

中国AI大模型工业应用指数

中国工业互联网研究院
香港科技大学 中国经济信息社

2023年10月

工业互联网大厦

中国工业互联网研究院

1. 背景：人工智能发展政策体系不断完善

■ 国家高度重视人工智能发展

党的二十大报告

- 构建新一代信息技术、**人工智能**、生物技术、新能源、新材料、高端装备、绿色环保等一批新的增长引擎。

中共中央政治局会议（2023年4月28日）

- 要重视**通用人工智能**发展，营造创新生态，重视防范风险。

2023年二十届中央财经委员会第一次会议

- 要把握**人工智能**等新科技革命浪潮，推进产业**智能化**、绿色化、融合化。

■ 工信部持续出台政策文件深化人工智能在工业领域的应用

《工业互联网创新发展行动计划（2021-2023年）》

- 鼓励工业企业综合运用**人工智能**等技术，发展**智能化**制造，提升各要素智能化水平，实现全流程.....

《“十四五”信息化和工业化深度融合发展规划》

- 深化**人工智能**和工业融合应用，通过全面感知、实时分析、科学决策和精准执行，提升生产效率.....

《“十四五”智能制造发展规划》

- 提出加快研发**人工智能**在工业领域的适用性技术，推动**人工智能**等新技术在制造环节的深度应用.....

2. 中国AI大模型工业应用指数意义与体系

■ 指数意义

政府支持：服务于国家决策，帮助判断AI宏观发展趋势。

打造通用人工智能与工业融合发展的“风向标”，帮助了解人工智能与工业融合应用的整体与局部发展态势。

产品迭代：服务于技术革新，促进大模型应用迭代升级。

反映大模型工业应用的优势与不足，促进技术落地、产品升级。

业界指导：服务于行业发展，为企业智能化发展指明方向。

通过发布该指数，可研判大模型在工业领域各行业应用的成熟度，为行业企业智能化升级提供可行性建议。

为政府、研究机构、工业企业和社会公众提供一个权威、科学的综合评估，反映大模型技术在工业中的应用程度、产生效益和问题挑战。

■ 指数体系^[1]

