

2023 中国人工智能大模型企业综合 竞争力 50 强研究报告



赛迪工业和信息化研究院（集团）四川有限公司

2023 年 12 月

目 录

第一章 现状分析：人工智能大模型发展现状	3
1.1 参数规模突破，模型能力不断提高	3
1.2 应用边界拓展，覆盖领域持续扩大	3
1.3 产业创新驱动，市场规模稳步增长	4
1.4 产业结构优化，配套体系日趋完善	4
第二章 综合榜单：人工智能大模型企业 50 强评价	6
2.1 指标体系	6
2.2 评价方法	7
2.3 50 强榜单	9
2.4 竞争力分析	11
第三章 明星案例：大模型应用典型案例分析	17
3.1 工业领域大模型应用案例	17
3.2 金融领域大模型应用案例	18
3.3 交通领域大模型应用案例	20
3.4 医疗领域大模型应用案例	21
3.5 教育领域大模型应用案例	23
第四章 智领未来：人工智能大模型技术发展趋势	26
4.1 多模态大模型成为趋势	26
4.2 知识图谱与大模型融合	27
4.3 强化学习与大模型融合	27
4.4 生成对抗网络（GAN）与大模型融合	27

4.5 模型压缩与硬件加速技术结合	28
第五章 建言献策：人工智能大模型高质量发展建议	29
5.1 对政府.....	29
5.2 对企业.....	30
5.3 对投资机构	30

第一章 现状分析：人工智能大模型发展现状

当前，全球科技革命和产业变革加速推进，人工智能已成为国家间竞争的关键领域。人工智能作为战略性新兴产业，是推动科技跨越发展、产业优化升级、生产力整体跃升的重要战略性资源。大模型作为人工智能的核心组成部分，以其强大的学习能力和巨大的应用潜力，正在重塑千行百业。

1.1 参数规模突破，模型能力不断提高

大模型通常基于深度学习算法，通过大量的数据训练，使得模型能够更好地学习和理解数据，从而胜任各种任务。近年来，人工智能大模型的参数规模不断突破，带来了模型能力的不断提升。例如，GPT-1 到 GPT-3，模型的参数规模从 1.1 亿增长到 1750 亿个，GPT4 模型参数规模据悉已达到万亿级。随着参数规模的增加，大模型可以捕捉到更多特征和模式，处理更多数据和更复杂的网络结构，显著提高处理语言、图像和其他复杂任务的能力。

1.2 应用边界拓展，覆盖领域持续扩大

随着大模型的发展，其应用领域也在持续扩展和深化。最初，大型人工智能模型主要应用于语言处理任务，如文本翻译、情感分析和自然语言理解。目前，大模型的应用范围已经延伸到医疗诊断、金融风险评估、智能制造、轨道交通等多个行业，在提高生产效率、优化用户体验、辅助决策等方面展示出巨大潜力。例如，在医疗领域，大模型可以通过

分析医学影像和患者数据，提供更准确的疾病诊断和治疗建议。在金融领域，可被用于评估信贷风险和市场趋势，帮助企业和个人做出更明智的投资决策。

1.3 产业创新驱动，市场规模稳步增长

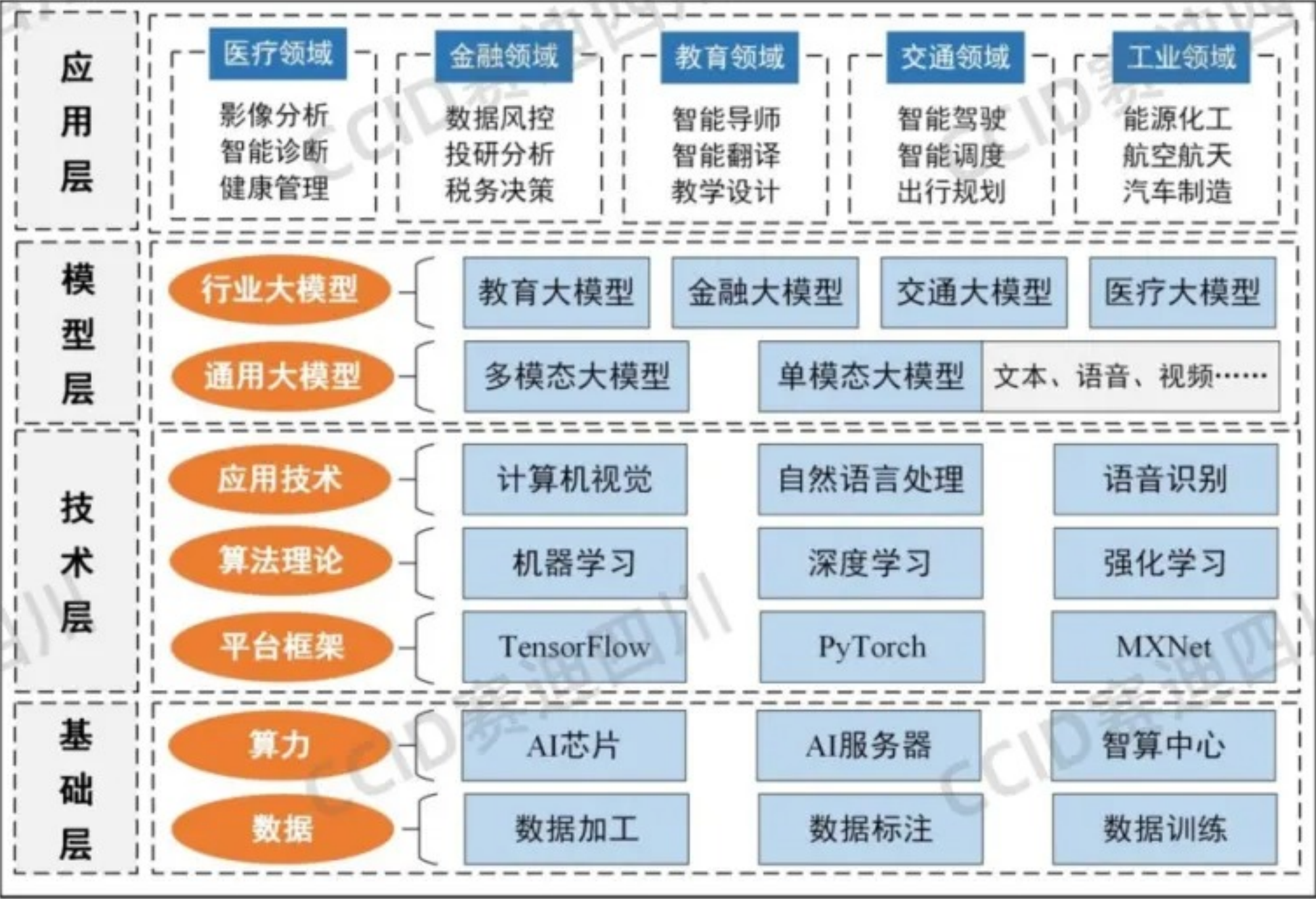
伴随着人工智能技术的不断成熟和应用领域的持续拓展，越来越多的企业开始采用人工智能技术来优化运营，提升自身竞争力。工信部最新统计数据显示，我国人工智能核心产业规模已达 5000 亿元，企业数量超过 4400 家。全球人工智能市场规模预计将在未来几年内持续高速增长。人工智能产业的创新发展将加速推动人工智能大模型的商业化和产业化进程。同时，政府也在积极推动人工智能大模型产业的发展，出台了一系列政策和规划，为人工智能大模型的发展提供政策支持和保障。

