2.0	OCR 文字识别	端到端文档 OCR 表征学习预训练模型，在 5 项不同文档图像理解任务上刷新记录
VIMER-UMS	商品识别、多模态搜索与推荐、零售快消数字化等	基于海量的互联网商品图文信息训练，显著改善商品视觉检索和商品多模态检索体验

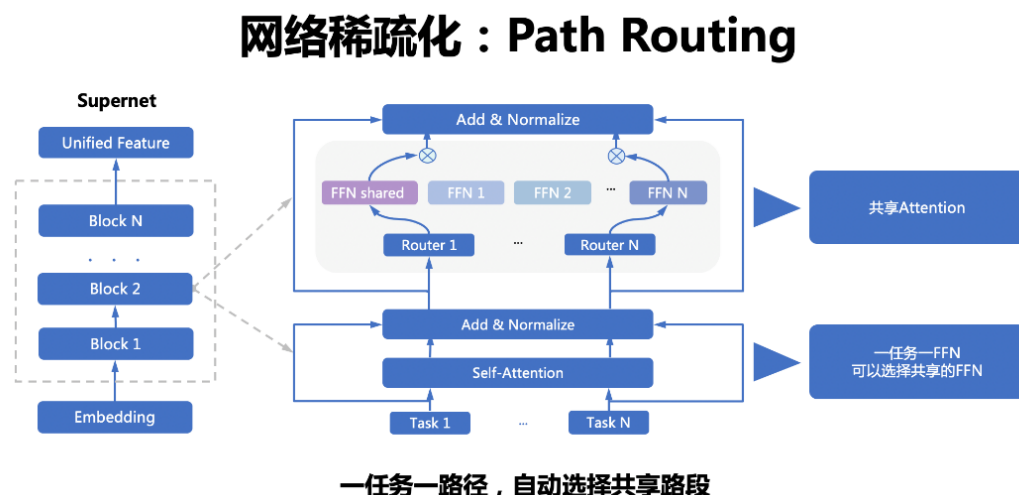
来源：百度文心，国金证券研究所

VIMER-UFO 2.0 是目前百度文心旗下最先进的视觉大模型，其创新性的提出了视觉多任

务的超网络与训练方案，使模型无需对下游任务进行微调，便可直接应用处理多个任务，以智慧城市场景为例，VIMER-UFO 2.0 可以用单模型实现人脸识别、人体和车辆多重识别等场景任务。

VIMER-UFO 2.0 在底层架构上采用超网络结构，超网络结构是一种稀疏神经网络，模型在进行推理时只需激活对应的部分参数，为每个独立任务自动分配各自特定的参数路径和共享路段，可有效减少推理算力消耗。VIMER-UFO 2.0 模型参数为 170 亿，而在进行任务推理时只需激活部分参数，实际计算量仅约为 6 亿参数，加速比接近 30 倍。

图表6: VIMER-UFO 2.0 采用超网络结构



来源：百度文心，国金证券研究所

超网络的结构设计还使得模型具备灵活、可伸缩的弹性部署能力，当用户仅需处理单个场景任务时，可以根据具体任务独立抽取对应的子网络进行部署，而无需部署完整大模型，大大降低了 AI 的落地成本。

2.3 文心跨模态大模型

文心跨模态大模型基于知识增强技术，对跨模态语义信息理解较好，可实现跨模态检索、图文生成、图片文档的信息抽取等跨模态场景应用的快速搭建，助力 AI 技术产业化落地。