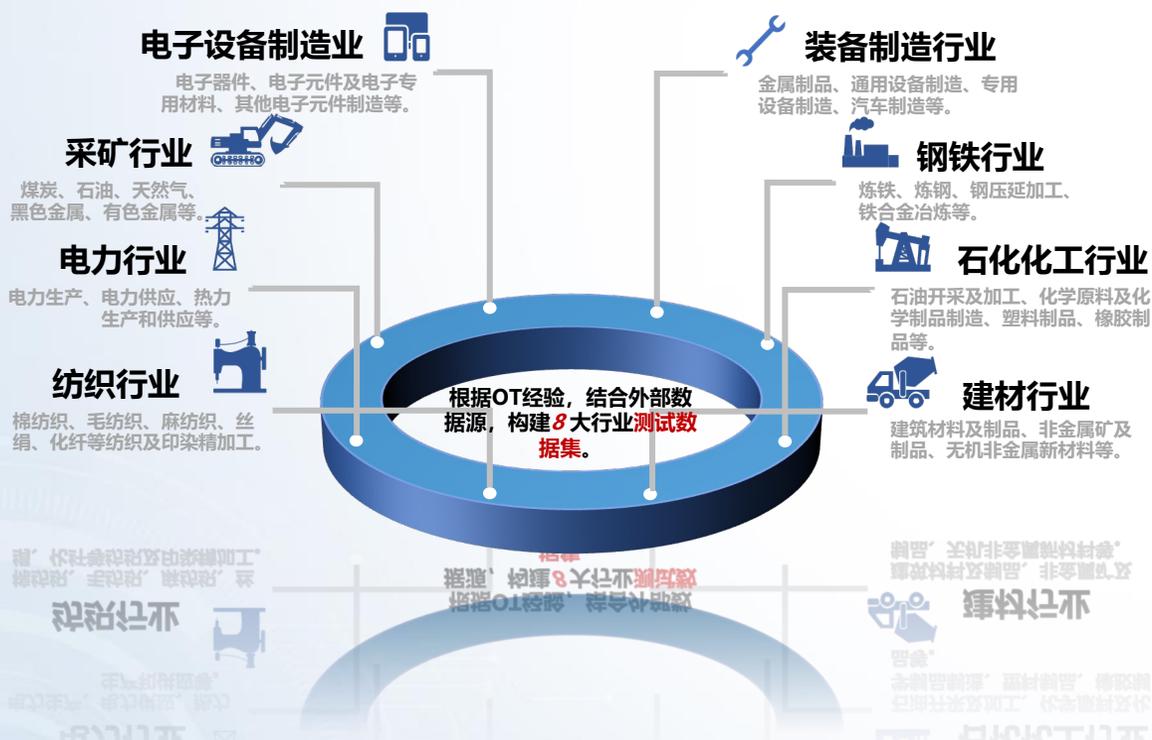


指数体系包括准确性指数，稳定性指数等细分指数。

[1] 指数体系会根据大模型应用发展不断迭代，当前版本关注大模型在工业领域中知识问答场景。

3. 中国AI大模型工业应用指数数据来源

■ 数据来源：八大行业



选取工业领域规上企业收入总值前八的行业
汇集知识问答数据。

■ 指标数据：准确性与稳定性报告

1. 准确性测评报告^[1]

通用人工智能大模型 工业领域知识问答性能评估

中国工业互联网研究院

2023年6月

2. 稳定性测评报告^[2]

人工智能大模型在工业领域知识问答： 稳定性评测

中国工业互联网研究院

香港科技大学

2023年9月

指标数据由前期发布的准确性与稳定性
实测值计算生成。

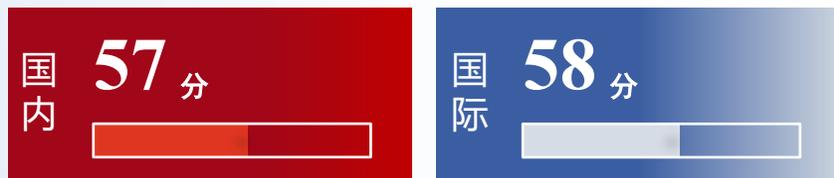
[1] 《通用人工智能大模型工业领域知识问答性能评估》，<https://china-aii.com/newsinfo/6080469.html> 4

