

# 2023年中国大模型 行研能力评测

(摘要版)

AI变革行业创新发展

2023 China Large Language Model Industry Research Evaluation

2023年中国大規模モデル産業研究能力評価

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，头豹研究院保留采取法律措施、追究相关人员责任的权利。头豹研究院开展的所有商业活动均使用“头豹研究院”或“头豹”的商号、商标，头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表头豹研究院开展商业活动。

## 摘要

- 大模型是自然语言处理领域中深度学习技术的高级应用。通过对海量文本数据进行深度训练，能够吸收和提炼出丰富的语言学知识和语义内涵。这种大模型通常参数规模庞大，动辄数亿甚至数十亿，从而具备理解和生成自然语言的能力，与人类沟通几乎无障碍。它的应用场景广泛，包括但不限于机器翻译、智能问答、文本摘要等多个领域。从技术发展的视角来看，语言模型的演进历程颇具启发性。起初，基于深度神经网络进行训练；随后，业界逐渐采用预训练结合微调的模式来提升性能；而如今，最先进的范式已经转向预训练结合提示训练，这无疑标志着模拟人类思维交流方面的又一重大突破。这一发展轨迹不仅凸显了技术的迅速迭代，更揭示了自然语言处理领域巨大的探索空间和创新潜力
- 大模型能够显著增强研究报告的编写效率，通过采纳精细化的询问策略，分析师在研究过程中能够实现效率的最优化。此外，这一模型还能够提供全面的文本编辑支持，包括校正、查重、以及文风润色等功能，从而确保分析师提交的工作成果达到更高的质量标准
- 大模型技术潮流兴起催生中国大模型百花齐放。沙利文与头豹研究院联手对12个大模型进行综合评估，以深入梳理中国大模型在行研领域的应用表现，为各界提供清晰深入的视角，以应对预训练大模型引发的人工智能浪潮

当前，自然语言处理技术中的预训练大模型正在全球范围内引领一场前所未有的人工智能革命。自ChatGPT推出以来，仅在中国地区，涌现出超过80个不同的预训练语言大模型。这些模型的参与者涵盖了中国顶尖的学术研究机构 and 领先的互联网科技企业，均希望能在这场技术浪潮中抢占先机。值得一提的是，过去一年里，中国学术和产业界也取得了显著进展，商汤的“商量”、百度的“文心一言”等大模型持续迭代，推动了中国大模型产业的快速发展。

沙利文与头豹研究院合作，基于数字行研解决方案的研究和实践经验，凭借百人分析师团队的匿名投票机制，筛选出12个大模型进行综合评估。这一评估旨在全面了解并梳理中国大模型在行研领域的应用表现。通过这份专业分析，各界能获取到清晰、深入的视角，以更好地理解 and 应对由预训练大模型引发的人工智能浪潮。

# 内容目录

## 1 大模型行研能力评测背景及方法

05 页

- 行研背景概述
- 传统行研发展痛点
- 数字行研革新
- 大模型赋能行研
- 评测大模型参与者
- 评测方法与指标
- 报告撰写能力
- 行研基础能力
- 行业理解能力

## 2 大模型行研能力评测结果 15 页

- 评测结果总榜
- 报告撰写能力总榜
- 报告撰写能力综合表现
- 报告撰写高阶难度模块表现
- 报告撰写低阶难度模块表现
- 报告撰写能力分布
- 行研基础能力总榜
- 行研基础能力综合表现
- 行研基础能力细分模块
- 行研撰写能力分布
- 行业理解能力总榜
- 行业理解能力综合表现
- 行业理解能力细分行业

# 研究目标

## 研究目的

基于数字行研解决方案的研究和实践基础，头豹研究院联合沙利文凭借百人分析师团队匿名投票机制，筛选了12个大模型，进行了多维度的综合评估，旨在全面了解并系统梳理中国大模型参与者在行研领域的应用表现。

## 研究目标

- 了解中国大模型在行研领域的应用价值
- 探析中国大模型在行研领域的综合表现
- 梳理中国大模型的参与者现状
- 评估中国大模型行研领域的发展潜力

## 本报告的关键问题

- **参与者发展现状：**中国大模型的参与者有哪些？在行研领域的发展现状如何？
- **行研应用表现：**中国大模型在行业研究领域的实际应用表现如何？
- **报告撰写能力：**中国大模型在撰写行业研究报告方面的表现能力如何？
- **模型基础能力：**中国大模型逻辑推理能力、类比迁移能力等基础能力表现如何？
- **行业理解能力：**目前中国大模型在不同行业下的理解能力以及表现如何？



