

2024年中国大语言模型能力评析（三）

行业应用能力评测结果

企业标签：百度、腾讯、商汤

AI变革行业创新发展

China Large Model Industry

中国大规模モデル産業

撰写人：王利华

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。

团队介绍

头豹是国内领先的行企研究原创内容平台和创新的数字化研究服务提供商。头豹在中国已布局3大研究院，拥有近百名资深分析师，头豹科创网(www.leadleo.com)拥有20万+注册用户，6,000+行业赛道覆盖及相关研究报告产出。

头豹打造了一系列产品及解决方案，包括数据库服务、行企研报服务、微估值及微尽调自动化产品、财务顾问服务、PR及IR服务，研究课程，以及分析师培训等。诚挚欢迎各界精英与头豹交流合作，请即通过邮件或来电咨询。

报告作者



袁栩聪
首席分析师
oliver.yuan@Leadleo.com



王利华
行业分析师
lihua.wang@leadleo.com

头豹研究院

咨询/合作

网址：www.leadleo.com

电话：15999806788（袁先生）

电话：18916233114（李先生）

深圳市华润置地大厦E座4105室

摘要

在行业应用能力领域，文心一言、通义千问及混元大模型等凭借其强大的泛化能力和深度学习技术，能够灵活适应并高效解决各行业复杂多变的实际问题，在各类行业能力评测中均展现出卓越的表现和广泛的应用潜力。尤其是在一些关键行业的能力评测中，这些头部大模型展现出了显著的领先优势，例如文心一言与通义千问已经在超过10个以上的行业中确立了其专业能力的领先地位。然而，对比之下，目前排名较为靠后的大模型在应对不同行业领域的专业问题时，其处理能力和适应性略显不足，尤其在面对行业细分场景和复杂任务挑战时，表现一般。这意味着，对于这些大模型而言，还有很大的空间去加强其行业特性和专业能力的优化提升，以满足更多行业需求，实现更全面、深入的应用落地。

- 中国大模型在行业应用与专业知识储备上表现出色，文心一言、通义千问尤为突出，并在道德伦理安全方面领先国际平均水平

从专业知识储备维度来看，文心一言、腾讯混元、通义千问、商量等大模型表现超国际水平，仍有部分大模型的专业知识储备能力处于中国均值以下，专业知识储备能力有待进一步提升。

从行业应用能力维度来看，中国多数大模型表现较好，展示了优秀的行业理解与应用能力，处于中国均值以上，其中文心一言、通义千问凭借超强的行业应用能力，表现亮眼，超国际均值。

从行业道德伦理安全维度来看，中国的大模型表现具有较大的差异，其中文心一言、通义千问、天工等大模型表现优异，超国际均值，少数大模型由于无法识别和理解行业中的道德伦理安全，表现一般。

研究框架

◆ 中国大语言模型能力评析	-----	5
• 行业能力评测结果	-----	5
• 大模型优势行业分析	-----	6
• 专业知识储备细分维度	-----	7
• 行业应用能力细分维度	-----	8
• 道德伦理安全细分维度	-----	9
• 行业能力细分行业	-----	10
◆ 方法论	-----	27
◆ 法律声明	-----	28

名词解释

- ◆ **大模型**：通常指的是在深度学习中使用的具有大量参数和复杂结构的模型。这些模型需要大量的计算资源和数据来进行训练。
- ◆ **深度学习**：深度学习是机器学习的一个子领域，它基于神经网络（尤其是深度神经网络）来学习数据的表示。这些网络通常由多个处理层组成，可以学习数据的多层次抽象。深度学习已成功应用于图像识别、语音识别、自然语言处理等许多任务。
- ◆ **数据孤岛**：指的是在组织或系统中，数据被分隔成多个独立、不兼容或难以整合的部分。这种情况是由于技术、组织或政策障碍造成的。数据孤岛导致信息不一致、重复工作和决策失误。
- ◆ **智能客服**：是利用人工智能技术（如自然语言处理、机器学习等）来自动回答客户问题、提供支持和帮助的系统。智能客服可以处理常见问题，减轻人工客服的负担，并提高客户满意度。
- ◆ **知识图谱**：是一种表示实体（如人、地点、概念等）及其之间关系的图结构。它可以用于组织、理解和共享网络知识。知识图谱在搜索引擎优化、推荐系统、问答系统等领域有广泛应用。
- ◆ **自然语言处理**：是人工智能和语言学的交叉领域，旨在让计算机理解和生成人类语言。NLP技术包括文本分析、情感分析、机器翻译、语音识别、文本生成等。这些技术使计算机能够处理和理解人类语言的复杂性。

中国大模型能力评析——行业能力评测结果

- 在行业能力板块，文心一言、通义千问、混元大模型能灵活适应并高效解决各行业实际应用场景中遇到的问题，在多行业能力评测中表现优异，而排名末位的大模型面对大多数行业的专业能力表现一般

大模型行业能力评测结果

模型名称	得分	传媒	电商	电信	法律	泛娱乐	房地产业	工业	互联网科技	教育	金融	旅游	能源	农业	线下零售	医疗	运输	政务
文心一言	7.23																	
通义千问	7.13																	
腾讯混元	7.00																	
日日新·商量	6.88																	
360智脑	6.87																	
百川智能	6.60																	
Moonshot (Kimi.ai)	6.57																	
智谱AI	6.47																	
豆包	6.22																	
天工	6.08																	
讯飞星火	5.95																	
雅意	5.87																	
MiniMax	5.75																	
紫东太初	5.67																	
面壁露卡	5.21																	