[2] 《人工智能大模型在工业领域知识问答稳定性测评》，<https://china-aii.com/newsinfo/6382460.html>

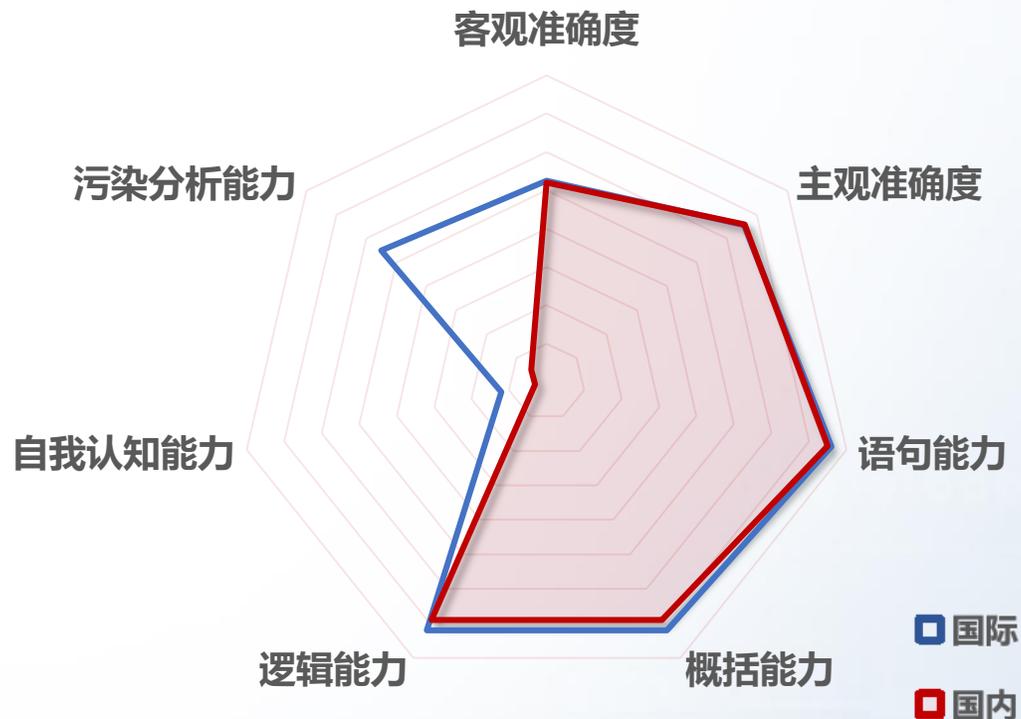
4. 中国AI大模型工业应用指数-准确性

准确性：在知识问答中能够正确回答问题的能力。

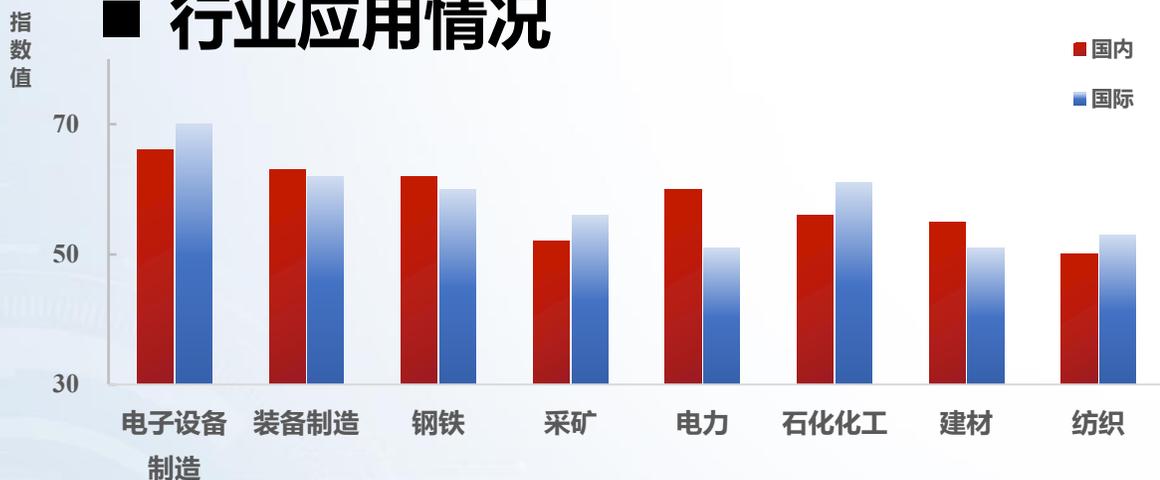
■ 国内外综合对比



■ 细分能力对比^[1]



■ 行业应用情况



- 整体上，国内大模型能力和国际大模型的能力接近；
- 行业上，国内大模型在电力、钢铁、建材等行业上优势明显，其他行业与国际水平相比有进一步提升空间；
- 细分能力上，国内大模型在自我认知能力和污染分析能力有一定差距，在其它能力维度表现接近。

[1] 污染分析能力、自我认知能力在指数计算时未单独计算，在细分能力对比图中展示仅为说明国内外大模型在这两方面的差异。

4. 中国AI大模型工业应用指数-准确性

■ 行业能力对比

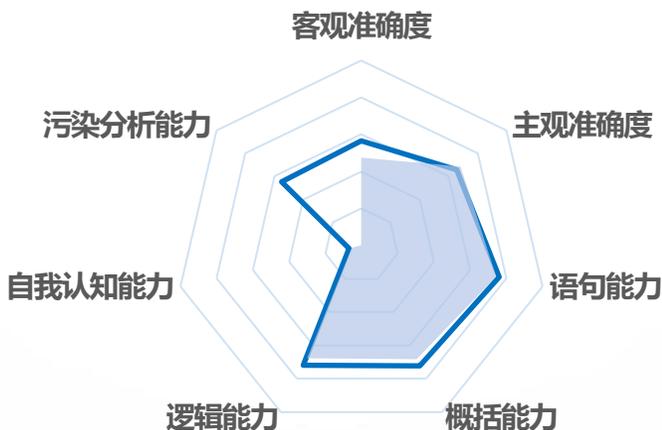
工业领域	国内外对比*
电子设备制造业	66:70
装备制造业	63:62
钢铁行业	62:60
采矿行业	52:56
电力行业	60:51
石化化工行业	56:61
建材行业	55:51
纺织行业	50:53

*国内: 国际

■ 国内大模型准确性第一梯队 (排名不分先后)

文心一言(百度)