# Chapter 1

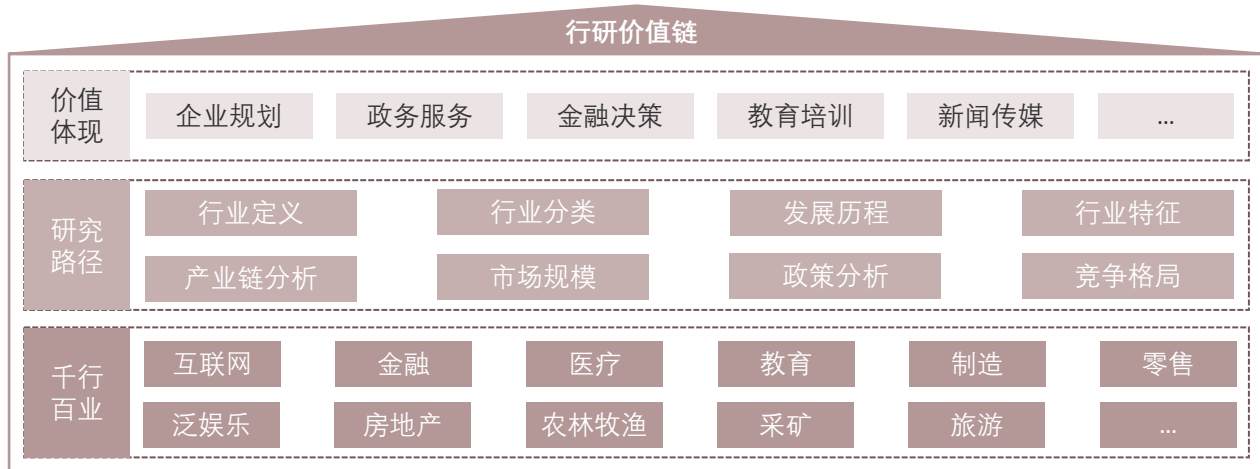
## 大模型行研能力评测背景及方法

- 本次大模型行研能力测试覆盖1,800+题目，由20人资深研究分析师团队经过严格的双盲评测流程，围绕研究报告撰写能力，模型基础能力以及行业综合理解能力进行综合评测
- 评测方法通过双盲机制最大程度保证公允性，每名测试人员会随机分配N个模型进行答案搜集，彼此在答案搜集期间互相禁止分享信息，以保持答案评测阶段的公允性；在评测阶段，每个问题相对应的12个模型答案顺序会随机打乱，保证评测人员对答案不存在任何偏见
- 本次大模型行研能力评测分为三个核心模块，分别为报告撰写能力、模型基础能力以及行业理解能力。报告撰写为本次评测的核心能力，基础能力以及行业理解能力为衍生能力，顾赋予40%/30%/30%的评分权重
- 本次大模型行业研究能力评测截止于11月30日，所选模型基于目前开放可用的范围。本次评测未涵盖企业尚未开放的大模型，评测结果仅反映当前公开可用模型的数据

# 中国大模型行研能力评测——行研背景概述

- 行业研究通过分析特定行业的定义、竞争格局、市场规模等关键方面，产出深刻洞察和观点。方法论涵盖从宏观的产业层到微观的产品层的分析，对企业战略、政策制定和金融决策等产生显著影响

## 行研背景概述



- 行业研究通过详尽方法论输出价值观点，对企业策略、金融决策等领域的价值显著。行业研究是深入探讨分析特定行业的发展现状和市场动态的全面过程，包含行业定义、分类、竞争格局、市场容量等关键维度。分析师通过分析，生成具有深度的洞察和价值观点，对企业战略规划、政策制定、金融投资决策和教育培训等多个领域产生重要影响。

行研范围	外部维度				内部维度				
	政策	经济	环境	法律	发展历程	产业链分析	发展特征	市场规模	竞争格局
产业层	重要程度高	重要程度高	重要程度高	重要程度高	重要程度低	重要程度低	重要程度低	重要程度低	重要程度低
行业层	重要程度低	重要程度低	重要程度低	重要程度低	重要程度高	重要程度高	重要程度高	重要程度高	重要程度高
产品层	重要程度低	重要程度低	重要程度低	重要程度低	重要程度低	重要程度高	重要程度低	重要程度低	重要程度低