图表7: 文心跨模态大模型支持多场景应用

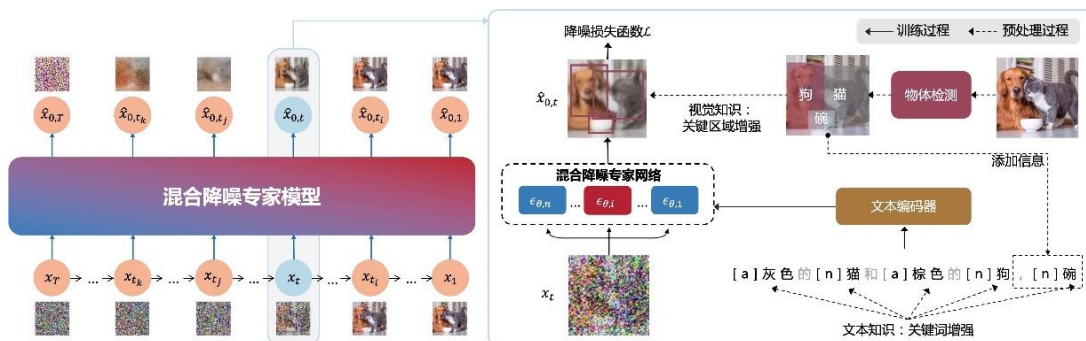
模型名称	应用场景	特点
ERNIE-ViLG 2.0	图像生成、艺术创作、虚拟现实、AI 辅助设计等	百度文心旗下最先进的 AI 画图模型，在文本生成图像任务中刷新世界最好效果
ERNIE-ViL	视觉常识推理、视觉问答、跨模态检索、引用表达式理解等	业界首个融合场景图知识的多模态预训练模型，在多项经典多模态任务中刷新了世界纪录
ERNIE-Layout	文档分类、信息抽取、文档问答等	采用文档布局知识增强技术，在 4 项文档理解任务上刷新世界纪录
ERNIE-SAT	语音编辑、语音生成、语音克隆等语言-文本跨模态场景	采用语音-文本联合训练，同时支持中文和英文
ERNIE-GeoL	地理-语言跨模态场景	模型充分学习地理和语言之间的关联，已在百度地图 POI 检索等场景实现落地

来源：百度文心，国金证券研究所

文心跨模态大模型继承了文心语言大模型的技术路线，将知识增强算法引入跨模态大模型中，使模型具备了更加精确的语义理解能力。其中，ERNIE-ViLG 2.0 作为百度文心旗下

最先进的 AI 画图模型，基于混合降噪专家网络构建，在图像生成中的不同阶段选择不同的专家网络，有效地解决了 AI 作图中不同阶段对模型能力要求不一致的问题。ERNIE-ViLG 2.0 还引入了基于语言和图像知识的知识增强算法，使模型更关注图像与文本中的主体要素，大大提升了生成图像的质量。ERNIE-ViLG 2.0 在文本图像生成任务中表现已超越了 OpenAI 于 2022 年 4 月推出的 DALL-E 2。

图表8: ERNIE-ViLG 2.0 模型引入知识增强算法



来源: 百度文心, 国金证券研究所

2.4 文心生物计算大模型

文心生物计算大模型融合了自监督和多任务学习方法，将生物领域中化合物分子、蛋白质分子等研究对象的特性融入模型，构建了先进的生物计算领域预训练模型，有望赋能生物医药行业。

图表9: 文心生物大模型在多任务中已达到世界领先水平

模型名称	应用场景	特点
HelixGEM-2	小分子药物研发	<ul style="list-style-type: none"> 业界首个考虑原子间多体交互、长程相互作用的模型，在量子化学属性预测和虚拟筛选双场景上达到领先效果
HelixFold-Single	蛋白质结构预测	<ul style="list-style-type: none"> 业界首个开源的基于单序列语言模型的蛋白结构预测大模型，预测效果达世界领先水平
HelixFold	蛋白质结构预测	<ul style="list-style-type: none"> 全国产化软硬件适配的高效蛋白质结构分析大模型，在精度持平的情况下，模型性能显著优于 AlphaFold

来源: 百度文心, 国金证券研究所

HelixGEM-2 是业界首个考虑原子间多体交互、长程相互作用的模型，创新性地采用多轨机制对化合物的不同阶的多体集合进行长程建模，在量子化学属性预测和虚拟筛选双场景上达到领先效果。2021 年 11 月，百度文心与国际知名生物制药公司赛诺菲达成协议，赛诺菲将利用百度文心生物计算平台在生物计算领域相关技术，优化与加速 mRNA 疫苗和药物的设计研发工作。

2.5 文心行业大模型

文心行业大模型是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，建立的具有行业特色的 AI 基础设施。目前，文心行业大模型已成功与 11 家企业或部门完成合作，各自形成了具有行业特色、性能优良、有落地潜力的行业大模型，涵盖能源、金融、航天、制造、传媒、城市、社科、电影 8 个行业领域。

