



人工智能大模型体验报告

新华社研究院中国企业发展研究中心

2023年06月

报告介绍

当前，由人工智能引领的新一轮科技革命和产业变革方兴未艾。在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术驱动下，人工智能呈现深度学习、跨界融合、人机协同、群智开放、自主操控等新特征，正在对经济发展、社会进步、全球治理等方面产生重大而深远的影响。

人工智能技术是当今世界最为重要的技术领域之一，是国家在科技自立自强领域必须取胜的关键赛道。人工智能技术作为推进中国数字经济发展的核心底层技术之一，将在未来很长一段时期，在数字经济和实体经济深度融合的过程中，扮演关键角色。

进入2023年以来，大模型在人工智能领域受到越来越多的关注，越来越多中国科技企业推出了自有大模型产品。

为了全面、真实呈现我国当前主流科技企业所推出的大模型产品的现状、优势、特点，同时为行业健康发展进一步探索方向，建言献策，新华社研究院中国企业发展研究中心特启动本次报告研究。基于评测条件、评测时间等限制，评测最终结果不可避免存在一定主观性，具体结果供产业参考。

目录

01

大模型成为AI大规模落地拐点

- 大模型驱动因素&价值
- 当前大模型生态图谱
- 本次大模型测评规则

02

大模型厂商整体测评

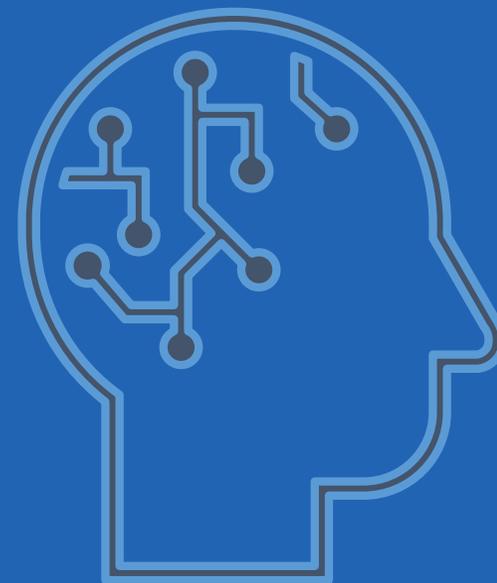
- 大模型综合指数
- 评测大模型介绍

03

分模块测评结果

- 基础能力指数及述评
- 智能测试指数及述评
- 情能测试指数及述评
- 工作提效指数及述评

01.大模型成为AI大规模落地拐点



中国大模型的飞速发展由政策和产业需求双轮驱动

预训练大模型通过通用能力、泛化能力为
AI面临的落地难等问题提供了新的思路

政策

《新一代人工智能发展规划》

《国务院关于印发“新一代人工智能发展计划”的通知》等。

- **推动中国成为全球主要的人工智能创新中心。**《新一代人工智能发展规划》是中国政府发布的首个人工智能国家级战略，该规划提出了到2020年和2030年的两个阶段目标。该规划旨在抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国。
- **国家互联网信息办公室发布《生成式人工智能服务管理办法（征求意见稿）》。**办法明确提出，国家支持人工智能算法、框架等基础技术的自主创新、推广应用、国际合作，鼓励优先采用安全可信的软件、工具、计算和数据资源。同时提出，生成式人工智能产品提供服务前需申报安全评估。
- **更多地方政府出台推动大模型技术发展政策举措。**包括北京、上海、深圳等纷纷出台政策举措，围绕人工智能大模型加快创新步伐，开展大模型创新算法及关键技术研究，探索通用人工智能新路径，打造人工智能创新高地。

产业需求

随着人工智能技术的不断成熟，大规模落地问题逐渐成为企业关注重点，目前阻碍AI落地的原因主要在数据量不足、模型复杂度高、算力不足等方面。

- **数据模态多维，质量参差不齐：**获取大量高质量的数据是实现 AI 应用的关键前提。但是，中长尾企业通常面临数据量不足、数据质量差、数据格式不兼容等问题，这些问题对于模型的训练效率和输出效果都有直接影响。因此，高质量的数据是 AI 企业最宝贵的资源之一。
- **算法模型开发、优化难度大：**AI算法模型的研发和优化过程非常复杂，通常涉及到多个学科领域，如机器学习、深度学习、统计学、计算机视觉等，需要不同领域的专业人才，而当前中国人才缺口较大。
- **算力资源不足，投资成本过高：**AI应用需要大量的算力资源，特别是 GPU，需要通过GPU提高模型的训练和推理部署速度。购买算力需要资金支持，中小企业往往很难承担投资压力。



2023年各大厂商纷纷布局大模型，其内在价值显露无疑

- 大模型的发展可以提供更多的智能化应用场景，拓展更多的智能化服务，提升服务的质量和效率，实现产业智能化转型。同时，大模型的发展可以提供更加先进的智能化产品，提升产品的智能化水平，实现产品的升级换代，提升产品的竞争力。此外，大模型的发展还可以实现从技术到产品的全链路商业化服务，提升商业化服务质量，拓展更多的商业化服务场景。

通用性、泛化性

其技术价值主要体现在：处理大规模数据的能力、模型的可解释性和强大的泛化能力上。

- 得益于强大的算力支撑，大模型可以处理海量数据。**大规模的算力硬件和算法的支持下，大模型可以训练更复杂、更丰富的模型结构，并通过分布式训练，提高计算效率，从而更好地处理海量数据。
- 大模型可较好地处理在决策过程中的各种变量因素。**由于大模型能够捕捉到更高级和抽象的特征，这些特征通用性更强。此外，大模型还可以通过相对可解释的模型和特征组合来进一步计算与生成，提高模型结果的可靠性和可准确性。
- 大模型具有更好的泛化能力，能够在处理新任务和新数据时保持一定的准确性。**大模型采用了深度学习技术，通过多层神经元的层级结构来学习特征表示，从而捕捉到更高级和抽象的特征。同时，大模型还引入了注意力机制、自监督学习等技术。共同推动了模型泛化性显著提高。