在电子设备制造、钢铁、建材等行业位于第一梯队，有较强的语句能力、逻辑能力、概括能力。

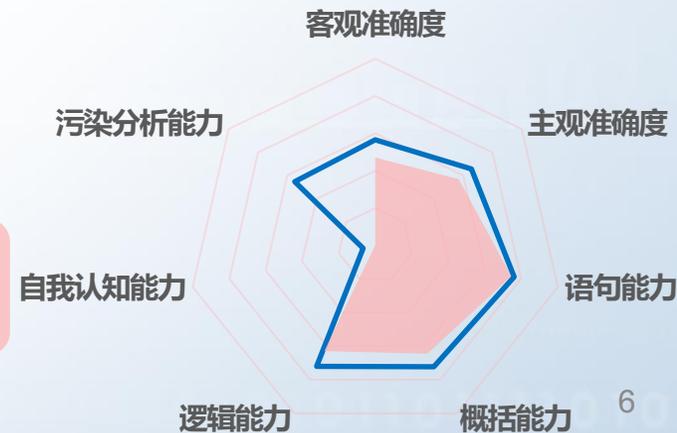


ChatGLM (智谱)

在装备制造、采矿、电力行业位于第一梯队，有较高的主观准确度，有较强的语句能力和概括能力。

360智脑 (360)