图表10: 文心行业大模型拥有丰富的合作案例

合作案例	行业	合作对象	应用场景
国网-百度·文心	能源电力	国家电网	<ul style="list-style-type: none"> 设备技术标准知识库、电力知识问答系统、电力文档智能分析等多个能源行业应用场景

合作案例	行业	合作对象	应用场景
浦发-百度·文心	金融	浦发银行	<ul style="list-style-type: none"> 智能对话、智能文档分析、智能双录等多个金融行业应用场景
航天-百度·文心	航天	中国航天与探月工程	<ul style="list-style-type: none"> 航天故障部位信息抽取、航天工程文档情报分类、航天技术文献检索与摘要生成等多个航天领域应用场景
人民网-百度·文心	传媒	人民网	<ul style="list-style-type: none"> 新闻内容分类、新闻情感分析、新闻摘要生成等多个媒体场景
冰城-百度·文心	城市管理	哈尔滨市政府	<ul style="list-style-type: none"> 市长热线工单分类、投诉工单信息抽取
电影频道-百度·文心	电影	中央电视台电影频道	<ul style="list-style-type: none"> 视频修复及清晰度提升
深燃-百度·文心	能源	深圳燃气	<ul style="list-style-type: none"> 燃气行业中的场站槽车安全作业、园区安全生产、营业厅着装规范、安全隐患预警等
吉利-百度·文心	汽车	吉利汽车	<ul style="list-style-type: none"> 智能客服系统、车载语音系统、领域知识库构建等汽车行业应用场景
TCL-百度·文心	制造	TCL	<ul style="list-style-type: none"> 电子制造行业各种不同产线工艺场景的缺陷检测
辞海-百度·文心	社科	上海辞书出版社	<ul style="list-style-type: none"> 知识挖掘、词条管理、版权保护等人文社科行业应用场景

来源：百度文心，国金证券研究

通过与各行各业携手构建文心行业大模型，百度文心积累了旗下文心 NLP 大模型、文心 CV 大模型、文心跨模态大模型在多个行业、多种应用场景中落地的相关经验，形成了较为成熟的 AI 赋能模式。

图表11：文心行业大模型融合知识增强技术



来源：百度文心，国金证券研究所

随着百度新一代生成式对话机器人“文心一言”开放测试，AI 大模型落地进程将有望大大加速。目前已有 650 余家知名企业宣布接入文心一言，涵盖企业服务、金融 IT、汽车、传媒、互联网、教育、电器、金融等十余个领域。此外，百度自身主流业务也将接入文心一言，包括百度搜索，以及基于百度智能云的智能语音助手小度、智能驾驶 Apollo 等。3 月 16 日起，“文心一言”首批用户可通过邀请测试码接入体验，同时百度智能云将面向企业客户开放“文心一言”API 调用接口，有望进一步丰富百度 AI 生态，赋能更多行业伙伴。

图表12：已有 650 余家公司接入百度“文心一言”

领域	百度“文心一言”合作公司
企业服务	<ul style="list-style-type: none"> 致远互联、金蝶国际、明源云、用友网络、福昕软件、汉得信息、慧博云通、东软集团、远光软件、彩讯股份、北信源、竞业达、飞天云动、悦商集团、力龙信息、爱点击、神思电子、荣联科技、金财互联、华宇软件、利欧股份等
金融 IT	<ul style="list-style-type: none"> 恒生电子、宇信科技、凌志软件、京北方、软通动力、博彦科技、银之杰、科蓝软件、信雅达、新致软件、中科金财、新晨科技、艾融软件等
汽车	<ul style="list-style-type: none"> 吉利汽车、东风汽车、长城汽车、长安汽车、海马汽车、零跑汽车、红旗汽车等
传媒	<ul style="list-style-type: none"> 风语筑、太平洋网络、超媒体控股、财讯传媒、华策影视、蓝色光标、天娱数科、视觉中国、36 氪、浙文互联、汽车之家、中原网、金融界、值得买、百姓网、新浪新闻、迪生数字、凡拓数创、返利科技、博瑞传播等
互联网	<ul style="list-style-type: none"> 爱奇艺、前程无忧、智联招聘、掌阅科技、巨人网络、携程、Soul 等
教育	<ul style="list-style-type: none"> 博瑞传播、高顿教育、高途、学大教育、昂立教育、达内教育等
电器	<ul style="list-style-type: none"> 联想集团、四川长虹、TCL、美的集团、国光电器、海信视像、创维数字等
金融	<ul style="list-style-type: none"> 邮储银行、兴业银行、中信银行、江苏银行、广发证券、长江证券、国海证券、泰康保险、重庆银行等
算力	<ul style="list-style-type: none"> 浪潮信息、中科曙光、紫光股份、光环新网等
其他	<ul style="list-style-type: none"> 航天宏图、绿地控股、孩子王等