完整版登录 www.leadleo.com

搜索 《2024年中国大语言模型能力评析（三）：行业应用能力评测结果》

来源：沙利文、头豹研究院

中国大模型能力评析——大模型优势行业

- 在行业能力板块，大模型所拥有的优势行业呈现明显特征，头部大模型在多个行业中表现出领先优势，如文心一言、通义千问拥有超10个以上的优势行业，末位大模型的行业能力仍需进一步提升

大模型优势行业分析

模型名称	优势行业
文心一言	电商、电信、泛娱乐、房地产、工业、教育、旅游、医疗、运输、线下零售
通义千问	电商、电信、泛娱乐、房地产、教育、能源、农业、医疗、运输、政务、线下零售
腾讯混元	传媒、法律、泛娱乐、政务、教育、金融、旅游、线下零售、互联网科技
日日新·商量	房地产、法律、农业、教育、金融、旅游、线下零售
360智脑	传媒、法律、农业、教育、金融、旅游、线下零售
百川智能	法律、农业、互联网科技
Moonshot (Kimi.ai)	电商、法律、能源
智谱AI	金融、能源
豆包	传媒、互联网科技
天工	泛娱乐
讯飞星火	房地产、教育
雅意	泛娱乐
MiniMax	泛娱乐

完整版登录 www.leadleo.com

搜索《2024年中国大语言模型能力评析（三）：行业应用能力评测结果》

来源：沙利文、头豹研究院

中国大模型能力评析——专业知识储备细分维度

- 从专业知识储备维度来看，文心一言、腾讯混元、通义千问、商汤日日新·商量等大模型表现超国际水平，仍有部分大模型的专业知识储备能力处于中国均值以下，专业知识储备能力有待进一步提升

大模型行业能力专业知识储备维度评测结果

专业知识储备

..... 国际均值: 6.69

---- 中国均值: 6.12



完整版登录 www.leadleo.com

搜索 《2024年中国大语言模型能力评析（三）：行业应用能力评测结果》

- 文心一言、腾讯混元、通义千问、商汤日日新·商量以及360智脑超国际均值，但中国仍有部分大模型表现一般，在专业知识储备能力仍有待提升

文心一言、通义千问、商汤日日新·商量以及360智脑之所以能够达到如此高的知识储备水平，背后是海量的数据支撑和先进的算法驱动。以文心一言为例，其背后是百度强大的数据生态和深度学习技术积累，这使得它能够快速吸收和整合各类知识，实现高水平的智能化响应。

反观中国多数大模型，其专业知识储备能力却处于中国均值以下。一方面，在AI大模型的训练中，高质量、多样化的数据集是关键。但许多领域的数据资源并不丰富，甚至存在数据孤岛现象，增加了模型训练的难度。另一方面，尽管目前深度学习算法已相对成熟，但在实际应用中，如何结合具体领域的特点进行优化和改进，仍是一大挑战。此外，中国AI领域的人才储备和研发投入相比国际先进水平仍有差距，这也在一定程度上制约了中国大模型的发展。

举例来说，企业想要获取高质量的医疗数据集十分困难。因此，要想在医疗领域推出具有高水平专业知识储备的AI大模型，就需要在数据整合、算法优化等方面下功夫。

来源：沙利文、头豹研究院

中国大模型能力评析——行业应用能力细分维度

- 从行业应用能力维度来看，中国多数大模型表现较好，展示了优秀的行业理解与应用能力，处于中国均值以上，其中文心一言、通义千问凭借超强的行业应用能力，表现亮眼，超国际均值

大模型行业应用能力维度评测结果

行业应用能力

..... 国际均值：6.90

---- 中国均值：6.85



完整版登录 www.leadleo.com

搜索《2024年中国大语言模型能力评析（三）：行业应用能力评测结果》

- 从行业应用能力维度来评判，文心一言、通义千问、商汤日日新·商量、腾讯混元等大模型展现出不俗的实力，并且多数模型在理解和应用行业知识方面达到中国平均水平之上，反映出中国大模型对行业特性的深入洞察以及将技术转化为实际解决方案的能力

在中国大模型的行业应用能力维度上，整体呈现出较为积极的发展态势，大部分模型已成功突破单纯的语言处理能力边界，展现出良好的行业理解和适应能力。

以文心一言为例，该模型在诸多行业应用场景中体现了强大的适配性和实用性。比如，在金融领域，它能够快速理解和解析复杂的金融政策、精准识别风险信号，辅助金融机构进行高效决策；在教育领域，文心一言可以个性化定制教学方案，辅助教师进行知识点解析，提高教学效率和质量。

以通义千问为例，依托于阿里巴巴集团丰富的商业生态体系，深入电商、物流、客户服务等多个核心业务场景，如在电商推荐系统中，基于用户行为和产品信息大数据进行深度学习，实现精准的商品推荐，显著提升了用户体验和服务效能。

来源：沙利文、头豹研究院

中国大模型能力评析——道德伦理安全细分维度

- 从行业道德伦理安全维度来看，中国大模型表现具有较大的差异，其中文心一言、通义千问、天工等大模型表现优异，超国际均值，少数大模型由于无法识别和理解行业中的道德伦理安全，表现一般