在装备制造、采矿、石化行业位于第一梯队，有较高的客观准确度，有较强的语句能力。

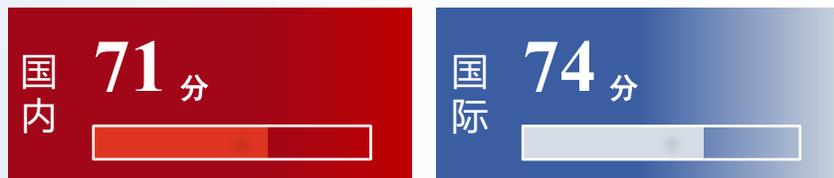


— 国外领先大模型均值

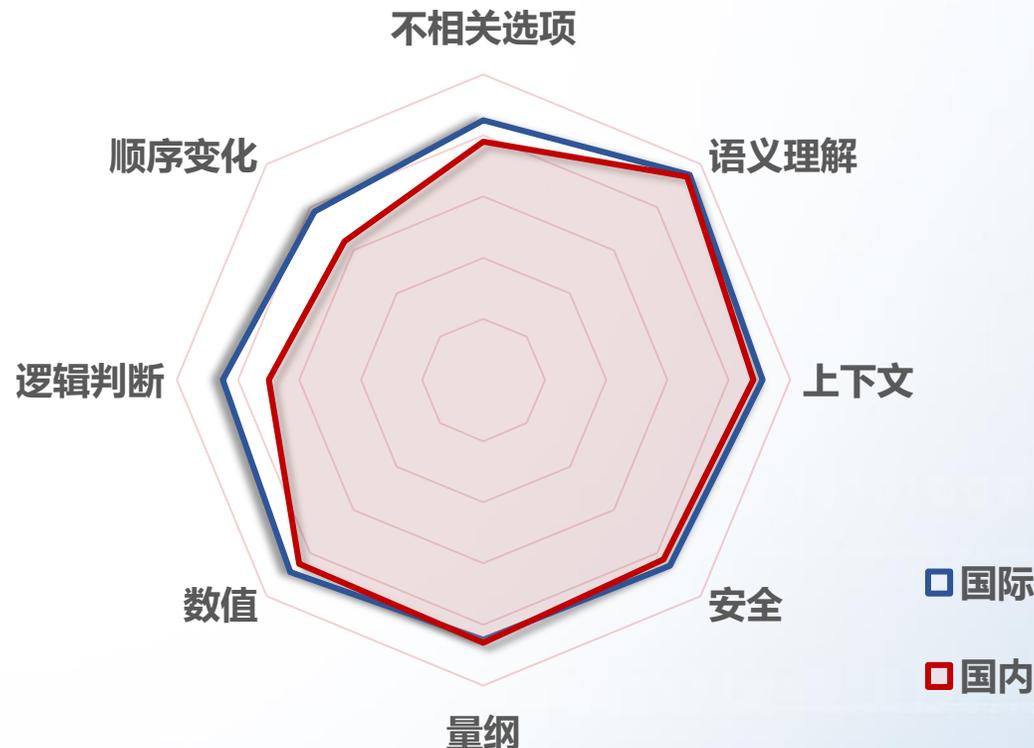
4. 中国AI大模型工业应用指数-稳定性

稳定性：在原问题回答正确的前提下，受到噪音干扰后仍然能输出正确答案的能力。

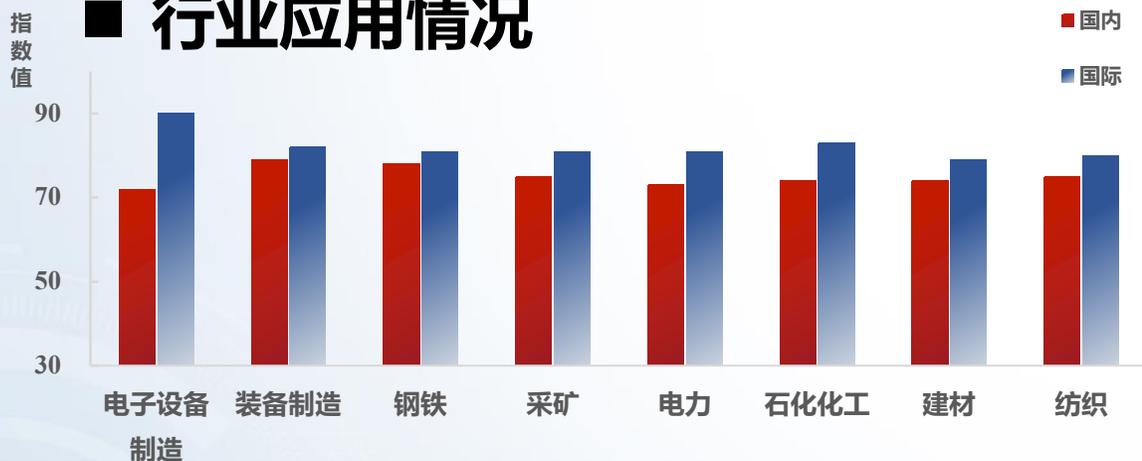
■ 国内外综合对比



■ 细分能力对比



■ 行业应用情况



- 整体上，国内大模型略低于国际大模型；
- 行业上，国内在电子设备制造、电力、石化行业距离国际水平有一定差距，其他行业较为接近；
- 细分能力上，国内模型在语义理解、上下文理解等语言任务上表现出色，但在逻辑判断、应对顺序变化、过滤无关信息等方面仍存在局限性，需要结合更多数据集和最新的提示工程技术进行改进。

4. 中国AI大模型工业应用指数-稳定性

■ 行业能力对比

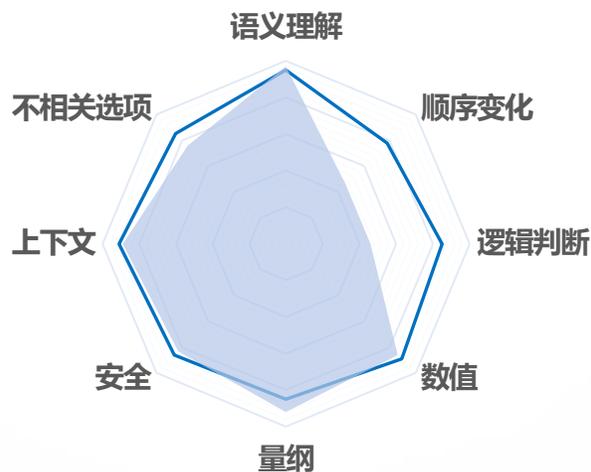
工业领域	国内外对比*
电子设备制造业	72:90
装备制造业	79:82
钢铁行业	78:81
采矿行业	75:81
电力行业	73:81
石化化工行业	74:83
建材行业	74:79
纺织行业	75:80

*国内: 国际

■ 国内大模型准确性第一梯队 (排名不分先后)

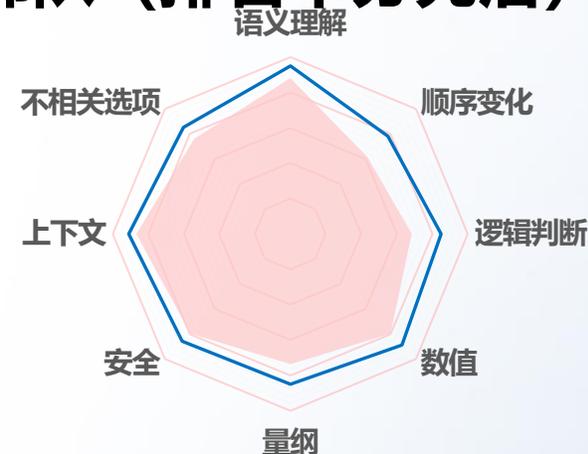
星火认知 (讯飞)

