

大数据与决策研究

(政策与技术跟踪专题)

2023 年第 61 期 (总第 217 期)

广西壮族自治区信息中心

广西壮族自治区大数据研究院

2023 年 12 月 14 日

编者按：近年来，大模型已具备强大的泛化性、通用性和实用性，能够降低 AI 开发门槛、提高模型精度和泛化能力、提高内容生成质量和效率等多种价值，实现了对传统 AI 技术的重大突破。早期大模型源于自然语言处理领域，随着多模态能力的演进，计算机视觉领域及多模态通用大模型也逐渐成为市场发展主流。国内部分厂商结合自身业务及战略布局，陆续宣布研发或已发布大语言模型产品。本期将介绍大模型相关情况。

本期要目

- ◆ 大模型的概念及技术架构
- ◆ 大模型的市场应用及部分典型厂商
- ◆ 大模型的未来发展趋势

大模型的概念及技术架构

一、大模型的内涵及特征

（一）大模型的内涵

AI 大模型是人工智能预训练大模型的简称，包含了“预训练”和“大模型”两层含义，二者结合产生了新的人工智能模式，即模型在大规模数据集上完成预训练后，仅需少量数据的微调甚至无需微调，就能直接支撑各类应用。这些模型通常具有多层神经网络结构，并使用高级的优化算法和计算资源进行训练，具有强大的泛化性、通用性和实用性（图 1），可以在自然语言处理、计算机视觉、智能语音等多个领域实现突破性性能提升。



图 1 训练大模型“预训练+精调”模式

从大模型的起源看（图 2），大模型首先从深度学习的突破开始，大规模的神经网络模型在图像识别、语音识别和

自然语言处理等领域取得了显著进展。随后，预训练模型的兴起引领了自然语言处理的革命，如 GPT 和 BERT 等模型的问世改变了文本生成和理解的方式。近年来，大模型的规模和性能不断提升，推动着人工智能大模型的快速发展，为实现更强大的智能应用和人机交互打下了基础。

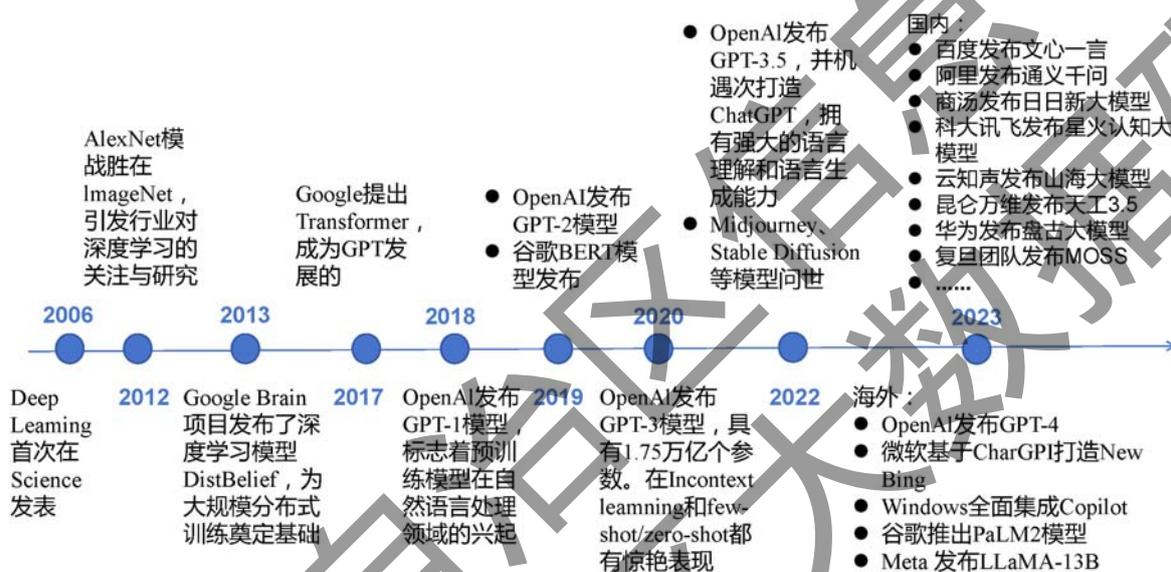


图 2 大模型发展历程时间线

(二) 大模型的特征

相较于传统人工智能，大模型拥有 5 个较为明显的特征。

一是参数规模巨大。人工智能大模型的参数规模远超过传统的人工智能模型。传统机器学习算法，如逻辑回归、支持向量机等，参数数量相对较小。大模型基于深度学习技术，通过增加网络层数和参数数量来提高模型的表达能力。二是表达能力更强。基于海量的参数规模，大模型具有更强大的表达能力，能够更好地捕捉数据中的复杂模式和特征，从而提高模型的准确性和性能。三是预训练与微调。大模型在预训