1.4 产业结构优化，配套体系日趋完善

人工智能大模型的产业链结构不断优化，涵盖了从数据处理、算力平台搭建、算法研发到大模型应用服务等各个环节，涌现出一批从事大模型研发应用及提供相关配套服务的企业和机构。例如，英伟达和华为正在研发更为高效的处理器，打造强有力的算力底座来支持大型模型的运算需求。亚马逊 AWS 和阿里云提供了强大的基础设施来部署和运行人工智能模型。产业链结构的不断优化和配套体系的日趋完善

为人工智能大模型的发展提供了坚实的基础，并将促进技术的快速迭代和优化。

图 1 人工智能大模型产业链全景图



资料来源：赛迪四川，2023.12

第二章 综合榜单：人工智能大模型企业 50 强评价

2.1 指标体系

基于对人工智能大模型的研究，首次提出对人工智能大模型企业综合竞争力评价指标体系，综合考虑企业经济实力、技术水平、技术创新、合作生态、行业影响力 5 个一级指标，市场营收、融资与风险、开放平台等 19 个二级指标、形成企业竞争力综合评价指标体系，推动人工智能大模型产业高质量发展。

表 1 人工智能大模型企业 50 强综合竞争力评价指标体系

一级指标	二级指标	指标说明
经济实力	市场营收	开发大模型的经济支撑
	融资与风险	持续投入大模型相关业务的经济潜力
技术水平	人才梯队	研发创新团队的支持力量
	模型规模	大模型参数
	开放平台	开发者或企业接入大模型服务的资源建设情况
	开发工具	大模型工程化能力
	数据支撑	大模型训练的知识语料及特定行业的专用语料
	自有算力	涉及自主研发 AI 芯片、AI 计算设备或采买 AI 算力设

		备
技术创新	论文数量	科研成果呈现
	专利数量	研发成果呈现
	研发投入	对大模型相关业务的重视程度、资金支持程度
合作生态	生态合作影响力	生态建设、合作开放程度
	生态体系建设	
	科研合作开放度	
	运维保障能力	服务支持、大模型稳定性支持能力
	开源社区建设	技术交流推动技术发展与创新
行业影响力	行业标准贡献度	行业标准带头/参与制定情况
	垂直行业覆盖度	大模型落地所涉行业的类别
	行业认可度	大模型在各行业应用的深度、广度

资料来源：赛迪四川，2023.12

2.2 评价方法

在评价工作开展过程中，按照科学合理的方式对评价指标体系进行拆解、分析，具体包含以下步骤：数据收集整理、

数据清洗、确立并赋值指标权重、数据处理、建立计算模型、计算评分、结果分析等。以下重点展开说明确立并赋值指标权重、数据处理、建立计算模型、计算评分四大步骤的过程。

2.2.1 确立并赋值指标权重

根据综合评价指标体系中的一级指标、二级指标，运用德尔菲（Delphi）法专家赋分并结合层次分析（AHP）方法，得到每一项一级指标、二级指标的相对权重。

根据全国所有人工智能大模型企业的实际发展情况，得到每一项一级指标、二级指标的具体数值，并对数据进行无量纲化处理，得到每一指标的实际赋值情况。

2.2.2 数据处理

为消除各项指标间单位不一、数值差异过大的问题，需对数据进行无量纲化处理。根据不同的指标数据类型，选择不同的无量纲化处理方式。记各项评估指标的原始数值为 X_{ei} （ e 为指标对象， i 为指标编号），无量纲化后的值为 Z_{ei} ，指标 i 的计算基值为 $\bar{X}_i$ 。

基值的计算：各项指标体系的基值为人工智能大模型企业数据的平均值，计算操作如下：

$$\bar{X}_i = \frac{\sum_{i=1}^n X_{ei}}{n}$$

数值类指标的处理：为避免各项数值类指标间数值范围差异过大的问题，采用取对数的方法对此类指标进行无量纲化处理，计算操作如下：

$$Z_{ei} = \left(\ln \left(1 + \frac{X_{ei}}{X_i} \right) \right)$$

指数类指标的处理：此类数据各数值间差异较小，直接进行数据归一化处理即可，计算操作如下：

$$Z_{ei} = \frac{X_{ei}}{X_i}$$

2.2.3 建立计算模型

根据专家已打分（各级指标体系权重总分为 100）赋值的指标权重，确立最终的指标权重的平均值为 λ_i 。每个细分指标的各级指标指数的计算均采用加权平均法，即：

$$Z_i = \frac{\sum \lambda_i Z_{ei}}{\sum \lambda_i}$$

2.2.4 计算评分

根据指标的权重和具体赋值情况，运用计算模型得出最终计算结果，反映人工智能大模型企业发展情况。

2.3 50 强榜单

根据人工智能大模型竞争力综合评价模型，对全国人工智能大模型企业进行评价分析，具体结果如下。

表 2 人工智能大模型企业 50 强综合竞争力榜单

排名	企业名称	大模型名称	注册地址
1	百度	文心一言	北京
2	华为	盘古	深圳
3	阿里巴巴	通义千问	杭州
4	商汤科技	日日新	北京
5	科大讯飞	星火	合肥
6	腾讯	混元	深圳
7	第四范式	式说	北京
8	浪潮信息	源	济南
9	拓尔思	拓天	北京
10	中国电信	TeleChat, 启明, 星辰	北京
11	星环科技	无涯, 求索	上海
12	抖音	豆包	北京
13	云从科技	从容	广州
14	新华三	百业灵犀	杭州
15	智谱华章	智谱清言	北京
16	网易	子曰	杭州
17	大华股份	星汉	杭州
18	恒生电子	LightGPT	杭州
19	中科闻歌	雅意	北京
20	思必驰	DFM-2	苏州
21	达观数据	曹植	上海
22	云知声	山海	北京
23	百川智能	百川	北京
24	云天励飞	天书	深圳
25	澜舟科技	孟子	北京
26	幻方 AI	DeepSeek Coder	杭州
27	京东	言犀	北京