在电力、建筑、纺织行业位于第一梯队，有较高语义理解和顺序变化稳定性。



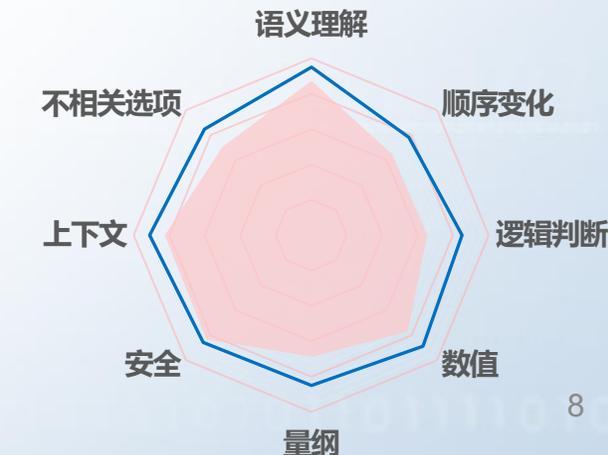
通义千问 (阿里)

在装备制造、钢铁行业位于第一梯队，有较高的语义理解、量纲处理和数值稳定性。



360智脑 (360)

在电子设备制造、石化、建材行业位于第一梯队，有较高的安全稳定性和不相关选项稳定性。



— 国外领先大模型均值

5. 中国AI大模型工业应用指数后续规划

动态更新指数报告

周期性更新模型指数。

不断完善工业语料库

完善工业语料库，扩充不同行业、不同领域、不同场景语料，为大模型应用指数提供测算基础。

开展行业模型测评

面向行业大模型的工业应用准确性与稳定性开展评测工作，在重点工业领域遴选一批优秀的行业大模型。

持续扩展测评行业

在现有的八大行业基础上不断增加测评行业，挖掘新的模型应用场景。

6. 专家组成员



唐立新，中国工程院院士，IEEE Fellow，东北大学副校长，第十四届全国人大代表，工业智能与系统优化国家级前沿科学中心主任和首席科学家。



王国栋，中国工程院院士，钢铁行业专家，博士生导师。



刘云浩，ACM Fellow，IEEE Fellow，长江学者，ACM主席奖、CCF王选奖获得者，清华大学教授，创新学院院长。



於志文，哈尔滨工程大学党委常委、副校长，教育部“长江学者”特聘教授，国家杰出青年科学基金获得者。



张涛，IET Fellow，清华大学教授，信息科学技术学院副院长，自动化系系主任。



张大庆，欧洲科学院院士，IEEE Fellow，北京大学讲席教授，CCF普适计算专委会主任。



刘江川，加拿大工程院院士，IEEE Fellow，加拿大西蒙菲莎大学计算机学院正教授和大学杰出教授。



刘劫，IEEE Fellow，ACM杰出科学家，哈尔滨工业大学讲席教授，人工智能研究院院长。



张燕咏，IEEE Fellow，中国科学技术大学教授，计算机科学与技术学院副院长，ACM中国副主席，科技部下一代人工智能重大项目负责人。

6. 专家组成员



贾佳亚, IEEE Fellow, 香港中文大学终身教授, 全球计算机视觉、人工智能领域权威专家, 国家科技部“新一代人工智能”重大项目核心专家。



杨铮, IEEE Fellow, 清华大学副教授、博士生导师, 国家万人计划。



黄河燕, 北京理工大学计算机学院教授, 国家科学技术进步奖一等奖获得者。



聂再清, 清华大学国强教授, AIR首席研究员, 微软学术搜索发起人, 曾任微软自然语言理解平台LUIS技术负责人。



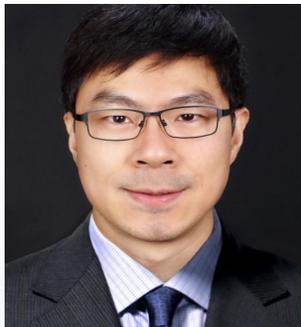
沈阳, 清华大学教授, 博士生导师, 中宣部“文化名家”暨“四个一批”入选者、教育部新世纪人才计划入选者。