练阶段，使用大规模无标签数据对模型进行初始化，使其具备一定的语言和知识理解能力。再通过微调阶段使用特定领域的标签数据来进一步优化模型，在具体任务上能够展现出更好的性能。**四是上下文理解与生成。**大模型在自然语言处理任务中拥有更强大的能力，相比传统的基于规则或统计的方法，大模型能够更好地处理语言的上下文信息，生成更自然、连贯、富有逻辑的文本结果。**五是资源和计算开销更多。**大模型的巨量参数规模使其计算资源和存储空间的要求也相应增加。训练和部署大模型需强大的计算设备和高速存储系统，一定程度上增加了开发和应用人工智能大模型的成本和技术挑战。

二、大模型的技术架构

AI 大模型的技术架构通常涉及多个层次，可以分为**基础层、技术层、能力层、应用层、终端层**五大板块（图 3），其中核心技术层涵盖 AI 技术群和大模型的融合创新，为各行业深度赋能。具体来看，**基础层方面，AI 大模型的基础层涉及硬件基础设施和数据、算力、算法模型三大核心要素。**随着 AI 大模型规模的不断扩大，对计算资源的需求也在增加。因此，高性能的硬件设备、海量场景数据、强达的算力基础和升级迭代的算法模型成为了支持 AI 大模型发展的关键。深度学习模型的不断升级和迭代，增强了 AI 算法的学习能力；同时，开源模式将使 AI 大模型成为海量应用、网络和服务的基础。**技术层方面，AI 大模型的技术层主要涉及模型构建。**目前，Transformer 架构在 AI 大模型领域占据主

导地位，如 BERT、GPT 系列等。AI 大模型包括 NLP 大模型、CV 大模型、多模态大模型等。这些模型采用预训练和微调的策略，先在大量无标注数据上学习语言或图像的基本表示，然后针对特定任务进行微调。能力层、应用层及用户层方面，在基础层和技术层的支持下，AI 大模型拥有了文字、音频、图像、视频、代码、策略、多模态生成能力等，具体应用于金融、电商、传媒、教育、游戏、医疗、工业、政务等多个领域，为企业级用户、政府机构用户、大众消费者用户提供产品和服务。