28	中科创达	魔方 Rubik	北京
29	佳都科技	佳都知行	广州
30	中国电子云	星智	武汉
31	小米	MiLM-6B	北京
32	航天宏图	天权	北京
33	虎博科技	Tigerbot	上海
34	联汇科技	欧姆	杭州
35	安恒信息	恒脑	杭州
36	九章云极	DataCanvas	北京
37	面壁智能	Luca	北京
38	智臻智能	华藏	上海
39	超对称	乾元	北京
40	惟远智能	千机百智	深圳
41	智慧眼	砭石	长沙
42	元年科技	方舟 GPT	北京
43	理想汽车	MindGPT	北京
44	汉王科技	天地	北京
45	实在智能	塔斯	湖州
46	众合科技	UniChat	杭州
47	开普云	开悟	东莞
48	云问科技	云中问道	南京
49	慧安股份	蜂巢知元	北京
50	清博智能	先问	北京

资料来源：赛迪四川，2023.12

2.4 竞争力分析

2.4.1 行业应用分布

2023 中国人工智能大模型 50 强行业应用共涉及 13 个领域，显著集中于金融行业，随后为工业、政务和交通等行业。

从大模型的应用领域来看，位居前三的行业具有以下几点共性。**一是**行业数据丰富且数据结构化程度较高，具有易于收集和加工处理的特性；**二是**行业体系成熟度高，具有将数据转换为行业知识的明确路径；**三是**行业本身市场前景广阔，具有巨大发展潜力。

图 2 2023 中国人工智能大模型 50 强企业行业应用分布



资料来源：赛迪四川，2023.12

2.4.2 区域分布格局

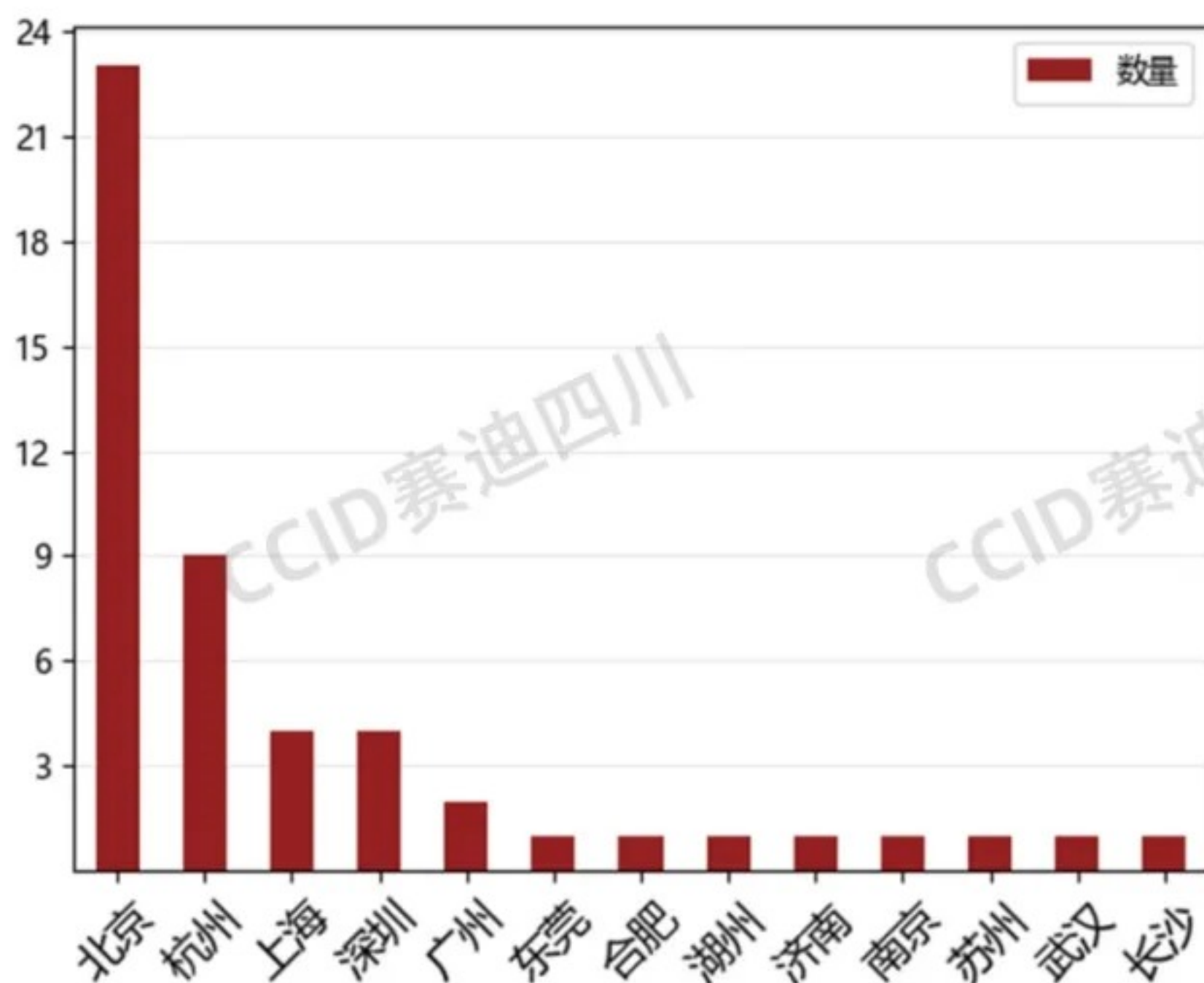
在区域分布上，2023 中国人工智能大模型 50 强企业在京津冀和长三角区域的集聚效应明显。

城市分布方面，北京占有明显领先地位，杭州、上海、深圳、广州位列第二梯队，其余城市呈现长尾分布。形成此种格局的原因主要有以下几方面：

一是该区域政府前瞻布局战略意识高，人工智能产业发

展决心强，出台相关产业优惠政策多；**二是**该区域拥有较为完善的人工智能上下游产业链，信息技术和高端制造领域具有较强竞争力，为人工智能的发展提供了坚实基础和需求牵引；**三是**该区域具备新兴产业人才优势，拥有众多国内顶尖的高校和科研机构，为人工智能产业发展提供人才储备。