刘云新, 清华大学国强教授, 智能产业研究院首席研究员, 前微软亚洲研究院主任研究员, 国家重点研发计划项目负责人。



张晓明, 北京雁栖湖应用数学研究院工业与应用首席研究员, 美国麻省理工学院博士。



任磊, 北京航空航天大学教授, 博士生导师, 复杂产品智能制造系统技术全国重点实验室专委会副主任, 国家杰青。



王帅, 香港科技大学计算机系助理教授, 博士生导师, 谷歌学者。

下一步，中国工业互联网研究院、港科大、中国经济信息社持续更新“中国AI大模型工业应用指数”，加速推动人工智能技术与实体经济融合创新。在大模型问答性能、应用安全、标准制定等方面展开研究，欢迎交流合作。

转载请注明来源：中国工业互联网研究院。

联系人：叶老师 13661350566 吴老师 15210188099

地址：北京市朝阳区酒仙桥北路
甲10号403号楼

电话：010-87901276

网址：www.china-aii.com



附录1：中国AI大模型工业应用指数-准确性评价标准

评分标准：

1. 每个行业抽取138道题目进行测试，其中：
 - 单选题(105)，判断题(30题)，简答题(3题)；
2. 对于客观题
 - 回答正确得分（1分），错误不得分（0分）；
3. 对于主观题
 - 按照基础能力，语句能力，概括能力，逻辑能力四个维度进行评价，每个维度5分；
 - 对于每个维度，分五个等级，分别对应1-5分。

准确性指数定义：

$$\text{模型维度准确性指数} = \frac{\text{模型维度全部题目得分}}{\text{全部题目总分}} \times 100$$

$$\text{各能力维度准确性指数} = \frac{\text{能力维度第一梯队模型总分}}{\text{第一梯队模型数量}} \times 100$$

$$\text{各行业维度准确性指数} = \frac{\text{行业维度第一梯队模型总分}}{\text{第一梯队模型数量}} \times 100$$

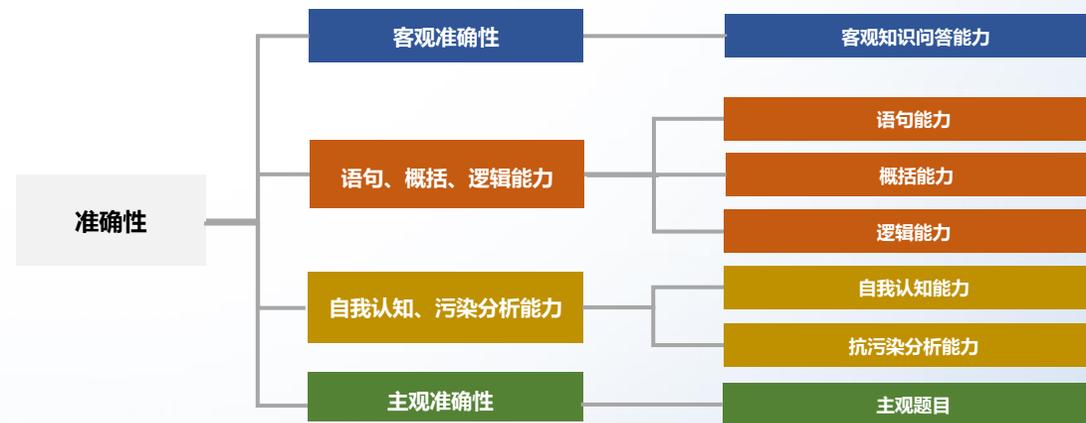
$$\text{国家/行业综合准确性指数} = \frac{\text{第一梯队模型总分}}{\text{第一梯队模型数}} \times 100$$