重要程度高    重要程度低

- 行业研究涵盖从宏观的产业层到微观的产品层，各层级决定着相应的研究方法，研究方法论囊括外部宏观因素和内部微观细节的全面分析

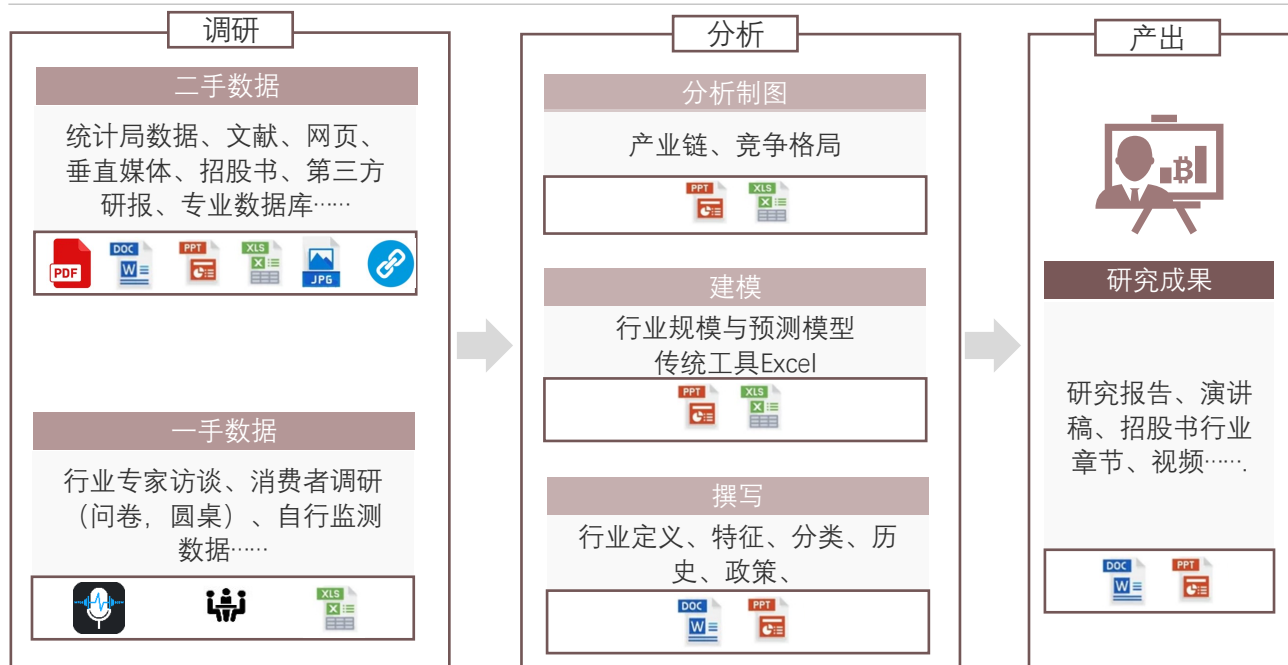
在行业研究中，产业层、行业层和产品层各自代表宏观经济的不同层次：产业层包含具相似特征的行业群体，行业层着重于特定行业的市场和企业情况，而产品层深入到具体产品或服务的设计、功能和市场定位。研究方法论根据这些层级的宏观到微观差异而有所不同，外部维度考虑政策、经济、环境等因素，而内部维度则包括发展历史、产业链分析等更细致的方面。

来源：沙利文、头豹研究院

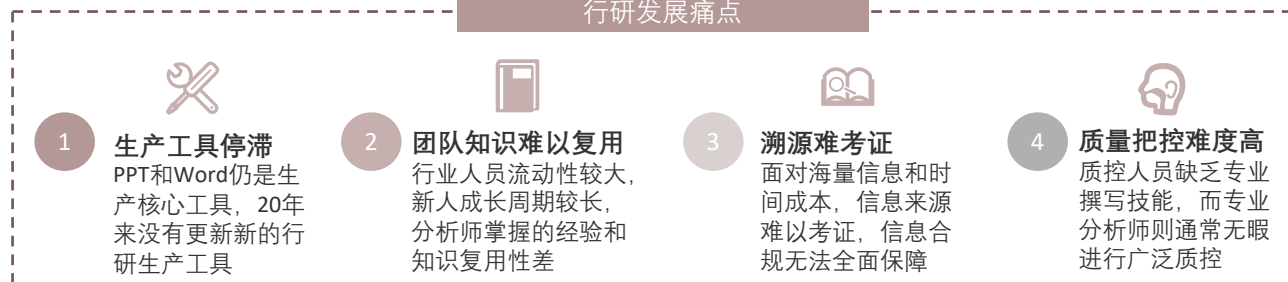
## 中国大模型行研能力评测——传统行研发展痛点

- 从基础数据收集到深度分析输出，传统行业研究的流程面临着工具革新滞后、团队知识难以传承、信息溯源复杂性以及研报质量控制的重挑战，共同影响行业研究的产出效率 and 创新能力，限制了其发展潜力