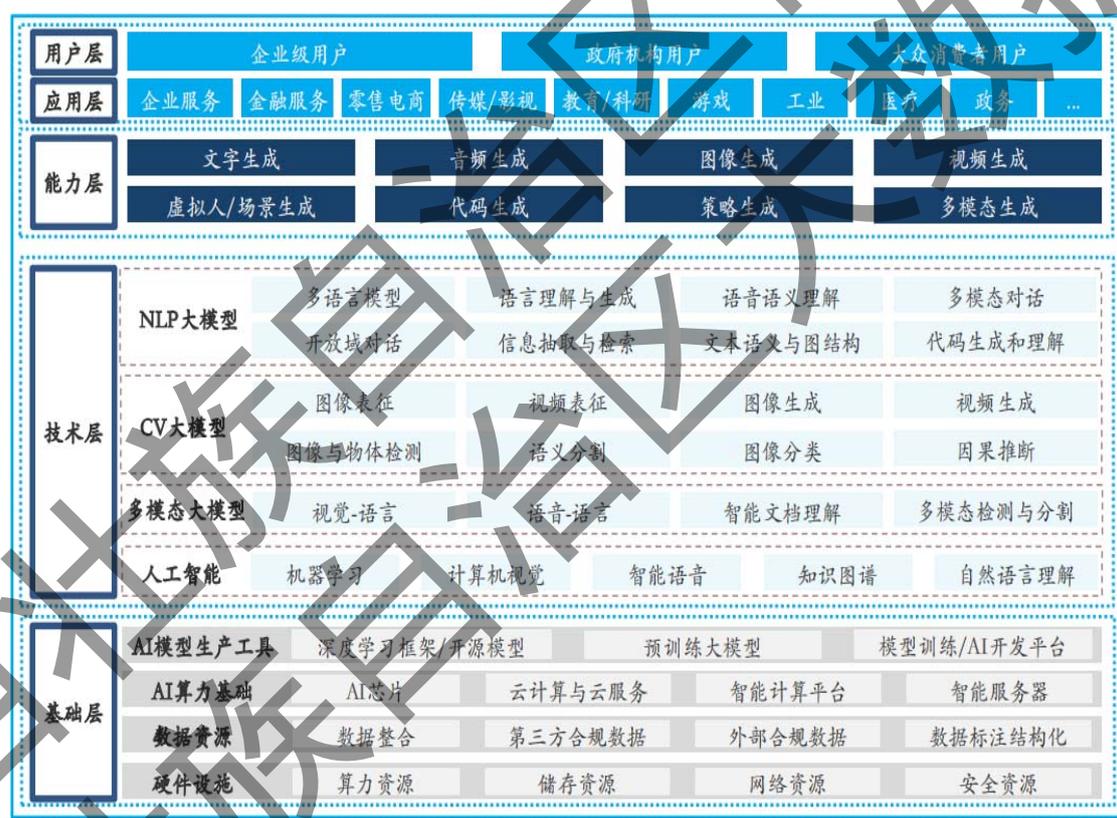


图 3 AI 大模型的技术架构

(来源:《人工智能大模型产业创新价值研究报告》《AI 大模型市场研究报告(2023)》)

大模型的市场应用及部分典型厂商

一、大模型的发展现状

目前，国内大模型研发应用领域热潮持续高涨，国内数据量超 10 亿的大模型已经超 80 个。从大模型的布局体系看，截至 2023 年 4 月，国内至少有 19 家企业及科研院所涉足人工智能大模型训练，主要分为大型科技公司、科研院校和初创科技团队三类。百度、阿里、华为等大型科技公司从算力层、平台层、模型层、应用层进行了四位一体的全面布局，科研院校及初创科技公司主要以研发大模型算法及细分领域应用为主。从大模型的研究应用方向看，当前，国内主要围绕研究层面和应用层面开展实践探索，涉及产学研多个领域，主要集中在通用领域，涵盖自然语言处理（NLP）、计算机视觉（CV）和语音识别（ASR）等技术领域。在研究层面，中国大模型在科研方面已形成涵盖理论方法、软硬件的体系化能力，学术界和产业界共同参与，形成更加完善的理论框架。在应用层面，目前大模型已开始渗透到各行各业实现商业化应用，超过半数的大模型已实现开源，大部分企业前期以内部应用为主，后续主要向 B 端企业拓展服务，预计少数企业将在 C 端市场形成规模。从大模型的市场融资角度看，截至 2023 年上半年，与“大模型”直接相关的融资事件超 20 起；从被投企业的地域分布来看，北京、上海以及杭州位居前列，特别是北京，融资事件数量超 10 起，凸显

出北京在人工智能领域的领先地位。从大模型的业界评估看，业界普遍认为，国内大模型与 GPT-4 有较大差距，但科技大厂具备追赶实力。目前第一梯队当属百度、阿里巴巴、腾讯、华为四家。四家企业不不仅在大模型研发投入、技术能力和人才团队实力较强，且在商用推进方面均依托现有业务领域，更容易形成大模型应用规模效应。

二、大模型的应用场景

大模型具备强大的自然语言与多模态信息处理能力、超强的迁移学习和少样本学习能力，可以实现对不同领域、不同数据模式的适配，这些特点使得大模型较容易赋能其他行业，提升行业效率。如在信息检索领域，大模型可以从用户的问句中提取出真正的查询意图，检索出更符合用户意图的结果，还可以改写查询语句从而检索到更为相关的结果；在新闻媒体领域，大模型可以根据数据生成标题、摘要、正文等，实现自动化新闻撰写。在智慧城市领域，大模型能够实现城市—园区—商圈—社区—网点级别的智能计算与研判，为线上线下数实融合的智能决策和场景交互提供具有 AI 自学能力的“空间 AI 专家顾问”服务。在智慧办公领域，大模型可以协助用户撰写各类文档，实现文档创作、编辑和总结等功能，用户只需用自然语言提出需求，即可以快速生成或修改文档内容。此外，大模型还可以根据用户要求自动生成演示文稿幻灯片，以及完成数据统计分析，并将结果以图表形式清晰可视化呈现。在影视制作领域，大模型可以应用于剧本创作、角色设计和音乐配乐，为影视制作带来更多