准确性测试流程：



准确性评测能力维度体系：

准确性评测的能力维度，围绕AI大模型在客观准确性、主观准确性等子维度展开深入评估。



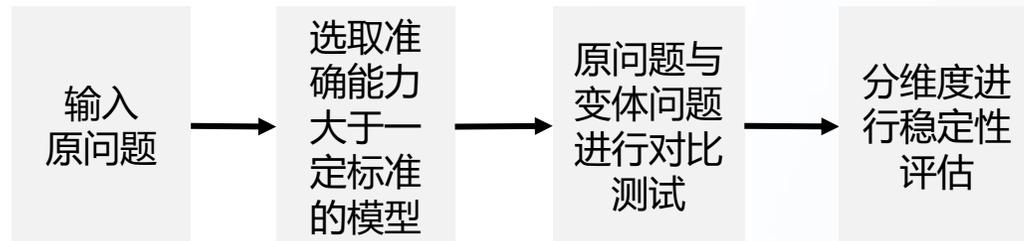
- 注：1. 主观题评价标准部分参考高考作文评分标准；
2. 主观题由专家组成评审委员会，针对大模型的回答举行集中评审；
3. 本研究仅针对公开通用大模型进行直接测评，对微调后的模型不做评价。

附录2：中国AI大模型工业应用指数-稳定性评价标准

评分标准：

1. 将原问题与对应变体问题进行对比测试，题型包含单选题和判断题两种类型题目；
2. 变体题目从准确测试题目中经过变换得到；
3. 若变体与原问题回答一致时，得分（1分）；
4. 若变体与原问题回答不一致时，不得分（0分）。

稳定性测试流程：

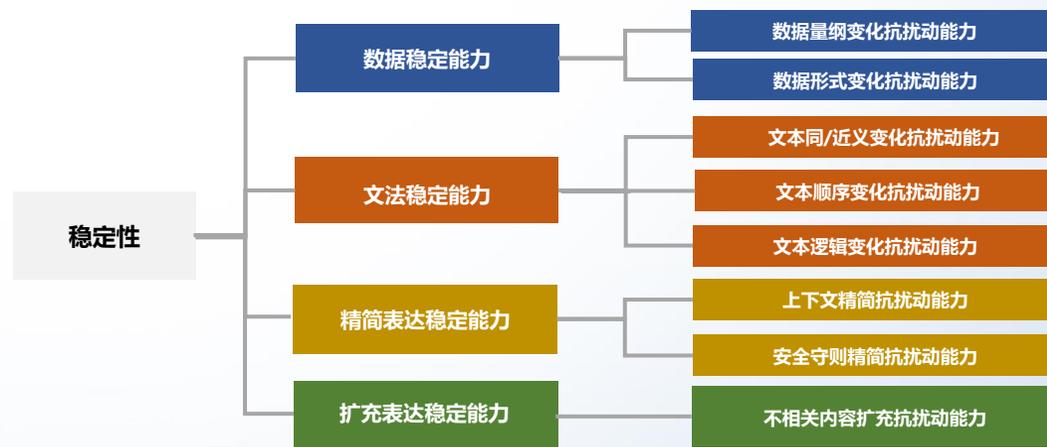


稳定性能力维度体系：

稳定性评测围绕AI大模型在数据、语法、精简表达、扩充表达四大类八个子维度展开深入评估。

稳定性指数定义：

模型维度稳定性指数	=	$\frac{\text{模型各能力维度总分}}{\text{能力维度数}}$	X 100
能力维度稳定性指数	=	$\frac{\text{能力维度第一梯队模型总分}}{\text{第一梯队模型数量}}$	X 100
各行业维度稳定性指数	=	$\frac{\text{行业维度第一梯队模型总分}}{\text{第一梯队模型数量}}$	X 100
国家综合稳定性指数	=	$\frac{\text{第一梯队模型总分}}{\text{第一梯队模型数}}$	X 100



注：经调研及评测专家组认定，该评测体系符合大模型工业日常应用稳定性基本逻辑。

附录3：指数报告涉及的大模型评测版本

大模型名称	公司名称	版本信息	Web链接
文心一言	百度	v2.2.2	https://yiyan.baidu.com
星火	科大讯飞	v2.0	https://xinghuo.xfyun.cn
ChatGLM2	智谱华章	6b	https://open.bigmodel.cn
通义千问	阿里巴巴	v1.0.5	https://qianwen.aliyun.com
360智脑	360	V4.0	https://chat.360.cn
GPT-3.5	OpenAI	gpt-3.5-turbo-0301	https://chat.openai.com
GPT-4	OpenAI	gpt-4-0613	https://chat.openai.com

注：1. 本研究实测模型只限于通用大模型；
2. 国际以GPT3.5和GPT4 为测量标准。