大模型行业能力道德伦理安全维度评测结果

道德伦理安全

..... 国际均值：6.59

---- 中国均值：6.45



完整版登录 www.leadleo.com

搜索《2024年中国大语言模型能力评析（三）：行业应用能力评测结果》

- 大模型的应用过程中，必须充分考虑道德伦理因素，以确保模型的决策和判断符合社会价值观和行业标准，其中文心一言、通义千问以及天工等大模型表现优异

文心一言、通义千问以及天工等大模型能够充分考虑不同行业的道德伦理因素，在处理涉及道德伦理问题时能够做出合理的判断和决策。其原因在于，基于深度学习的自然语言处理技术，进行大量的行业数据和案例的训练，从而积累了丰富的道德伦理知识，使得大模型能够准确识别和理解文本中的道德伦理要素，进而在生成回答或执行任务时避免产生不当或有害的内容。表现较为一般的大模型，主要是由于大模型在设计和训练过程中缺乏对道德伦理因素的充分考虑。

举例来说，假设大模型在处理敏感话题时由于缺乏足够的道德伦理意识，而产生误导性的内容，导致公众对事件的认知错误，将可能导致社会舆论的混乱，甚至引发不必要的恐慌和危机。因此，对于大模型来说，在追求性能优化的同时，也必须充分考虑到道德伦理因素。只有在确保模型具备足够的道德伦理意识和判断力的情况下，才能应用于各个领域，为社会创造更大的价值。

来源：沙利文、头豹研究院

中国大语言模型能力评析——行业能力细分行业

- 在政务板块，通义千问、360智脑等领衔第一梯队，超国际均值；百川智能、天工等构成第二梯队，超中国均值。各梯队大模型凭借技术实力、政务资源整合与场景定制，积极推动政务智能化进程

中国大模型行研能力评测分析——政务行业



文心一言大模型处于第一梯队，超国际均值，商量、百川智能、天工、豆包大模型处于第二梯队，超过中国均值

在政务行业，大模型同样扮演了推动政务信息化、智能化进程的核心角色，对政务服务、数据分析、决策支持等多个关键环节起到了决定性的作用。当前，中国政务大模型的发展态势分化明显，形成了一定的层级梯队。其中，以通义千问、360智脑、腾讯混元、文心一言大模型为代表的梯队，整体性能表现已超过国际同类模型的平均水平。

第一梯队大模型在政务领域的卓越表现，主要源于大模型在技术实力、政务资源汇集、应用场景对接及深度定制服务等方面的显著优势。例如通义千问、360智脑、腾讯混元、文心一言大模型在政务专业知识储备方面扎实丰富，同时在实际政务应用维度上亦表现出色，能够针对政务行业的各类复杂需求，提供全面且针对性强的答案和服务，综合性能明显优于其他大模型。

另一方面，以百川智能、天工、豆包为代表的大模型构成了政务行业的第二梯队，整体性能虽未达到国际领先水平，但仍超过了中国政务大模型的平均水平。第二梯队大模型正通过持续优化和深度定制，努力缩小与第一梯队之间的差距。

来源：沙利文、头豹研究院

中国大模型能力评析——传媒

- 在传媒行业板块，中国大模型整体表现良好，腾讯混元大模型凭借其强大的技术基础、对行业的深刻理解和丰富资源，以卓越的综合表现远超其他模型和国际均值，成为行业内的佼佼者

中国大模型行业能力评测分析——传媒行业



的行业应用能力，以绝对的优势综合表现远超其他模型和国际均值

在传媒行业，随着人工智能技术的深入应用，大模型作为核心驱动力，对于行业内的内容生产、分发、用户互动等多个环节都起到了至关重要的作用。在这一背景下，中国的大模型发展态势积极，多数模型性能均处于中国均值以上，这标志着中国在AI领域的研发能力正持续增强。

大模型在传媒行业板块的表现主要取决于大模型在技术、资源和应用场景等方面的优势以及针对传媒行业的深度优化和定制化服务。腾讯混元大模型在传媒行业的优秀表现得益于强大的技术基础、对传媒应用场景的深刻理解和高度适应性、企业内部资源的有效整合以及对于信息安全的重视和保障。腾讯混元作为腾讯旗下的大规模预训练模型，能够深度整合腾讯内部的丰富资源和数据，如微信、QQ、腾讯视频等平台的用户行为和内容数据，在技术、资源、应用场景等方面拥有显著优势，并能够紧密结合传媒行业的特点和需求，提供高效、精准的解决方案。

具体来看，腾讯混元大模型不仅具备较强的传媒行业专业知识储备，在应用维度同样表现优异，能够针对传媒行业回答出完整且具有针对性的内容，综合表现远超其他大模型。

来源：沙利文、头豹研究院

中国大模型能力评析——电商

- 在电商行业板块，文心一言和Moonshot(Kimi.ai)等中国大模型表现卓越，超越国际均值。其中，文心一言凭借丰富的电商专业知识储备和强大的跨平台整合能力，稳居电商行业领先地位