元化和个性化的创意，还能用于视频内容分析，实现内容标签化和智能推荐，提升观众的观影体验。此外，大模型还可以应用于生物科技、智慧军事、智能教育等领域。

三、部分典型大模型厂商及应用产品

1. 科大讯飞—星火：“讯飞星火认知大模型”是科大讯飞推出的新一代认知智能大模型，具有文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力、多模态能力7大核心能力。拥有跨领域的知识和语言理解能力，能够基于自然对话方式理解与执行任务。

2. 百度—文心一言：文心一言是百度打造出的人工智能大语言模型，具备跨模态、跨语言的深度语义理解与生成能力。文心一言有五大能力，包括文学创作、商业文案创作、数理逻辑推算、中文理解、多模态生成，其在搜索问答、内容创作生成、智能办公等众多领域都有更广阔的想象空间。

3. 阿里—通义千问：阿里通义大模型是阿里云推出的超大规模的语言模型，功能包括多轮对话、文案创作、逻辑推理、多模态理解、多语言支持，是阿里云在人工智能领域的重要成果之一。该模型在多个领域得到了广泛应用，为业务提供了强大的基础框架和工具，推动了人工智能技术的发展和应用。

4. 商汤—商量：商汤科技以“坚持原创，让AI引领人类进步”为使命，是第一家将人脸支付技术集成至地铁售票系统中的公司，北京大兴国际机场正式投运由商汤科技AI技术支持的58套“智能旅客安检系统”，为全球旅客提供服务。

5. **360—360 智脑**：360 公司主营 360 安全卫士，是中国领先的互联网安全软件与互联网服务公司。目前，“360 智脑”为 360 自研千亿参数大语言模型。后续，360 公司将坚持自研核心技术和通过大模型在城市、政府和企业端布局和落地的“两翼齐飞”大模型战略，进一步服务产业数字化和智能化。



图 4 部分典型通用大模型及垂直大模型厂商及应用产品

(来源：《AI 大模型市场研究报告（2023）》《人工智能大模型体验报告 2.0》《中国人工智能系列白皮书——大模型技术（2023 版）》《我国人工智能大模型发展动态》)

大模型的未来发展趋势

一、大小模型协同进化，推动端侧化发展

未来几年，大模型和小模型将会协同推动人工智能的发展，实现明确分工，高效率低成本地解决业务问题。大模型负责向小模型输出模型能力，小模型更精确地处理自己“擅长”的任务，再将应用中的数据与结果反哺给大模型，让大模型持续迭代更新，形成大小模型协同应用模式，达到降低能耗、提高整体模型精度的效果。大规模参数并不是产业所追求的重点，更少的标注数据依赖、更优的模型效果、更高的模型性能以及便捷的部署方式将是未来研究的重点。

二、大模型通用性持续加强，实现 AI 开发“大一统”模式

大模型通过无标注数据进行自监督学习，从而降低标注数据的人力要求，GPT-3、文心大模型均展示出了在未标记数据中的学习成果，并展示了在不同任务与行业领域上的通用性。同时，多模态大模型也逐渐兴起，数据形态差异化问题也将得到解决，未来大模型将进一步致力于构建通用的人工智能底层算法框架，融合多领域的模型能力，在不同场景中“自我学习”，使得通过一个大模型能够解决产业中各种问题。目前，在通用模型的基础上，各行业正利用精调或 prompt 的方式加入任务间的差异化内容，从而极大地提高了模型的利用率，推动 AI 开发走向“统一”。

三、大模型从科研创新走向产业落地，通过开放的生态持续释放红利

大模型最重要的优势，是推动 AI 进入大规模可复制的产业落地阶段，仅需零样本、小样本的学习就可以达到很好的效果，以此大大降低 AI 开发成本。目前，大模型已经开始与领域、行业深度融合，例如，工业质检、蛋白质结构预测等领域的大模型，验证了大模型已不仅在科技企业中应用，也迈出了走向各行各业的步伐。同时，ChatGPT 专注于强化内容创造能力，将生成式 AI 应用到实际业务中，为大模型带来新的产业落地机遇。具体来看，微软已经宣布将会全线整合 ChatGPT，将大模型嵌入搜索引擎和办公软件，进一步推动 AI 能力的全面赋能和产业落地。

（来源：《2022 中国大模型发展白皮书》）

编辑部地址：南宁市体强路 18 号广西信息中心 1412 号房

联系电话：0771-6113592

电子邮箱：dsjyjs@gxi.gov.cn

网 址：<http://gxxxxz.gxzf.gov.cn/>



扫描二维码获取
更多决策参考信息