图 3 2023 中国人工智能大模型 50 强企业城市分布



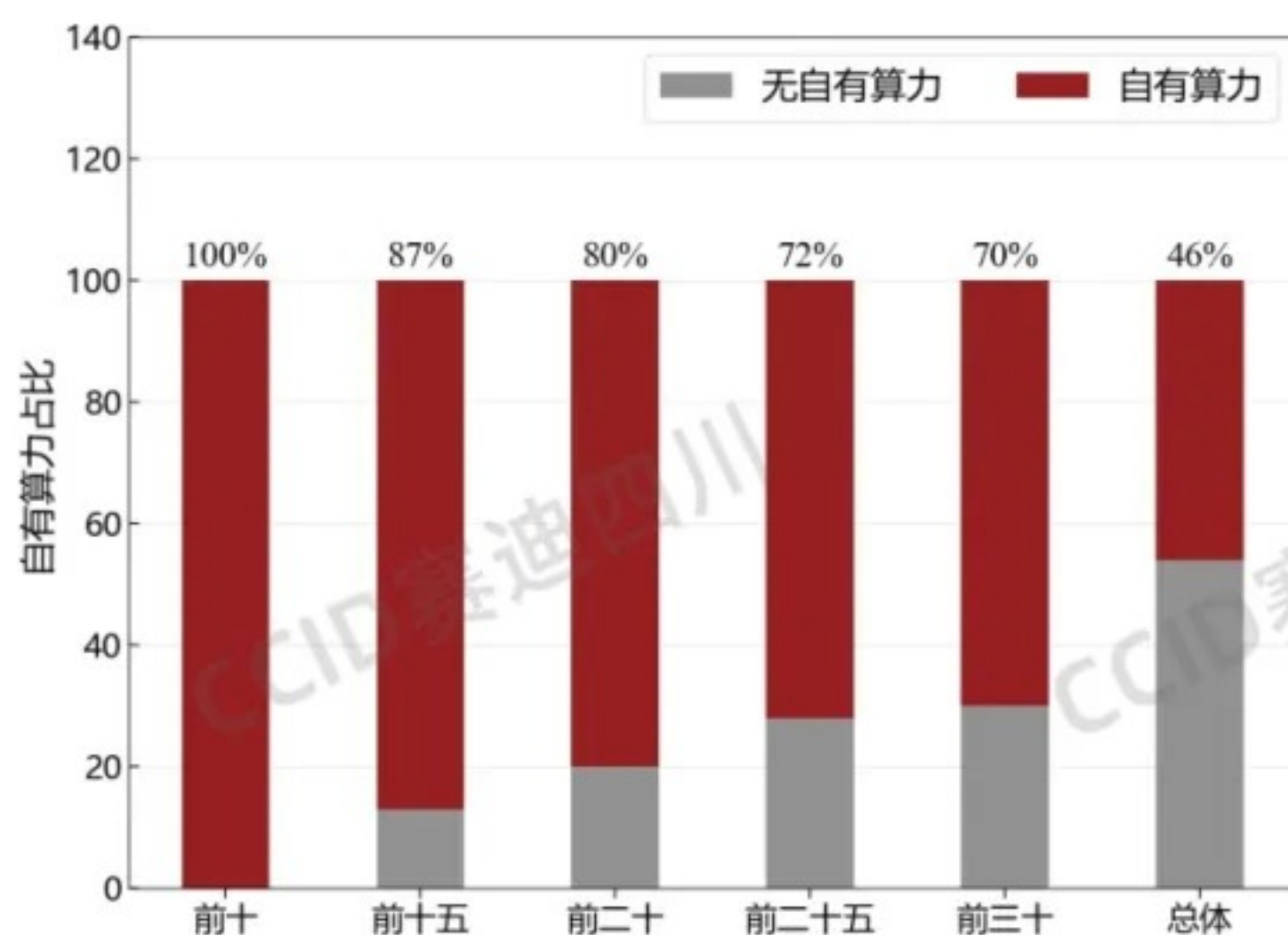
资料来源：赛迪四川，2023.12

2.4.3 自有算力布局

2023 中国人工智能大模型 TOP10 企业中，自有算力的企业占比达 100%（本报告中所指自有算力包含自主研发芯片、自有计算硬件或采买算力设备等）；TOP15 企业中，自有算力占比为 87%；50 强企业整体自有算力占比为 46%。从整体情况来看，不足半数的企业拥有自主掌握的算力资源，表明我国在整体基础实力方面仍有待提高。从不同排名分阶的

自有算力占比来看，自有算力的企业具备更强的竞争力。强大的算力资源不仅可以加速模型训练，提升市场响应速度，还能支持更复杂、更深层次的模型训练，从而提高模型的预测精度和整体性能。