### 传统行研发展痛点



### 行研发展痛点



- 传统行业研究在工具更新、知识复用、信息溯源和质量控制方面面临显著挑战，影响其效率和准确性

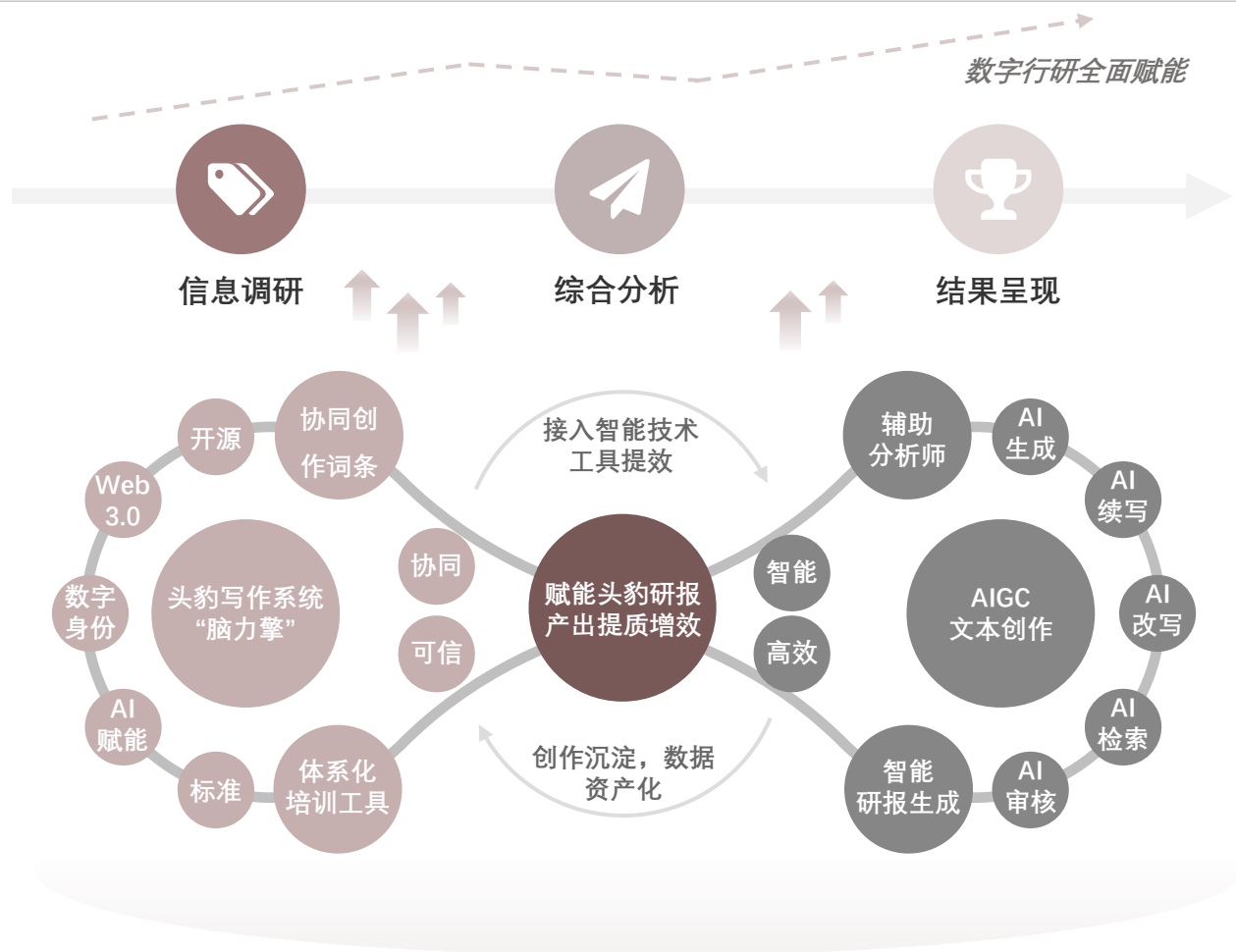
传统行业研究的产出流程包括三个步骤：首先是基础调研，聚焦于一手和二手行业数据的收集；其次是数据加工，涉及整理逻辑、验证数据真实性，并对关键信息进行可视化处理；最后是产出结果，确保全文逻辑一致、可视化清晰并且观点合理。在实际操作中，行业研究面临多个挑战：1) 工具更新停滞，自互联网兴起以来，行研主要依赖网络检索和办公软件，近20年未见显著革新；2) 团队知识难以复用，由于高人员流动性和新成员培养周期长，分析师的经验和知识传承困难；3) 信息溯源和合规性考量复杂，在应对海量信息和时间成本压力下，信息来源和合规性难以全面保障；4) 质量控制难度高，质控人员缺乏专业撰写技能，同时专业分析师缺乏时间进行广泛质控，导致质量监控极为困难。这些核心挑战共同影响着行研的产出效率 and 创新，阻碍行研进一步发展。

来源：沙利文、头豹研究院

## 中国大模型行研能力评测——数字行研革新

- 数字行研解决方案，结合标准化工具和先进的大模型技术，有效克服了传统行业研究的核心制约因素，显著提升研究的精度和效率，并引领行业研究进入一个效率更高和质量更优的新产出范式

### 数字行研革新（以头豹脑力擎为例）



- 数字化行研解决方案通过标准化工具和大模型技术，解决了行业研究的核心难题，提高了研究质量和效率，为行业分析带来了高效、精准的新范式

数字化行研解决方案为行业研究带来范式革新。该系统通过引入一系列标准化工具，如精准的信息溯源系统和写作规范，极大降低了行业研究的门槛，提升了撰写效率，并简化了质量控制流程。此外，数字行研基于大模型的强大底层支持，为分析师提供AI辅助工具，包括专家访谈、智能检索、查重、校对和改写功能，有效减少信息检索的难度，提升文本的准确性和撰写效率，同时加速分析师的专业成长。