来源：iFind，各公司公众号，国金证券研究所

3. 文心平台提供一站式开发平台与工具

工具与平台层是百度文心平台生态构建中的重要一环，百度文心以 AI 落地为导向，为不同领域中的各级用户提供包括 AI 开发平台、AI 开发工具、大模型 API 在内的一站式 AI 开发服务。

3.1 AI 开发平台：零门槛平台 EasyDL 及全功能平台 BML

AI 应用开发者和 AI 专业开发者是 AI 落地环节中的两个典型角色：前者对 AI 底层算法和编程知识了解较为有限，希望能以高效、简单的方式应用 AI 技术为所属行业赋能；后者则专注于底层算法等 AI 专业领域，要求开发平台的功能强大且完善。为满足各级用户的使用需求，百度文心采用客户分级设计，向产业链上的不同客户提供功能各有侧重的 AI 开发平台，包括零门槛 AI 开发平台 EasyDL 以及全功能 AI 开发平台 BML。

- 零门槛 AI 开发平台 EasyDL 专为 AI 应用开发者打造，用户可以在不了解技术细节的前提下，以零代码、零门槛的方式快速实现 AI 模型的开发与部署。

图表13: EasyDL 提供一站式零代码 AI 模型开发



来源：百度文心，国金证券研究所

EasyDL 为用户提供了自然语言处理、图像、跨模态三个领域下的多种任务的一站式模型定制服务，并提供了工业制造、智慧城市、电子商务等多个真实场景下的 AI 开发范例，用户在进行 AI 模型开发时，仅需准备好场景数据集并依照提示操作，即可迅速完成模型的训练与部署。