图 4 2023 中国人工智能大模型 50 强企业自有算力占比分布

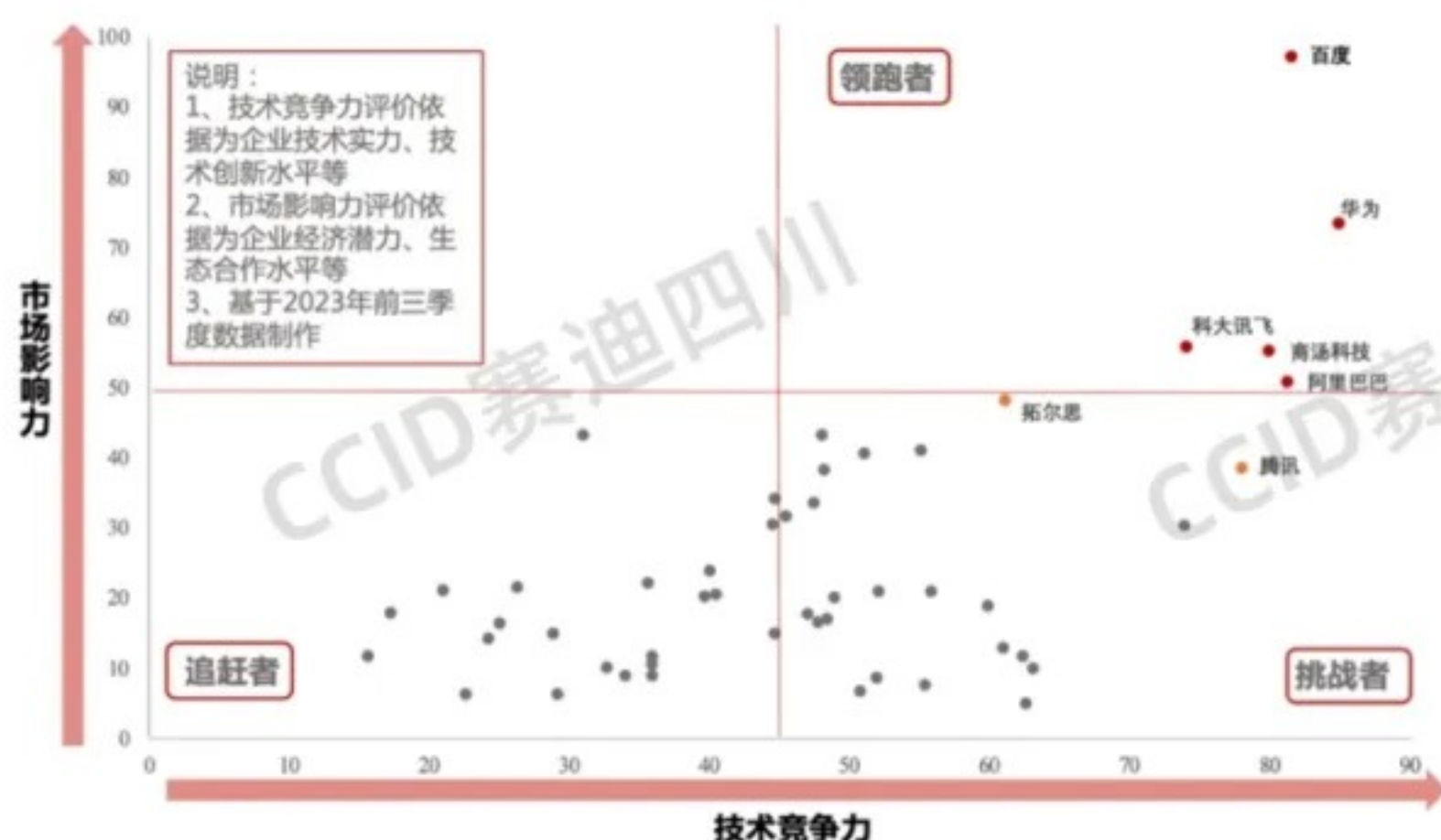


资料来源：赛迪四川，2023.12

2.4.4 技术市场关联度分析

整体来看，国内人工智能大模型竞争力排名前五十的企业可分为三类。第一类是领跑者，在国内市场上拥有绝对的优势；第二类是挑战者，即在技术竞争力或市场影响力方面具备一定的实力，在实际市场竞争中占据一定的位置；第三类是追赶者，他们与挑战者相比，技术实力并不显著，客户资源较为边缘或者主攻较为单一的细分行业。

图 5 人工智能大模型企业 50 强散点分布



资料来源：赛迪四川，2023.12

从散点图横纵向布局来看，技术竞争力方面，百度、华为、商汤科技、阿里巴巴、腾讯、科大讯飞处于明显的领先地位，其余企业与其存在较大差距。除技术领先企业外，大部分企业技术水平参差不齐，多集中于榜单中部位位置，头部、尾部分布较少。市场影响力方面，百度、华为、阿里巴巴等榜单头部企业远超其他处于榜单中部、尾部企业，中部、尾部企业间的市场影响力差异较小且整体市场影响力表现欠佳。

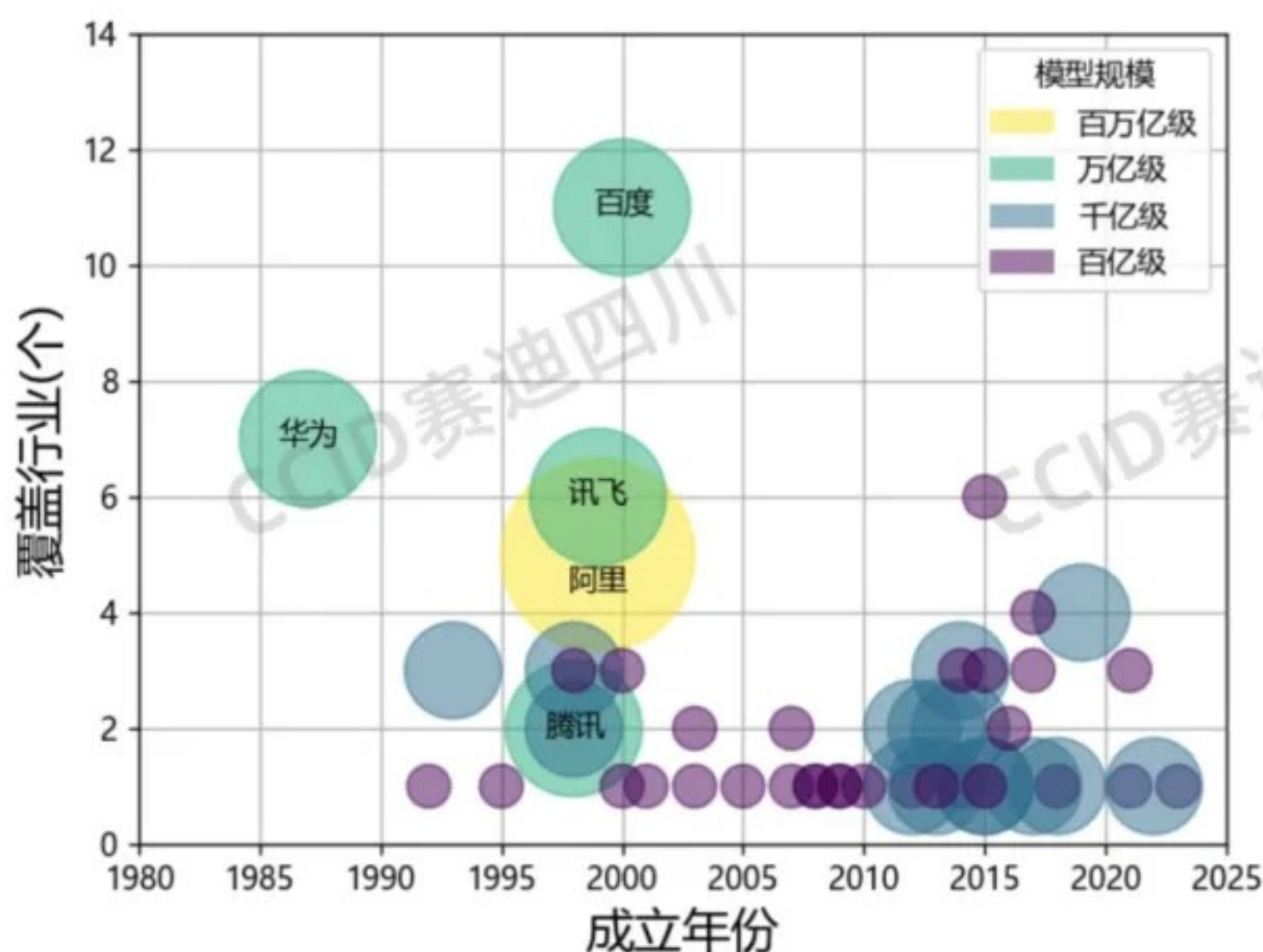
从散点图区域分布来看，领跑者为百度、华为、阿里巴巴、商汤科技、科大讯飞 5 家企业，位于榜单头部，遥遥领先于其他挑战者和追赶者。纵向来看，拓尔思、腾讯位于领跑者、挑战者临界区域，虽市场影响力稍弱但技术实力强劲，有望进军领跑者位置。横向来看，挑战者和追赶者之间的差距主要体现在技术竞争力方面，值得关注的是，十余家企业