数字行研解决方案不仅解决了传统行业研究中的关键痛点，如信息溯源困难、严峻的质控挑战、知识低复用率和工具更新滞后，还借助大模型技术推动行业研究质量和分析师成长速度，引领行业研究向更高效、精准的新范式转变。

来源：沙利文、头豹研究院

## 中国大模型行研能力评测——大模型赋能行研

- 大模型在数字行业研究中扮演核心角色，其功能特性极大提升了研究的效率和质量。本研究聚焦于挖掘中国大模型在行业研究中的实际应用和优势，了解大模型当前的能力边界，以推动行研领域的创新与变革

### 大模型赋能行研



- 大模型在数字行业研究中扮演核心角色，其功能特性极大提升了研究的效率和质量。本研究主要聚焦于中国大模型在行业研究中的实际应用和优势，知晓大模型的能力边界，旨在推动行研领域的创新与变革

作为数字行研的关键支撑，大模型通过其创作、生成、改写和检索等核心功能，全方位推动行业研究向前发展。首先，大模型扮演第三方AI专家角色，协助分析师在研究初期进行框架搭建和内容创作，有效减轻案头工作负担。其次，通过与分析师的有效互动，大模型助力生成结构化内容和洞察，显著提高基础内容的产出效率。进一步，它通过减少文本错误和重复内容，优化校对流程，提升产出质量。最后，大模型能够快速处理海量数据，提供实时信息检索，增强分析师在有限时间内获取全面信息的能力。

鉴于此，深入了解中国大模型行业的发展态势和技术创新对引领数字行业研究新范式至关重要。本研究的目标是全面评价中国领先的大模型技术，深入分析其在行业研究领域的实际应用和优势，全面审视大模型对行业研究发展的影响，以促进数字行业研究的创新和变革。

来源：沙利文、头豹研究院



## 中国大模型行研能力评测——评测大模型参与者

- 基于数字行研解决方案的研究和实践基础，沙利文联合头豹研究院依托百人分析师团队对12个大模型进行行研辅助能力的综合评估，旨在全面了解并系统梳理中国大模型参与者在行研领域的应用表现

### 大模型行研能力评测参与者



本次对大模型的行业研究能力进行评测的时间定于11月15日至11月30日，评选的模型基于目前市场上开放且可用的版本。评测结果将仅反映在评测时段内公开可获取的模型数据。在此次评测中，GPT模型选用的公测版本为3.5版本，文心一言模型也采用其3.5公测版本进行评估。

- 自ChatGPT推出后，中国在预训练大模型领域实现了显著进步，涉及顶尖学术机构和科技企业，沙利文联合头豹研究院对12个大模型进行综合评估，以全面了解中国大模型在行研领域的发展与应用

当前，基于自然语言处理技术的预训练大模型已在全球范围内掀起了有史以来最大的人工智能浪潮。自ChatGPT推出以来，仅中国地区就出现了超过80个不同的预训练语言大模型，参与者覆盖中国顶尖的学术研究机构以及互联网科技企业，旨在此番浪潮中拔得先机。过去一年中，中国学术与产业界也取得了实质性的突破，来自商汤的商量、百度的文心一言等前沿大模型不断升级，带动中国大模型产业的发展。