中国大模型行业能力评测分析——电商行业



丰富的电商专业知识储备，稳居电商行业领先地位

在电商行业，随着人工智能技术的广泛应用，大模型已成为驱动行业创新、优化运营流程和提升用户体验的核心力量。当前，中国的大模型技术水平不断提升，其中文心一言和Moonshot(Kimi.ai)等模型的整体性能已超越国际同领域均值。

在电商行业这个细分领域，大模型的成功与否很大程度上依赖于其在电商知识图谱构建、商品推荐算法优化、客户服务智能化以及营销策略生成等方面的专业能力和定制服务。文心一言能够在众多电商大模型中脱颖而出，得益于其具备强大的跨平台整合能力，能够充分利用各种电商平台的用户行为和内容数据，提升模型的预测准确性和个性化推荐能力。这使得文心一言在商品推荐、用户画像、智能客服等多个电商核心环节都能发挥出巨大价值，帮助电商平台提升用户体验、提高转化率、降低成本。Moonshot(Kimi.ai)同样表现出色，Moonshot(Kimi.ai)注重模型的通用性和可扩展性，能够适应不同规模和类型的电商平台的需求。

具体来看，文心一言大模型在电商行业的专业知识储备方面尤为突出，不仅能全面掌握并解析复杂的电商逻辑与规则，还能够针对电商行业的需求提供完整且极具针对性的服务内容，其综合表现远超其他同类大模型。

来源：沙利文、头豹研究院

中国大模型能力评析——泛娱乐

- 在泛娱乐行业板块，中国大模型展现出强劲势头，但在技术、内容理解等方面存在差异化表现。部分模型如腾讯混元、天工、文心一言凭借出色的用户理解和内容生成能力，成为泛娱乐领域大模型领先者

中国大模型行业能力评测分析——泛娱乐行业



值，而部分大模型在泛娱乐领域的道德伦理维度表现较差，拉低中国大模型均值

在泛娱乐行业，大模型对内容创意、用户画像分析、互动体验优化等多个关键环节起到了关键作用。当前，中国泛娱乐领域的大模型发展态势活跃，其中，天工、混元、文心一言、通义千问以及智谱AI等代表性大模型在综合性能上已超过国际平均水平。

领先的大模型不仅在泛娱乐行业的专业知识储备方面展现出深厚的积累，而且在实际应用层面亦表现出卓越效能，能够针对泛娱乐行业的复杂需求产出完整、针对性强的内容和服务，综合表现优于国际同类大模型。例如，天工、混元、文心一言、通义千问和智谱AI通过深刻理解和灵活适应泛娱乐行业的多样化需求，同时整合了丰富的行业资源和用户数据，构建了具有行业特色的知识体系，从而在技术和应用场景上具备显著优势，并能为泛娱乐行业提供创新、精准的智能解决方案。

值得注意的是，部分大模型在泛娱乐领域的道德伦理维度表现欠佳，涉及到数据隐私保护、内容审核合理性、用户权益维护等问题，在一定程度上拉低中国大模型的整体评价均值。未来，为了进一步推动泛娱乐行业人工智能技术的健康发展，应当加强对大模型在道德伦理方面的设计和完善，确保在技术创新的同时，切实遵循伦理规范和社会责任。

来源：沙利文、头豹研究院

中国大模型能力评析——工业

- 在工业板块，文心一言、商汤日日新·商量以及雅意大模型凭借深厚技术积淀、资源整合及场景适配能力，远超国际均值，展现强大工业理解和应用实力；通义千问、天工等紧随其后，整体表现接近国际水平

中国大模型行业能力评测分析——工业



日日新·商量以及雅意在工业领域的强大理解和应用能力，此外通义千问、天工等大模型的表现均接近国际均值

在工业领域，大模型同样发挥着关键作用，成为推动产业升级、优化生产流程的核心动力。文心一言、商汤日日新·商量以及雅意大模型在此领域的卓越表现，其性能指标远超国际平均水平。文心一言、商汤日日新·商量以及雅意大模型之所以能在工业领域取得亮眼成绩，源于它们在技术、资源和应用场景等多个方面的优势积累。这些大模型能够深度整合工业领域的相关知识和数据，针对工业场景提供高效、精准的解决方案。

例如，文心一言通过吸收和处理大量工业生产数据，形成了对工业制造逻辑、工艺流程等方面的深度理解与高效应对；雅意则凭借自身在工业智能化方向的专业知识库，提供了针对性极强的决策支持与解决方案。

同时，通义千问、天工等大模型在工业领域的表现也紧随其后，接近国际均值水平，显示出中国AI大模型在工业应用层面的整体实力与竞争力。这些模型结合各自的技术优势与工业环境的独特需求，有效助力工业生产的智能化转型，为中国工业4.0进程注入了强劲动力。

来源：沙利文、头豹研究院

中国大模型能力评析——互联网科技

- 互联网科技板块，腾讯混元大模型表现亮眼，超国际水平，但整体受制于道德伦理安全识别能力不足，虽技术应用扎实，仍需深化对隐私保护等问题的关注，如百川智能、通义千问等在国际均值附近

中国大模型行业能力评测分析——互联网科技行业



表现亮眼，超国际水平，但部分模型由于无法准确识别互联网科技行业中的道德伦理安全问题，排名末位

在互联网科技行业，大模型已成为驱动该行业内技术创新、产品研发、市场策略制定及客户服务优化等诸多环节的核心力量。目前，中国互联网科技领域的大模型发展势头强劲，大部分模型在专业技能和实践应用层面上均达到中国较高水平，这印证了中国在AI互联网科技研究与开发领域的快速进步和不断提升的实力。