位于追赶者和竞争者之间的临界点附近，表明企业间的技术竞争逐渐呈现白热化的局面。

2.4.5 综合竞争力分析

从企业成立年份来看，自 2012 年起，人工智能企业呈现出明显聚集趋势。2012 年作为深度学习的元年，在技术创新和政策推动的双重影响下，人工智能企业如雨后春笋般涌入市场。结合模型规模和成立年限来看，大模型的发展通常需要企业具有长期的技术沉淀，万亿级以上的模型都集中在成立 20 年以上的企业。从行业覆盖数量来看，百度遥遥领先，华为、讯飞、阿里紧随其后。通用大模型的行业落地与企业实力和生态能力紧密关联，尽管部分企业拥有万亿级的大模型，但由于入市较晚，目前并未位列第一梯队。

图 6 2023 中国人工智能大模型 50 强企业综合气泡图



资料来源：赛迪四川，2023.12

第三章 明星案例：大模型应用典型案例分析

3.1 工业领域大模型应用案例

3.1.1 文心大模型

文心大模型是百度智能云开发的一种强大的自然语言处理工具，是产业级知识增强大模型，因为它生于产业，服务于产业，以此成为各行各业的首选基座大模型之一。

百度智能云在化工、制造、能源以及汽车等工业领域逐渐开启了新的版图扩张。**在化工行业**，中国石化与百度签署战略合作框架协议，双方将在人工智能、互联网+新业态、数字化转型升级等领域深化合作。**在能源行业**，百度与其他大型央国企还有过深入合作。国家电网与百度联合发布知识增强的电力行业大模型，建设更加适配电力行业场景的 AI 基础设施，其模型基座也就是文心大模型，最终能够达到降低数据标注成本，提升细分场景模型应用的效果。**在航空航天行业**，中国探月、航天工程与百度开展合作，为文心大模型提供大量航天科研数据，使其从这些数据与知识中进行融合学习，助力航天工作效率提升。**在汽车工业领域**，中国华晨宝马与百度达成战略合作，共同探索 AI 技术与汽车制造业的融合创新。

3.1.2 魔方 Rubik

中科创达作为全球领先的智能操作系统产品和技术提供商，在智能汽车与智能硬件领域拥有丰富的经验。魔方

Rubik 作为其自主研发的大模型，与其主营业务深度结合应用。在智能硬件领域，中科创达将创达魔方大语言模型部署在 TurboX，在 2023 上海世界移动通信大会上展示了旗下首款集成魔方大语言模型的智能搬运机器人，在大模型的加持下，机器人实现了更自然的语言交互，可以更好地理解人类指令。在智能汽车制造方面，中科创达开发了以魔方 Rubik 大模型为基础的 Rubik GeniusCanvas 产品，实现了对智能汽车开发周期全过程的覆盖，包括对汽车的概念创作、3D 元素设计、交互与视觉、特效及场景制作、应用开发集成等方面，极大程度提高了汽车开发的效率。

3.2 金融领域大模型应用案例

3.2.1 Aisino 财税大模型

Aisino 是航天信息在税务领域所提出的一款财税大模型，致力于将税务服务智能化，实现税务服务的数字化转型。并且 Aisino 财税大模型可以应用于 AI 互动、税收预测、风险评估、税务决策等多个方面。

一是 AI 互动。如 AI 数字人助手、税务数字人交互终端、智能外呼助手、12366 智能热线电话咨询、人工即时通讯和音视频等，获得在线咨询办理、电话咨询办理、在线支付、在线查询、音视频连线等服务。二是税收预测。Aisino 财税大模型通过对历史数据和经济指标进行分析，可以自动挖掘数据中的模式和规律，提高预测准确度，为税务部门提供决

策依据。**三是风险评估。**Aisino 财税大模型能够分析大量数据和复杂的关联关系，比如，纳税人的历史纳税记录、财务数据、经营情况等，通过这些数据信息，能够识别出纳税人的异常行为和风险因素。**四是税务决策。**传统的税务决策往往基于经验判断和少量的数据，这使得决策过程容易受到主观因素的影响。而 Aisino 财税大模型则能够通过对大量数据和复杂关联关系的深度分析，为税务决策提供科学的决策支持。

3.2.2 LightGPT

LightGPT 是恒生电子自主研发的专为金融领域打造的大语言模型，该模型拥有专业的金融语料积累处理和高效稳定的大模型训练方式，使用了超 4000 亿 tokens 的金融领域数据（包括资讯、公告、研报、结构化数据等）和超过 400 亿 tokens 的语种强化数据（包括金融教材、金融百科、政府工作报告、法规条例等），并以之作为大模型的二次预训练语料，支持超过 80+金融专属任务指令微调，使 LightGPT 具备金融领域的准确理解能力，所以该模型在金融相关问题的理解和金融任务的处理上，比通用大模型更有优势。

一是基于 LightGPT 金融大模型所打造的智能助手“光子”，在实际应用中串联了“通用工具链+金融插件工具+金融数据+金融业务场景”的智能应用服务，可以为金融机构的投顾、客服、运营、合规、投研、交易等业务系统注入 AI 能力。

二是智能投研平台 WarrenQ 也是基于 LightGPT 金融大模型所打造的数智金融平台，该平台赋能“搜、读、算、写”投研全流程场景，旨在提供更加高效、智能和精准的投研服务。其衍生功能 WarrenQ-Chat 通过对话指令，轻松获得金融行情、资讯和数据，且每一句生成的对话均支持原文溯源，确保消息出处可追溯，还可以生成金融专业报表，为金融机构提供了更加便捷、快速、精准的投研服务。