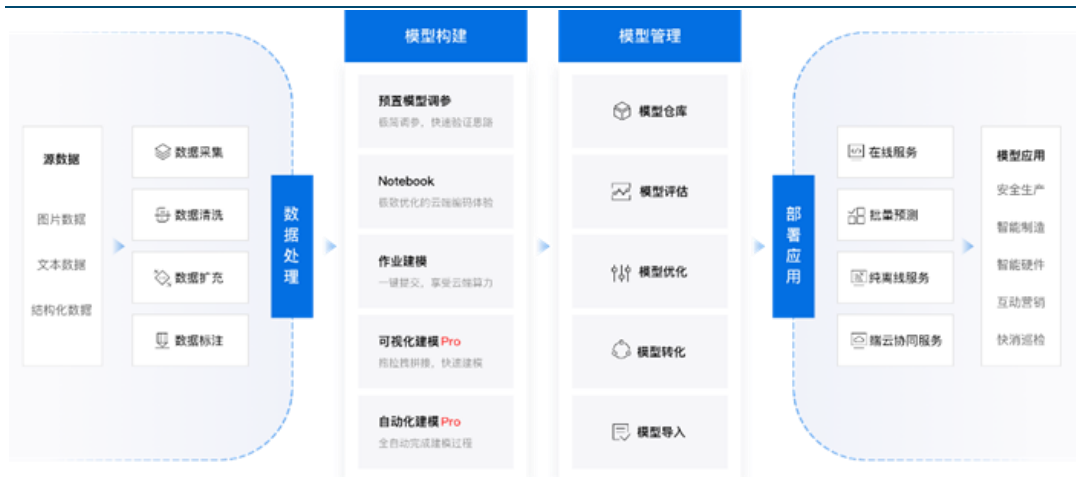
图表14: EasyDL 可支持多场景任务

领域	任务类型	应用场景
自然语言处理	文本创作	广告创意、文学创作
	文本分类	新闻推荐、评论分类
	情感分析	舆情分析
	文本实体抽取	金融研报信息识别、法律案件抽取
	文本实体关系抽取	行业知识图谱构建、行业知识问答
	评论观点抽取	辅助消费决策、舆情分析
	文本相似度计算	推荐系统
图像	图像分类	图片内容检索、制造业分拣或之间
	物体检测	医疗诊断、工业检测
	图像分割	医学图像分析、卫星图像分析
跨模态	图文匹配	图文素材推荐

来源：百度文心，国金证券研究所

- BML 全功能 AI 开发平台主要面向 AI 专业开发者打造，为 AI 专业开发者提供了从数据处理、模型训练、模型管理、模型部署与应用的生命周期管理服务，帮助用户更快、更好地构建、训练和部署 AI 模型。通过使用 BML 平台，开发者无需再关注底层资源的管理和维护，能够大大开发效率，加速 AI 赋能千行万业。

图表15: BML 全功能 AI 开发平台架构概览



来源：百度文心，国金证券研究所

EasyDL 和 BML 两个平台目前在训练阶段仅支持使用由百度文心提供的云端算力。目前，百度文心提供的公有云定价较高，以 GPU 提供的算力性能为标准统一计算，百度文心公有云的定价约为 Microsoft Cloud 的 6.3 倍。

图表16: EasyDL 与 BML 提供的模型部署方式

算力集群配置	价格
GPU: NVIDIA Tesla P4 8G CPU: 12 核 内存: 40G	13.4 元/小时
GPU: NVIDIA Tesla P4 24G CPU: 12 核 内存: 40G	17.3 元/小时
GPU: NVIDIA Tesla V100 16G CPU: 12 核 内存: 56G	21.6 元/小时
GPU: NVIDIA Tesla V100 32G CPU: 8 核 内存: 32G	24 元/小时

来源：百度文心，国金证券研究所

EasyDL 和 BML 两个平台在模型部署阶段灵活性较强，用户可以根据自身需要，灵活选择模型的部署方式，以满足不同开发者对价格、数据安全性、数据私密性等因素的个性化需求。

图表17: EasyDL 与 BML 提供的模型部署方式

部署方式	特点
公有云部署	<ul style="list-style-type: none"> 模型部署于飞桨公有云平台，按需按时获取算力资源，性价比高，适合中小型企业
私有化离线部署	<ul style="list-style-type: none"> 模型以私有化软件方式交付，客户可将模型部署于自有服务器或终端设备中，保障企业核心数据的私密性与安全性，能依据客户需要实现高级定制功能，适合大型企业

部署方式	特点
软硬一体部署	<ul style="list-style-type: none"> 模型部署至百度自研的软硬一体人工智能应用平台飞桨一体机中，飞桨一体机适配主流国产 AI 芯片与国产操作系统，支持自主可控，适合大型企业

来源：百度文心，国金证券研究所

3.2 AI 开发工具套件

AI 开发工具套件是百度文心生态构建中的基础环节，是百度文心旗下 AI 开发平台灵活易用、性能强大的基础，完善的 AI 开发工具生态将显著减少开发过程中“重复造轮子”的行为，加速 AI 开发与落地进程。

ERNIEKit 是百度文心专门面向自然语言处理领域研发的大模型开发工具，主要面向 NLP 工程师，提供全流程语言大模型开发与部署工具集，以全方位发挥语言大模型效能。

PaddleFleetX 覆盖大模型环境部署、数据处理、预训练、微调、模型压缩、推理部署全流程，并支持语言、视觉、多模态等多个领域的前沿 AI 算法。

ERNIEKit 和 PaddleFleetX 两种 AI 开发工具套件的优势在于以下几点：

- 简单易用，性能优异：用户能够以简单的方式利用 AI 开发工具套件的分布式运算能力，从而为模型训练和模型推理提供有力的算力支持。
- 模型训练、压缩、推理一体化：AI 开发工具套件支持模型训练、压缩、推理一体化等模型小型化策略，能够实现 AI 终端设备上的轻量化部署，大大降低部分场景下的 AI 落地成本。
- 多硬件支持，自主可控：AI 开发工具套件不仅支持 NVIDIA GPU 等主流的 AI 芯片，还支持多款主流国产 AI 芯片。