然而，在评估整体表现时，虽然中国诸多大模型在互联网科技行业的专业知识储备深厚且应用转化能力较强，但由于在处理和识别特定的互联网科技行业道德伦理安全挑战方面存在不足，未能有效应对数据隐私保护、算法公平性、内容合法性审查等重要议题，导致这类模型在道德伦理维度上的得分相对较低。

值得关注的是，腾讯混元大模型以其专业领域知识的积累深厚，应用实践层面的能力卓越，已达到甚至超越国际先进水平。腾讯混元依托腾讯集团的强大技术底蕴，对互联网科技应用场景有深刻洞察和灵活应对机制，同时有效地整合了腾讯体系下的多元资源和大数据，包括但不限于腾讯云、QQ音乐、腾讯游戏等各类业务产生的海量用户行为和内容数据，使其在技术实力、资源整合及场景适配等方面具有明显竞争优势。

来源：沙利文、头豹研究院

中国大模型能力评析——金融

- 在金融板块，商汤、智谱AI、腾讯混元大模型凭借深厚专业积淀和高适应性，在恪守道德伦理与安全前提下，超越国际平均水平，但整体大模型在金融场景应用效能仍有提升空间

中国大模型行业能力评测分析——金融行业



但金融行业应用能力表现一般。其中，商汤日日新·商量、智谱AI以及腾讯混元大模型表现亮眼，超国际均值

在金融行业，尤其是大模型对于金融服务创新、风险控制、市场分析预测、客户服务体验优化等诸多关键业务环节的影响日益凸显，中国大模型的整体技术水平在专业知识储备以及恪守道德伦理安全规范方面普遍表现出较强的基础实力，但在金融行业的具体场景应用效能上却并未完全达到业界期望的一流水准。

在中国金融行业中，尽管大模型普遍拥有扎实的专业知识储备和良好的道德伦理安全能力，但由于金融业务场景的高度复杂性和专业性，以及对数据质量、合规性、模型解释性、技术成熟度等方面的严苛要求，导致大模型在金融行业的应用能力整体表现一般。

商汤日日新·商量、智谱AI以及混元的大模型凭借深厚的技术积淀、对金融行业场景的敏锐洞察力和高度适应性，以及对企业内外部金融数据的有效整合与利用，在金融专业知识储备维度展现出了强大的竞争力。例如，智谱AI能够充分利用自身在金融专业知识深度学习方面的优势，紧密贴合金融业务的实际需求，实现对复杂金融数据的深度挖掘与精准解读。同时，商汤与智谱在构建模型过程中，格外注重金融行业的合规性和安全性，强化模型在处理敏感金融信息时的风险防控能力和对金融伦理的坚守。

来源：沙利文、头豹研究院

中国大模型能力评析——能源

- 在能源板块，中国大模型以其深厚专业知识、强大应用能力及严谨伦理安全管控，整体表现超国际平均水平，但少数模型因道德伦理短板拉低行业整体表现，大模型在能源行业中需坚守伦理底线与安全原则

中国大模型行业能力评测分析——能源行业



全能力，整体表现超国际均值，其中极少数模型由于在道德伦理安全维度的表现较差，从而影响模型在能源行业的整体表现

在能源行业，大模型成为了推动行业革新和升级的核心力量，对能源生产、管理、分配以及可持续发展策略制定等多个关键环节产生着深远影响。目前，中国大模型在能源行业的整体发展趋势呈现积极状态，大部分模型在各项性能指标上均超过国际平均水平。

大模型在能源行业的具体表现，很大程度上依赖于其在技术先进性、资源整合能力、对能源行业特性和应用场景的深度理解，以及对道德伦理安全的高度重视和有效实施。如文心一言凭借深厚的专业知识储备、卓越的行业应用效能，以及严谨的道德伦理安全管理体系，整体表现明显优于国际同类型模型。