3.3 交通领域大模型应用案例

3.3.1 佳都知行

佳都知行交通大模型是基于行业知识库、行业海量数据、轨交业务场景研发打造的超强逻辑推理、自然语言处理能力、反复训练应用到生产环境的“强人工智能”行业大模型。

该大模型在实际生活中运用于两个方向：**一是在轨道交通治理上**，佳都知行交通大模型能够学习轨道交通各类系统的特征、模式和规律，形成“领域知识库”和“领域工具库”，为城市智慧轨道交通各大业务系统赋能，如智能客服、智能运维、智慧节能、综合指挥、应急演练等。**二是在城市交通治理上**，佳都知行交通大模型在交管工作效率提升和业务问题解决上更加智能，能够对城市交通系统进行全面、深入的分析和预测，实时监测道路、公共交通、车道等各类交通设施的使用情况，精确评估交通流量、拥堵程度和出行需求，

为城市规划者、政策制定者和市民提供科学、准确的决策依据。

3.3.2 从容

从容大模型是一款基础大模型，能够根据给定的文本或语音数据进行学习和分析，并深度理解自然语言的含义和语义规则，具有高度智能化及自适应性，在 NLP 领域具有诸多优势。

在城市交通治理的实际应用中，从容大模型基于对城市交通行业知识进行持续训练和自学习，能够形成交通行业思维，在交管工作效率提升和业务问题解决上更加智能。更具体地说，基于从容大模型的应用工具，能够实现城市交通系统全面深入的分析和预测，对道路、公共交通以及车道等各类交通设施的使用情况进行实时监测，以此精确评估交通流量、拥堵程度和出行需求，为城市规划者、政策制定者和市民提供科学、准确的决策依据。

3.4 医疗领域大模型应用案例

3.4.1 砭石

近两年智慧医疗的概念在社会上不断被提起，砭石大模型应运而生。该大模型是世界医学领域首个支持多模态的大模型，它支持人脸视频输入、面舌诊输入、医学图像输入、病例文本输入、音频输入五种输入方式，实现智能问诊、辅助阅片、面诊舌诊、生理指标预测、睡眠数据监测等多样化

的任务。它的输出是结构化的文本，包括健康状态的简要描述、诊断描述和治疗建议等，可以方便医疗机构更好地理解分析，提高了医疗信息处理的效率和准确性。

在实际应用上，**一是以砭石医疗大模型为基底的智慧眼云慢病患者服务管理系统**，赋能于医保、医院、药房、科室和医护人员，将医疗健康服务延伸到院外，提供以患者为中心，贯穿院前管理、院中就诊、院后康复追踪的“医+诊+药+护”的连续性治疗服务闭环，向患者提供全流程、多方位的一站式全生命周期健康管理服务。**二是互联网医院**，一种基于人工智能+互联网技术的远程医疗服务平台，通过数字疗法，提供智能导诊、辅助诊断、智能用药等服务，贯穿预防、诊断、治疗和康复的核心环节，给患者提供分层、协同、联合、全程、连续的医疗数字化服务。

3.4.2 大医

大医是商汤科技发布的升级版医疗健康大模型，其功能覆盖智慧大健康、智慧患者服务、智慧临床以及数智建设四大领域，包括智能自诊、用药咨询、诊后随访管理、智慧病历、影像报告结构化等共 13 个细分医疗健康场景。**一是体检咨询助手**。面对日常健康咨询，大医提供疾病风险预测、检验检查分析、体检咨询、健康问答等健康管理服务，经过多轮对话后给出健康咨询建议。**二是陪诊助手**。面对患者的陪诊需求，大医可以更高效地实现智能导诊、预问诊等患者服

务功能。**三是医生助手。**面对医生的门诊，大医可以实时记录、整理、识别医患问诊对话内容，并将提取总结的病历信息上传至电子病历系统，解放医生双手，提高临床效率。**四是用药助手。**面对患者就诊之后，大医可为患者提供用药指导，并帮助患者建立随访档案及制定随访计划，提高医疗服务质量和患者获得感。

目前商汤科技已与郑州大学第一附属医院、上海交通大学医学院附属瑞金医院、上海交通大学医学院附属新华医院等机构开展合作，实现大医在具体医疗健康场景中的落地。除此之外商汤科技发布了以大医大模型为基底的“SenseCare®智慧医院”综合解决方案，覆盖智慧就医、智慧诊疗、智慧医院管理、智慧医学科研、智慧医疗云五大场景，赋能医院诊疗、科研、决策等各个层面，助力医院智慧化转型升级。

3.5 教育领域大模型应用案例

3.5.1 孟子

孟子是澜舟科技研发的大规模预训练语言模型，可处理多语言、多模态数据，同时支持多种文本理解和文本生成任务，能快速满足不同领域、不同应用场景的需求。孟子模型基于 Transformer 架构，仅包含十亿参数量，基于数百 G 级别涵盖互联网网页、社区、新闻、电子商务、金融等领域的高质量语料训练。相比于市面上已有的中文预训练模型，孟

子模型具有轻量化和易部署的特点，性能上超越同等规模甚至更大规模的模型。

目前澜舟科技已与多家企业合作，联合推出基于孟子GPT的相关产品。一是中文在线利用孟子预训练技术构建针对文学生产领域的专业模型，实现多属性可控文学生成提供续写、基于关键词生成、基于用户自定义模板生成、文学实体渲染等能力。二是新华智云将孟子大语言模型应用在文化数字化的全技术链路中，提供具备专业的历史文化知识的大模型服务及应用能力，在角色扮演、文化知识问答等实际业务场景进行验证。

3.5.2 子曰

相比于通用大模型，子曰大模型的定位从开始就是教育垂类大模型，主打场景优先，这也是国内企业推出的首个教育场景下类 ChatGPT 模型。该模型能够作为基座模型支持诸多下游任务，向所有下游场景提供语义理解、知识表达等基础能力。

在应用方面，覆盖翻译、作文、语法、口语、AIBox、文档问答共六种细分领域。比如，有专用于英语口语练习的虚拟人教练 Hi Echo，模拟的语境贴合实际情况，可以帮助用户更好地练习英语口语。此外，还有六大成果之一的“AI 作文指导”，不仅可以批改作文，还有指导写作的功能。针对学生在写作过程中面临的各种难题，如题目难以确定，写作主旨

不明等，该应用都能够给予指导，帮助学生逐渐掌握写作技巧。在批改环节中，AI 作文指导还会从表达、结构、内容深度、情感丰富度等维度全面提供改进建议。

第四章 智领未来：人工智能大模型技术发展趋势

大模型是人工智能行业发展的新范式、新引擎，是支撑产业数字化转型的重要基础设施。人工智能正逐渐成为数字经济的核心引擎，未来随着技术不断演进和升级迭代，大模型将与人工智能技术融合创新，实现更多产业端场景的落地。