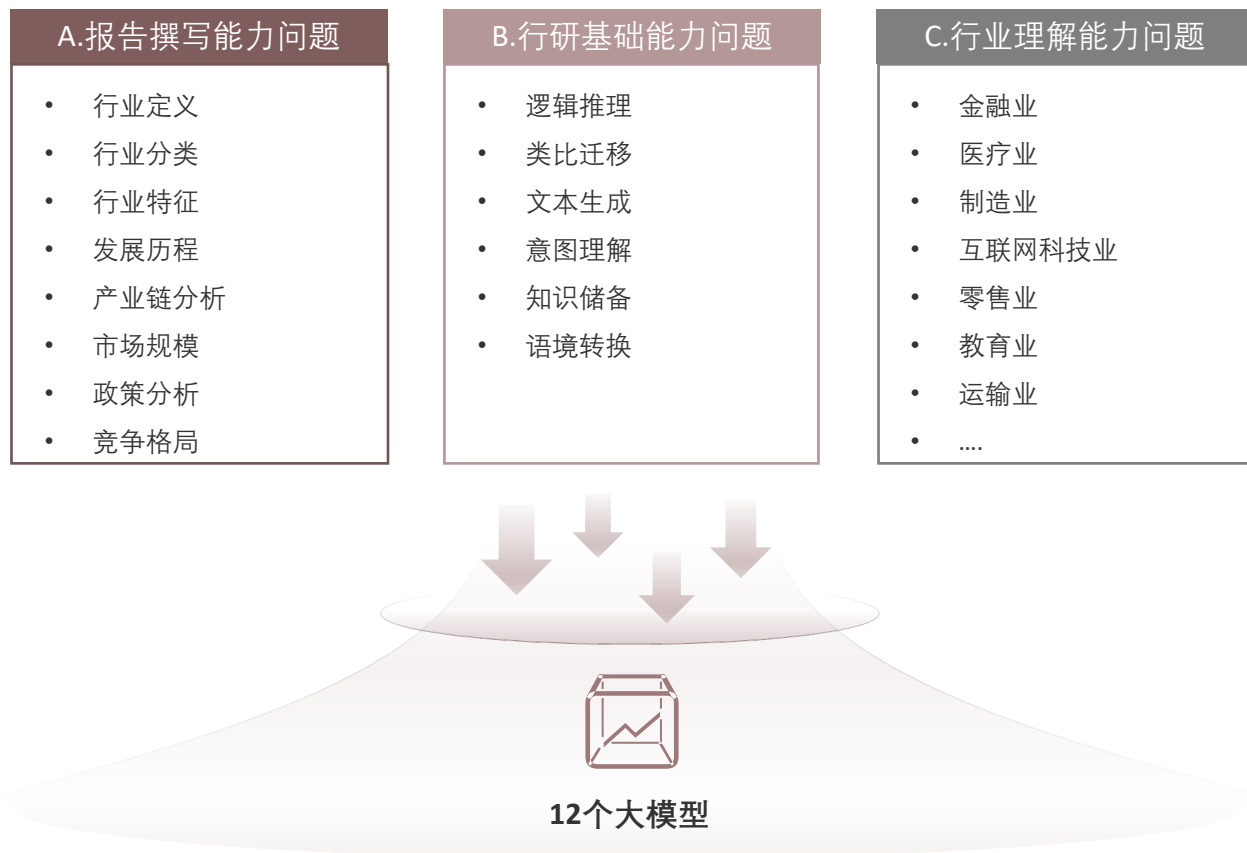
基于数字行研解决方案的研究和实践基础，沙利文联合头豹研究院凭借百人分析师团队匿名投票机制，筛选了12个大模型，进行了多维度的综合评估，旨在全面了解并系统梳理中国大模型参与者在行研领域的应用表现。

来源：沙利文、头豹研究院

## 中国大模型行研能力评测——评测方法与指标

- 本次大模型行研能力测试覆盖1,800+题目，由20人资深研究分析师团队经过严格的双盲评测流程，围绕研究报告撰写能力，模型基础能力以及行业综合理解能力进行综合评测

### 大模型行研能力评测方法



- 本次测试覆盖三大板块，1,800+道题目，由20人资深研究分析师团队经过严格的双盲评测流程得出，最大程度保证公允性

本次大模型能力测试围绕三大核心板块展开评测：分别是研究报告撰写能力，模型基础能力以及行业综合理解能力。通过模型在三大核心板块的表现力最终得出评测结果。其中，报告撰写覆盖8篇不同主体报告撰写，涵盖128道问题，分析师长期跟踪报告问题累积超1,500道题；模型能力覆盖6大文本产出核心能力，涵盖54道问题；行业理解覆盖14大核心行业，每个行业涉及12个问题，总计1,168题。三大维度合计1,800道题目。分析师团队由头豹研究院各团队资深分析师组成，且均具备超过8个月的大模型使用经历。