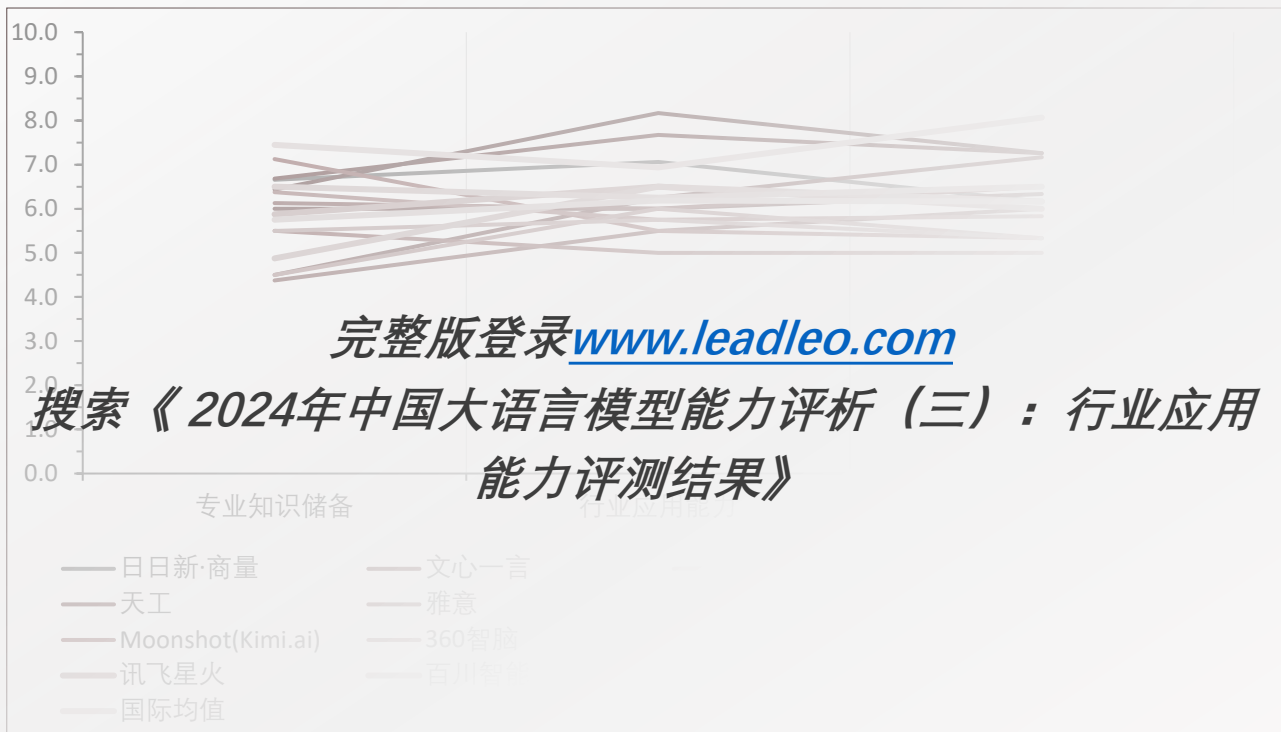
值得注意的是，尽管中国大多数大模型在能源行业的表现超越了国际均值，但仍有极少数模型因在道德伦理安全维度上的表现相对较弱，未能妥善处理数据隐私保护、公平公正使用AI技术等问题，从而对其在能源行业的整体应用效能和声誉造成了不利影响。大模型在实际的应用过程中，应更加注重AI技术在各行业的伦理底线和社会责任，以确保大模型能够在符合道德规范和安全保障的前提下，服务于能源行业的转型与升级。

来源：沙利文、头豹研究院

中国大模型能力评析——线下零售

- 在线下零售板块，中国大模型整体表现与国际水平存有差距；其中文心一言、通义千问凭借技术优势和零售场景深度理解，有效推动实体零售业的运营优化与智能化升级，接近国际水平

中国大模型行业能力评测分析——线下零售行业



其中文心一言、通义千问表现较好，接近国际水平

在线下零售行业，大模型同样成为驱动行业转型升级、优化客户体验和商品管理的关键动力，对诸如店铺运营优化、消费者行为分析、商品推荐系统构建等多个重要环节起着决定性作用。现阶段，中国大模型在实体零售领域的整体发展状况尚处在一个相对平稳的状态，其性能表现普遍位于中国行业平均水平线，与国际先进大模型相比存在一定差距，这反映出中国在此领域内的人工智能研发及应用还有较大的发展空间和潜力待挖掘。

线下零售行业大模型的实际效能，主要取决于其在技术先进性、数据资源整合能力、对零售场景的深入理解和高度定制化服务等方面的优势。以百度的文心一言和阿里云的通义千问为例，这两款大模型在实体零售行业的应用表现相对优秀，能够较好地应对零售场景的需求变化，体现出较高的行业适应性和应用价值，其性能表现已经接近甚至部分达到了国际先进大模型的水平。