3.3 大模型 API

除 AI 开发平台与 AI 开发工具外，百度文心还通过大模型 API 向用户提供开发服务。目前，百度文心各模型 API 尚处于起步阶段，均免费向 AI 开发者提供。

图表 18：百度文心提供大模型 API

模型名称	应用领域	API 开放模式
ERNIE-ViLG	AI 画图	收费调用，价格约为 0.15 元/张
PLATO	AI 聊天机器人	限量开放内测
ERNIE 3.0 Zeus	文本理解与创作	免费申请调用

来源：百度文心，国金证券研究所

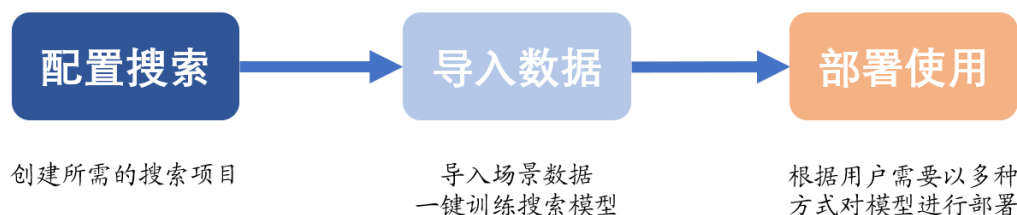
4. 文心平台提供搜索、创作等产品

产品与社区层建立于先进的百度文心大模型之上，是百度文心生态中的顶层，其旨在对外展示生成式模型的强大性能，并以此吸引爱好者投入创作，构建成熟的 AIGC 社区，助力 AI 模型迭代升级，并在实践中探索未来 AI 的商业化模式。

4.1 产业级搜索系统—文心百中

文心百中是百度文心基于旗下语言大模型开发的产业级搜索系统，基于文心 ERNIE 模型构建，目前已对外开放测试。相比传统搜索引擎，文心百中架构简单，且支持对多种类型的异构数据进行搜索。可以满足开发者和企业用户的多种垂直搜索、企业搜索需求。文心百中具有极简的逻辑架构以及强大的语义理解能力，可以极大降低搭建垂直搜索带来的数据成本与人力成本。

图表19: 文心百中使用流程简单



来源: 百度文心, 国金证券研究所

通过使用文心百中, 用户可以通过输入行业、企业知识库的方式以极低的成本构建所属领域的搜索系统, 从而实现企业内搜、知识搜索、知识推荐等搜索场景的相关任务。

4.2 AI 艺术和创意辅助平台—文心一格

文心一格是百度文心基于旗下文图生成跨模态大模型搭建的 AI 艺术和创意辅助平台, 目前已向用户开放, 用户可以通过文心一格体验 AI 作画, 并将优秀画作于平台中公开、分享创作经验, 为其他用户提供创意与创作经验, 由此形成平台的正向循环。

在版权方面, 文心一格允许用户将 AI 生成所得图片用于商业用途。文心一格旨在构建成熟的 AI 艺术和创意辅助平台, 为画师、设计师等视觉内容创作者提供灵感、创意和创作辅助, 为媒体、作家等文字创作者提供海量的高质量图片。文心一格目前采用收费调用模式, 根据用户需求不同提供了多种生成模式。

图表20: 文心一格收费价格 0.2-0.8 元/张

服务名称	服务内容	价格
普通 AI 作画	自定义图片描述、风格	0.2 元/张
高级 AI 作画	自定义图片描述、风格、尺寸, 多个可供选择的 AI 画师, 可上传参照图	0.2-0.8 元/张
AI 画作二次编辑	通过描述词实现图像内容删减	0.2 元/张