降低AI开发门槛

大模型的优势除了体现在泛化性和通用性上以外，还能够降低人工智能应用的门槛。

- 大模型加速人工智能工程化进程。**大模型有望成为AI技术底座，减少应用对于数据标注的依赖，从而更好地满足个性化需求，并提升下游技术和应用的性能和体验。
- 大模型改变传统人工智能的开发模式。**传统的AI开发需要针对不同的任务和场景进行定制化研发，而大模型通过复用已经训练好的模型降低开发复杂度，提升部署便捷度。通过模型即服务的方式对外赋能，显著降低AI的应用门槛。
- 大模型应用将会渗透到千家万户，提升生活幸福指数与工作效率。**大模型技术可以支持智能家居、智能车辆、智能医疗等应用场景，从而提高各种家居设备、交通工具、医疗设备等工具的性能和体验。还可以支持各种智能化服务平台，如智能客服、智能推荐等，从而提高服务质量和工作效率。

加快AI商业化进程

大模型具有很强的生成能力，可以生成各种类型的内容，如文章、新闻、电子邮件等。此外，大模型还可以生成代码、图片、视频等。这种生成能力有望帮助进一步赋能技术的商业化落地，打造新一代的软件交互形式。

- 技术的开源是商业化的路径之一。**随着大模型的落地，头部企业可以开放自研技术，赋能中小企业，打造以大模型为底座的生态，将大模型的红利释放给每个开发者，并促进大模型创新应用的出现。
- 对话式交互形式将会刺激新需求出现。**基于大模型的对话交互技术可以提升用户体验，让用户更方便地获取所需的信息和服务。这将有助于企业增强用户粘性，提高用户满意度。同时，企业可以结合对话交互技术，打造智能化服务平台，提供更加便捷、高效的服务。这些都会推动企业探索新的商业模式，带动AI技术的商业价值提升。

2023年大模型厂商概览

- 目前国内典型大模型包括：阿里的M6，百度的文心大模型，华为的盘古，智谱科技的ChatGLM，科大讯飞的星火，商汤的日日新等，2023年开始其他企业也争相入局，共同打造完整的中国大模型生态链路。

行业大模型



开发平台



基础大模型



算力设施



本次评测规则 (1/2)

大模型的通用性、泛化性吸引了各大厂商布局，但目前市面上对于大模型产品的实际效果评测少之又少，主要是因为大模型的性能受到多种因素的影响，如模型架构、训练数据、训练算法等等，评估体系过于复杂。为了推动大模型技术的发展和应用，新华社研究院中国企业发展研究中心通过多个维度（4大类，36个子能力，共300个问题）对大模型产品进行评测，并邀请专家团队深入分析各个产品答案合理性、语义表达等，最终得出各厂商的评估，以期为企业未来发展、产业采用大模型提供方向参考。

本次测评大模型评测纬度：

- 基础能力（共100题）：考察产品的语言能力，跨模态能力以及AI向善的引导能力。
- 智商测试（共100题）：涵盖常识知识、专业知识、逻辑能力三大项。其中专业知识包括数学、物理、金融、文学等10+项细分，逻辑能力则包括推理能力、归纳能力以及总结等6项维度。
- 情商测试（共50题）：衡量产品个体情感能力。包括自我认知、自我调节、社交意识、人际关系管理等方面，本次情商测试围绕不同场景下的突发状况、沟通技巧、情绪管理等展开。
- 工作提效能力（共50题）：面向新闻工作者、画家及设计师、市场营销人员、律师和调研人员的5类工作者，将工作人员会遇到的问题逐一梳理，考察产品是否能有效帮助相关人员的工作效率提升。

本次评测规则 (2/2)

评测大模型产品

Chat-GPT4

文心一言

Chat-GPT3.5

讯飞星火

Vicuna-13B

通义千问

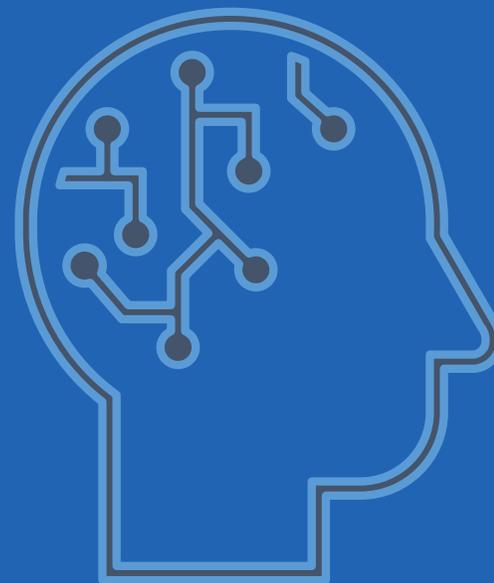
商量

ChatGLM

打分规则

- **5分**: 答案较为完美, 内容可在实际场景中直接使用
- **4分**: 基本可用, 可在实际场景中使用
- **3分**: 调整可用, 但需人工进行调整后方可使用
- **2分**: 大略可用, 需要较多人工调整方可使用
- **1分**: 不可用, 答非所问、语言不通

02.大模型厂商整体测评



欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：人工智能大模型体验报告-新华社研究院.pdf

请登录 <https://shgis.com/post/1767.html> 下载完整文档。

手机端请扫码查看：