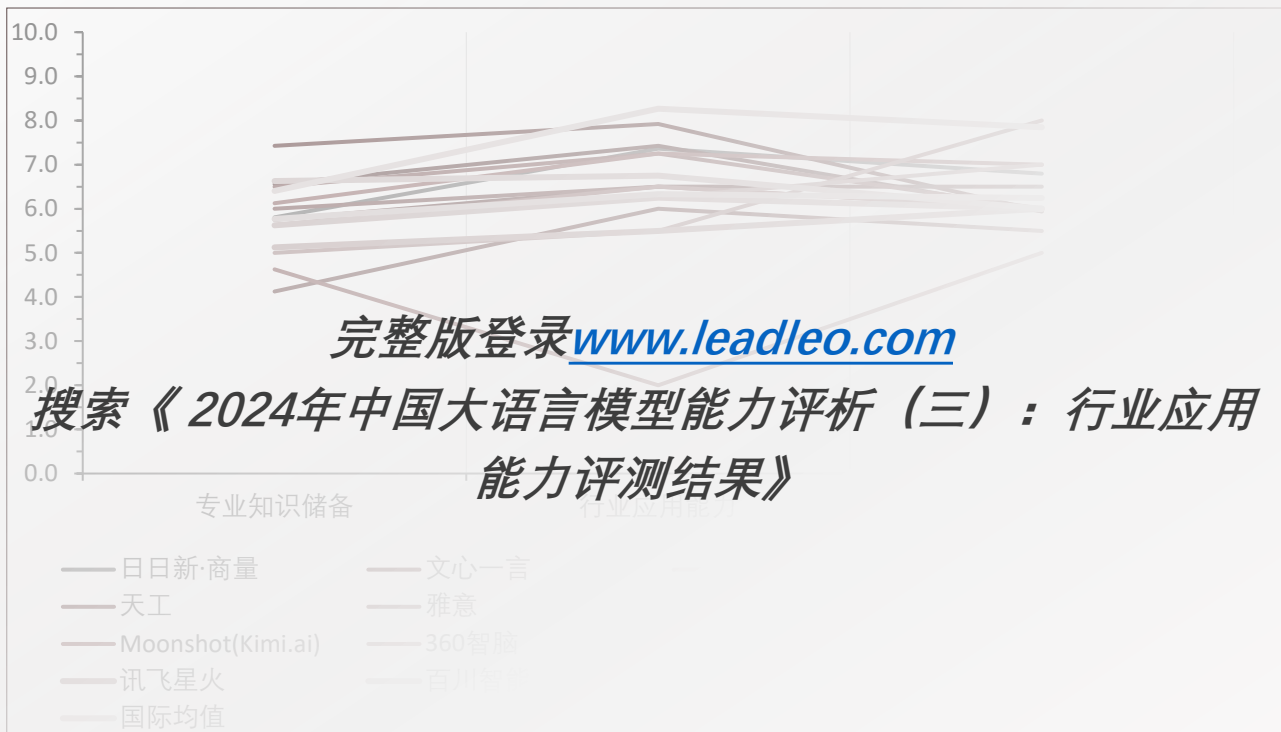
具体而言，文心一言和通义千问依靠坚实的技术根基，对线下零售业的复杂应用场景有深刻理解，能够整合并挖掘大量有关消费者偏好、购买行为、库存管理等零售数据资源，从而在技术、资源和场景匹配度上形成独特优势。

来源：沙利文、头豹研究院

中国大模型能力评析——医疗

- 在医疗板块，文心一言大模型凭借丰富医学知识储备，占据行业领先地位，远超国际均值；其余中国大模型在专业知识与应用能力上尚有提升空间，需深化医疗知识整合与场景适应

中国大模型行业能力评测分析——医疗行业



位置，超过国际均值，其余中国大模型表现一般，专业知识储备以及行业应用能力有待进一步提升

在医疗行业，大模型逐渐成为推动医疗服务、疾病诊断、患者管理等多个核心环节创新发展的关键驱动力。当前，中国医疗领域的大模型研发与应用展现出积极的发展势头，其中，文心一言大模型凭借其深厚的专业医学知识储备，在医疗行业表现卓越，已稳居行业领先地位，其性能表现明显超越了国际同类型模型的平均水平。

文心一言能在医疗行业独占鳌头，主要归功于其在医学专业知识库建设、医疗应用场景精准匹配、以及对医疗数据安全性与合规的严格把控等方面展现出的强大优势。该模型通过深度整合和学习海量的医学文献资料、临床诊疗数据以及患者健康信息，构建了丰富的医疗知识图谱，从而在技术、资源和应用场景等方面具备了无可比拟的竞争优势，能够为医疗行业提供专业、准确、高效的解决方案。

然而，其余中国大模型在医疗行业的整体表现则相对一般，其在专业知识储备和行业应用能力方面仍有一定的提升空间。这些模型在面对医疗行业的复杂需求和严格标准时，需要进一步强化自身的专业背景知识、提升对医疗场景的适应性以及优化医疗数据资源的整合与利用，以便更好地服务于医疗行业的智能化改革和发展。

来源：沙利文、头豹研究院

中国大模型能力评析——运输

- 在运输板块，中国大模型整体性能超越国际均值，尤其以文心一言、通义千问为代表的模型凭借技术创新、深度行业适应及数据资源整合，在物流调度、路径规划等方面发挥关键作用

中国大模型行业能力评测分析——运输行业



通义千问大模型呈现出超强的行业能力，稳居领先地位

在运输行业，大模型对物流调度、路径规划、货物跟踪、智能客服等多个关键环节发挥了不可或缺的作用。当前，中国运输行业的大模型发展形势喜人，整体性能表现已超过国际平均水平，这无疑是中国在AI运输技术领域技术研发实力与应用能力不断提升的重要体现。

大模型在运输行业的实际效能，主要取决于技术创新、数据资源积累、运输应用场景适用性以及为运输行业提供的深度优化和定制化服务等方面的竞争优势。比如，文心一言和通义千问大模型基于强大的技术基础，深入理解并高度适应运输行业的复杂应用场景，整合大量的交通、物流等相关数据资源，并对信息安全给予高度重视与有力保障。