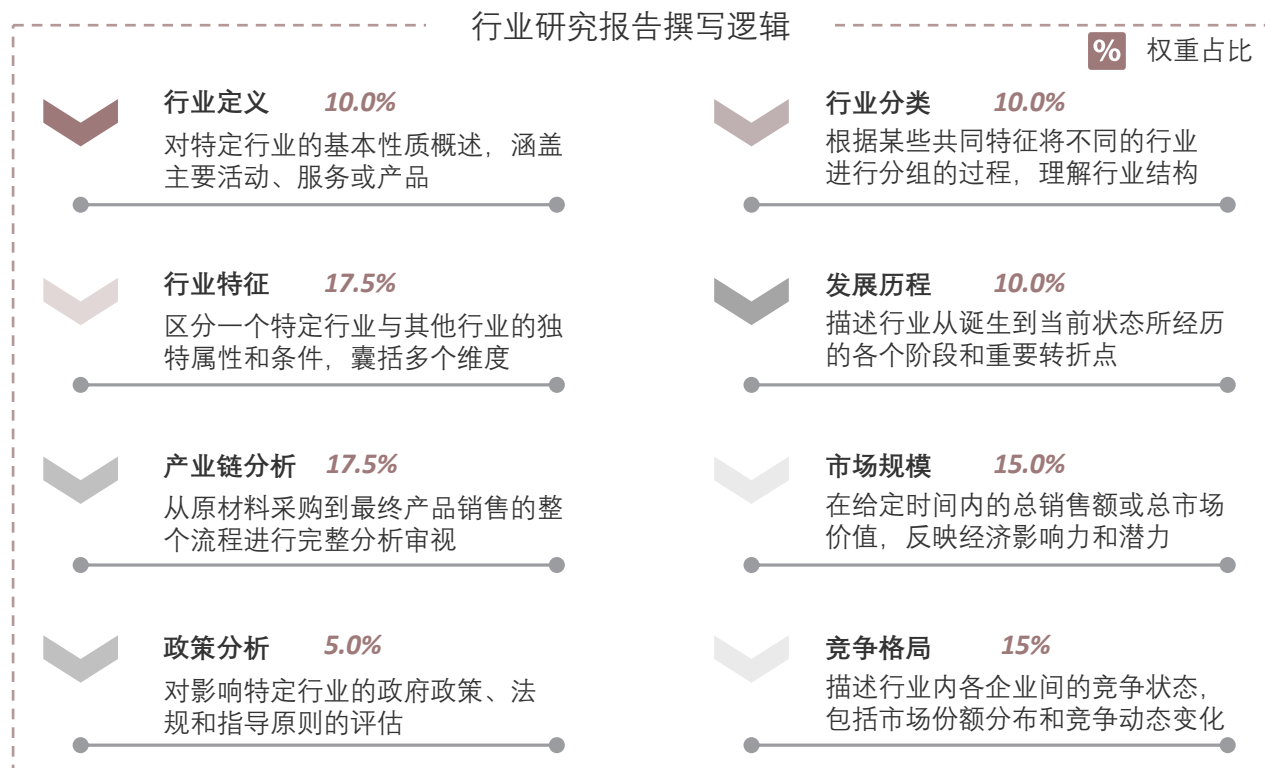
评测方法通过双盲机制最大程度保证公允性，每名测试人员会随机分配N个模型进行答案搜集，彼此在答案搜集期间互相禁止分享信息，以保持答案搜集阶段的公允性；在评测阶段，每个问题相对应的12个模型答案顺序会随机打乱，保证评测人员对答案不存在任何偏见。

来源：沙利文、头豹研究院

## 中国大模型行研能力评测——报告撰写能力

- 沙利文及头豹行企研究的8-D方法论，是一种全面系统的研究方法，包含了八大关键模块，用于对行业进行深入分析。在这一框架下，百名分析师研磨提炼一套高效的8D模块化大模型提示问题法，以对模型能力进行评测

### 大模型报告撰写能力



### 百名分析师经过八个月实践研磨提炼8D模块的提示问题，以测试12大模型的报告撰写能力

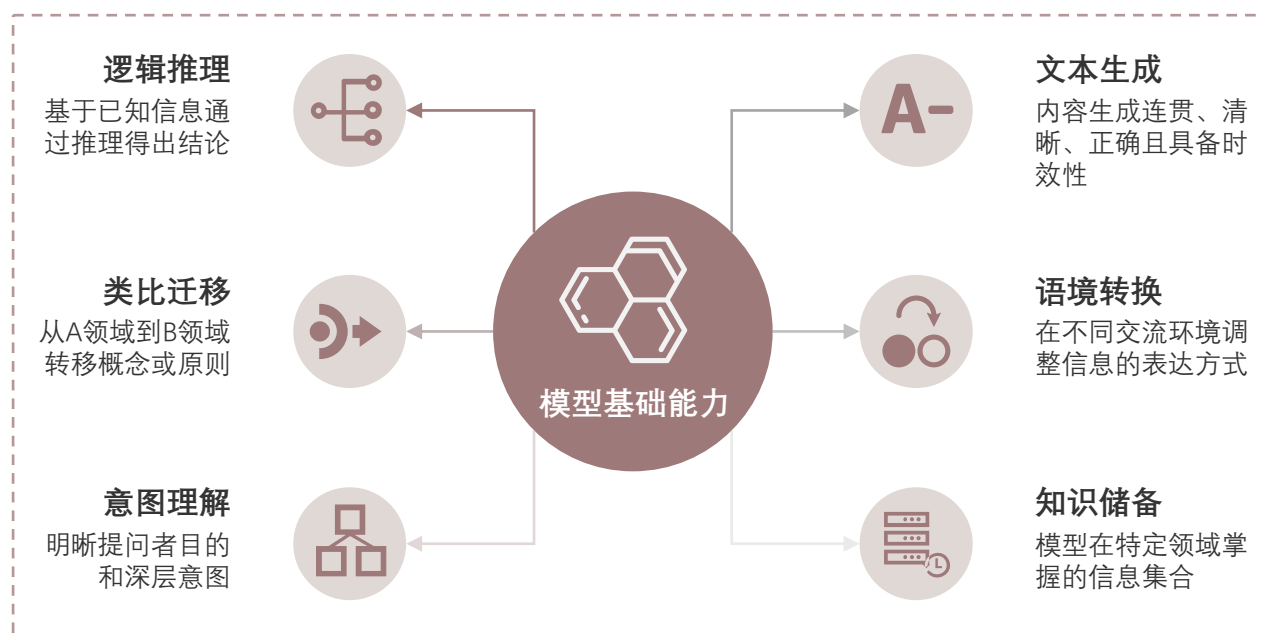
沙利文联合头豹企业研究所采纳的8-D方法论包括八个核心模块，构成了一套全面而系统的研究方法，专用于行业的深入分析。结合详实的数据和精准的分析，此方法能够提炼出深刻的观点和结论，显著提升行业研究内容的清晰度和数据的丰富性。在该方法论的指导下，百名分析师历经八个月的密集工作及多次优化，精制而成一套高效的8D模块化大模型提示问题法。这套精练的提问技巧已转化为评估工具，旨在通过对十二个主要模型的针对性提问，检验并评估模型报告撰写的效能。