来源: 百度文心, 国金证券研究所

4.3 大模型创意社区—晒谷社区

晒谷社区是百度文心基于旗下多领域文心大模型构建的创意社区, 旨在探索大模型的在各个领域与场景中的创意应用模式, 晒谷社区采用众创模式, 鼓励个人开发者基于文心大模型 API 针对特定场景训练 AI 模型, 训练得到的优质模型将由百度文心收录并向全体用户开放使用, 晒谷社区有望助力 AI 大模型的应用探索。

图表21: 晒谷社区部分创意应用

创意应用	应用场景
文本处理系统	<ul style="list-style-type: none"> 面向日常办公场景, 基于 PaddleOCR 中 OCR 技术, 可将拍摄到的图片进行文字提取的同时, 借助 ERNIE 3.0 Zeus 对 OCR 识别的文字进行二次修正, 避免出现因 OCR 而产生的文档错误。
听故事画图	<ul style="list-style-type: none"> 面向幼教场景, 利用语音转文字技术和 ERNIE 3.0 Zeus 提取故事摘要, 并借助 ERNIE-VLG 实现 AI 作画。
装修设计	<ul style="list-style-type: none"> 面向室内设计场景, 可选择多种装修风格一键生成装修效果图, 可辅助室内设计师与装修客户沟通。

来源: 百度文心, 国金证券研究所

5. 国内其他生成式语言模型

我们梳理了百度文心外，其余国内厂商在生成式语言模型领域的最新研究成果。文心 ERNIE 3.0 Zeus 大模型在参数规模、技术创新、应用落地方面均在国内同业中处于领先水平。

图表22：国内主流生成式语言模型对比

厂商	百度	阿里巴巴	华为	清华大学
模型名称	ERNIE 3.0 Zeus	PLUG	盘古	ChatGLM
发布日期	2022 年 5 月	2021 年 4 月	2021 年 4 月	2023 年 3 月
参数规模	千亿级	240 亿	1100 亿	1300 亿
模型结构	Transformer Encoder-Decoder	Transformer Encoder-Decoder	Transformer Encoder-Decoder	Transformer Encoder-Decoder
训练数据	海量文本数据与知识图谱	1TB 以上的高质量中文文本训练数据，涵盖新闻、小说、诗歌、问答等广泛类型及领域	40TB 的行业文本数据和 400 万小时的行业语音数据	海量中、英文本信息
特点	创新型提出知识增强方法，参数规模处于世界领先水平	基于纯中文文本训练	融合行业知识，基于大量行业语音和文本数据训练，	采用了人类反馈强化学习方法

来源：百度文心，阿里巴巴，腾讯，华为，清华大学，国金证券研究所

6. 投资建议

相较于 ChatGPT 等海外 AI 模型存在连接限制，百度文心平台可以为国内用户提供便捷的技术入口。“文心一言”的发布将极大加速国内生成式 AI 模型的落地应用节奏，有望帮助生态伙伴实现用户体验提升和生产降本增效。建议关注接入“文心一言”生态的公司，如恒生电子、用友网络、汉得信息、凌志软件、宇信科技等。

7. 风险提示

■ 海外基础软硬件使用受限

若因国际关系等原因，高算力 GPU 等基础硬件或计算框架等基础软件使用受限，可能会对国内人工智能算法应用产生影响。

■ 应用落地不及预期

若相关应用公司不能找到人工智能算法较好的商业应用落地场景，或相关场景客户没有较强的付费意愿，可能算法应用落地会不及预期。

■ 行业竞争加剧风险

若相关企业加快技术迭代和应用布局，整体行业竞争程度加剧，将会对行业内已有企业的业绩增长产生威胁。

行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。

特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

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、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-60753903	电话：010-85950438	电话：0755-83831378
传真：021-61038200	邮箱：researchbj@gjzq.com.cn	传真：0755-83830558
邮箱：researchsh@gjzq.com.cn	邮编：100005	邮箱：researchsz@gjzq.com.cn
邮编：201204	地址：北京市东城区建国门内大街 26 号	邮编：518000
地址：上海浦东新区芳甸路 1088 号	新闻大厦 8 层南侧	地址：中国深圳市福田区中心四路 1-1 号
紫竹国际大厦 7 楼		嘉里建设广场 T3-2402