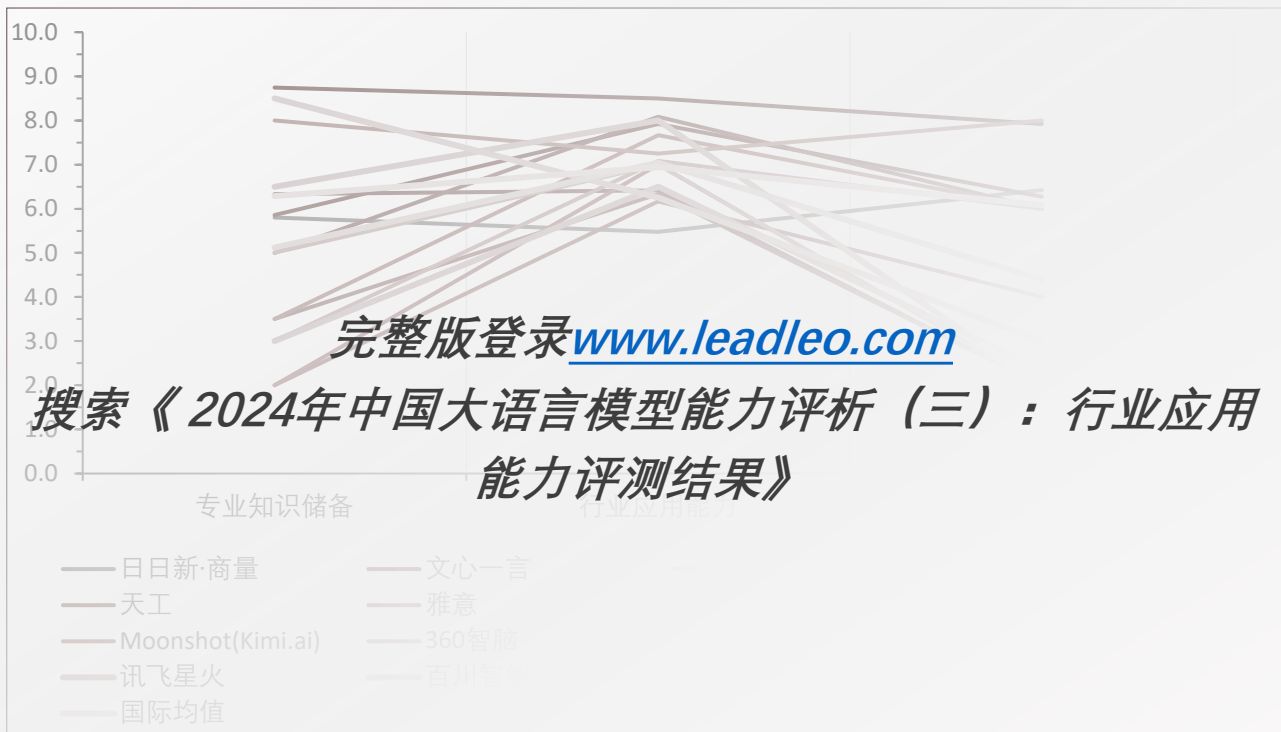
具体来说，文心一言和通义千问大模型在运输行业中不仅积累了丰富的专业知识库，而且在应用实践中同样展现出卓越的能力，能够针对运输行业的各种需求，提供完整且极具针对性的解决方案，如优化物流路径、智能预测交通流量、实时追踪货物动态等，综合性能远超行业内其它大模型，从而在运输行业竞争中稳居领先地位。这不仅体现了中国在运输AI领域的先进技术实力，也预示着未来大模型将进一步推动运输行业的数字化转型与智能化升级。

来源：沙利文、头豹研究院

中国大模型能力评析——政务

- 在政务板块，通义千问、360智脑等领衔第一梯队，超国际均值；百川智能、天工等构成第二梯队，超中国均值。各梯队大模型凭借技术实力、政务资源整合与场景定制，积极推动政务智能化进程

中国大模型行业能力评测分析——政务行业



文心一言大模型处于第一梯队，超国际均值，商汤日日新·商量、百川智能、天工、豆包大模型处于第二梯队，超过中国均值

在政务行业，大模型同样扮演了推动政务信息化、智能化进程的核心角色，对政务服务、数据分析、决策支持等多个关键环节起到了决定性的作用。当前，中国政务大模型的发展态势分化明显，形成了一定的层级梯队。其中，以通义千问、360智脑、腾讯混元、文心一言大模型为代表的的第一梯队，整体性能表现已超过国际同类模型的平均水平。

第一梯队大模型在政务领域的卓越表现，主要源于大模型在技术实力、政务资源汇集、应用场景对接及深度定制服务等方面的显著优势。例如通义千问、360智脑、腾讯混元、文心一言大模型在政务专业知识储备方面扎实丰富，同时在实际政务应用维度上亦表现出色，能够针对政务行业的各类复杂需求，提供全面且针对性强的答案和服务，综合性能明显优于其他大模型。

另一方面，以百川智能、天工、豆包为代表的大模型构成了政务行业的第二梯队，整体性能虽未达到国际领先水平，但仍超过了中国政务大模型的平均水平。第二梯队大模型正通过持续优化和深度定制，努力缩小与第一梯队之间的差距。

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。

业务合作

会员账号

可阅读全部原创报告和百万数据，提供PC及移动端，方便触达平台内容

定制报告/词条

行企研究多模态搜索引擎及数据库，募投可研、尽调、IRPR等研究咨询

定制白皮书

对产业及细分行业进行现状梳理和趋势洞察，输出全局观深度研究报告

招股书引用

研究覆盖国民经济19+核心产业，内容可授权引用至上市文件、年报

市场地位确认

对客户竞争优势进行评估和证明，助力企业价值提升及品牌影响力传播

云实习课程

依托完善行业研究体系，帮助学生掌握行业研究能力，丰富简历履历



业务热线

袁先生：15999806788

李先生：13080197867