4.1 多模态大模型成为趋势

多模态大模型主要包括三类：一是以语音、图像、文本为核心的多模态大模型，二是以视频为核心的多模态大模型，三是以图像、音频为核心的多模态大模型。未来，多模态大模型将会成为一种趋势，主要体现在以下几个方面：**一是数据模态的融合。**多模态大模型的前提是多种模态数据的融合，多模态大模型需要将不同模态的数据进行融合，以充分利用它们之间的信息。通过将不同模态的数据进行拼接、转换、融合等多种方式可实现数据模态的融合。**二是跨模态转换。**不同模态的数据具有不同的特征和表达方式，因此需要进行跨模态转换，以实现不同模态数据之间的转换和迁移。跨模态转换通常可以采用多种方式，例如使用自编码器进行跨模态生成、使用对抗生成网络进行跨模态转换等。**三是多任务学习。**多任务学习是指同时处理多个任务的学习方式，这些任务可以是不同的数据模态或者不同的应用场景。多任务学习可以使得模型具有更好的泛化性能和鲁棒性，同时也可以提高模型的效率和可用性。

4.2 知识图谱与大模型融合

知识图谱是一种语义网络，表示现实世界中各种实体、概念及其之间的关系，将其与大模型融合，可以有效地提高大模型的语义理解和推理能力。未来，知识图谱与大模型的融合将会成为一种趋势，主要体现在以下几个方面：**一是**知识图谱的嵌入。通过将知识图谱中的语义信息嵌入到大模型中，可以提高大模型的语义表达能力。**二是**联合学习。通过将知识图谱中的数据与大模型联合学习，可以提高大模型的泛化能力和推理能力。**三是**语义搜索。通过将知识图谱与大模型结合，可以实现更高效和准确的语义搜索。

4.3 强化学习与大模型融合

强化学习是一种通过试错学习的机器学习方法，可以训练出具有复杂行为能力的智能系统，将其与大模型融合，可以有效地提高大模型的自主学习能力。未来，强化学习与大模型的融合将会成为一种趋势，主要体现在以下几个方面：**一是**深度强化学习。通过将深度学习技术与强化学习技术结合，可以训练出更具复杂行为能力的智能系统。**二是**分布式强化学习。通过将多个强化学习智能体分布在不同设备上，可以实现更高效的分布式强化学习。**三是**可解释的强化学习。通过设计可解释的强化学习算法，可以提高大模型的透明度和可解释性。

4.4 生成对抗网络（GAN）与大模型融合

生成对抗网络是一种通过竞争游戏来生成新数据的技术，将其与大模型融合，可以有效地提高大模型的生成能力和创新能力。未来，GAN 与大模型的融合将会成为一种趋势，主要体现在以下几个方面：**一是**更大规模 GAN。随着计算资源和数据量的增加，GAN 的规模也将不断增大，以达到更高的生成质量和创新能力。**二是**多模态 GAN。通过将 GAN 与其他技术结合，可以扩展 GAN 的应用场景，如文本到图像的生成、视频生成等。**三是**可控 GAN。通过设计可控 GAN 的算法和技术，可以提高 GAN 生成结果的可控性和稳定性。

4.5 模型压缩与硬件加速技术结合

随着人工智能技术的广泛应用，模型压缩和硬件加速技术已经成为提高大模型性能和降低成本的重要手段。未来，模型压缩与硬件加速技术将会更加紧密地结合在一起，主要体现在以下几个方面：**一是**量化技术。通过将浮点数转换为整数进行计算，可以大大降低模型计算量和内存占用，提高模型的压缩比和计算效率。**二是**剪枝技术。通过去除模型中的冗余信息，可以大大降低模型复杂度和计算量，提高模型的压缩比和计算效率。**三是**硬件加速器。通过专门设计的硬件加速器来加速大模型的计算过程，可以大大提高模型的计算效率和性能表现。例如，GPU、FPGA 和 ASIC 等硬件加速器已经被广泛应用于人工智能计算领域。

第五章 建言献策：人工智能大模型高质量发展建议

5.1 对政府

加强顶层设计，健全政策保障体系。一是组织开展 AIGC 产业相关战略性研究，明确发展定位；二是构建产业图谱，探索大模型发展路径，分阶段制定战略规划和重点任务；三是整合本地资源，引导差异化发展，以制度创新释放发展红利。

匹配潜力要素，把控全局协同发展。一是打造科技创新优势，引导和支持企业跨区组建创新联合体，联合高校及科研院所打造协同创新平台，共同进行技术攻关和技术成果转化，推动产学研深度融合；二是打破数据壁垒，建立数据共享机制，协调推进数据基础制度建设；三是加强建设智算中心，提供算力资源服务，降低本地企业研发成本，推动产业集群化发展。

健全规范保障，营造安全发展环境。一是建立和完善人工智能安全政策法规，强化数据安全与隐私保护；二是搭建相关知识产权保护机制，解决生成式 AI 生成内容的版权归属与版权问题，保护其在商业领域的推广及使用；三是制定安全可信的准入标准，建立测评认证体系，避免商业恶性竞争。

5.2 对企业

重视产业布局战略，加快产品与新兴产业融合。企业应积极拥抱新兴技术，加速新兴产业与自身产品的结合，打造具有自身特色的产品，抢先获得未来市场的领先地位。研发上，建议企业因势利导选择合适细分赛道切入，灵活应对市场环境变化；小步快跑，不断迭代和改进模型，提升产品或服务的质量和性能，以满足客户需求并保持竞争力。

重视产业生态构建，降低使用成本与接入门槛。大模型作为 AIGC 产业的中游，需要与产业链上下游建立良好的技术适配机制；在算力资源方面，需加强与政府及龙头企业在算力资源方向的合作；在数据资源方面，需构建企业数据库，提高业务场景实际使用的适配度。

重视产业人才培养，培育招揽优化人才要素配置。大模型在研发环节和应用环节对于技术人员要求较高，企业应建立完善的人才招揽制度或人才培育体系，保障企业技术实力的不断提升。

5.3 对投资机构

大模型赛道技术潜力十足，企业成长空间巨大，建议重点聚焦涉及金融、工业、交通、政府行业的大模型企业，未来这四大领域将成为行业智能化的关键驱动领域。

大模型赛道投资回报周期长，建议关注落地场景和商业模式明确的企业；大模型优劣与优质数据资源相关性较强，

建议重视数据积累时间长、结构化程度高、带优质标签的企业。行业上重点关注金融、工业、政务和交通。

由于大模型“大”的特性对数据存储也有一定挑战，建议重点关注数据库厂商及分布式数据库厂商情况，由于企业大模型规模不断扩大，数据存储要求随之增高，会带来以上应用方向的爆发增长。