来源：沙利文、头豹研究院

## 中国大模型行研能力评测——行研基础能力

- 从AI辅助文本创作角度出发，结合文字生成基础核心能力，归总出对于行业研究报告撰写角度最核心的六大能力维度，包括逻辑推理、文本生成、类比迁移、语境转换、意图理解以及知识储备

### 大模型行研基础能力



- 从AI辅助文本创作角度出发，结合大模型基础核心能力，归总出对于行业研究报告撰写角度最重要的六大能力维度

**逻辑推理：**逻辑推理是指从已知信息出发，通过推论规则得出结论的过程。在内容评判中，关注信息组织、连接和推导的方式，以及结论是否合理、一致，且基于事实。

**类比迁移：**类比迁移是指从一个领域或情境中提取概念、原则或模式，并应用到另一个不同的领域或情境。在内容评判中，评估模型在不同概念、情境之间建立联系的能力，以及这些联系的适当性和创造性。

**文本生成：**文本生成是指创建连贯、相关和有意义的文本内容。在评判内容时，评估文本的清晰度、连贯性、原创性以及语言的正确性和表达能力。

**意图理解：**意图理解是指识别和理解用户或作者想要传达的目的和动机。在内容评判中，评估信息是否有效地传达了其预期的消息或意图，以及模型是否能清楚地识别这些意图。

**知识储备：**知识储备是指个体或系统所掌握的信息、事实、概念和理论的总和。在内容评判中，知识储备体现在信息的准确性、深度和广度，以及模型能否正确并有效地使用相关知识。

**语境转换：**语境转换是指根据不同的交流环境或对象调整信息表达方式。在内容评判中，评估信息是否适应特定的受众、文化背景或沟通场合，以及是否能有效地调整语气、风格和内容以满足不同场景下的写作需求。

来源：沙利文、头豹研究院



## 中国大模型行研能力评测——行业理解能力

- 头豹研究院成立至今汇集超5,000+行企报告，覆盖超14个大类行业以及上千个细分小类行业。本次评测汇集百余名各资深行业分析师，结合自身领域认知，对模型在14大行业的理解和产出能力进行评估

### 大模型行业理解能力



- 头豹研究院成立至今汇集超5,000+行企报告，覆盖超14个大类行业以及上千个细分小类行业。本次评测汇集头豹研究院百余名各资深行业分析师，结合自身特定领域认知与经验，对模型在14大行业的理解和产出能力进行评估

沙利文联合头豹研究院成立至今，平台共积累超12万+注册用户，5,000+行业企业研究报告积累，覆盖14个大类行业，以及上千个细分小类行业。沙利文联合头豹的精英分析师团队以及各项研究成果广泛受到金融、制造、互联网科技等各个行业领域的用户认可。当前，头豹已发展成为中国最大的行企研究平台之一，拥有行业覆盖程度广、报告库数量多、报告撰写效率高、行业知识精准度高四大优势。

在本次大模型评测中，头豹研究院联合其上海、南京和深圳分院，汇聚了跨越多个行业领域的百余名分析师，利用自身对竞争格局、发展趋势、制约因素、以及行业壁垒等关键知识领域的深厚理解，并结合丰富的行业报告撰写经验，向模型提出了针对14个主要行业的细致问题。通过对模型在各特定行业细分领域的纵向评估和全行业范围的横向比较，最终确定了12大模型在行业理解和内容产出方面的深度与能力。

来源：沙利文、头豹研究院

## 方法论

- ◆ 头豹研究院布局中国市场，深入研究19大行业，持续跟踪532个垂直行业的市场变化，已沉淀超过100万行业研究价值数据元素，完成超过1万个独立的研究咨询项目。
- ◆ 研究院依托中国活跃的经济环境，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ◆ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

## 法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。本报告所指的公司或投资标的的价值、价格及投资收入可升可跌。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本文所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本文所载资料、意见及推测不一致的报告和文章。头豹不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

首席分析师

袁栩聪

☎ 15999806788

✉ oliver.yuan@frostchina.com

研究总监

李庆

☎ 13149946576

✉ livia.li@frostchina.com

🖥 [www.frostchina.com](http://www.frostchina.com) ; [www.leadleo.com](http://www.leadleo.com)

📺 <https://space.bilibili.com/647223552>

👤 <https://weibo.com/u/7303360042>

©弗若斯特沙利文咨询（中国）

©头豹研究院

