



中国人工智能系列白皮书

——大模型技术（2023 版）

中国人工智能学会

二〇二三年九月

《中国人工智能系列白皮书》编委会

主 任：戴琼海

执行主任：王国胤

副 主 任：陈 杰 何 友 刘成林 刘 宏 孙富春 王恩东
王文博 赵春江 周志华

委 员：班晓娟 曹 鹏 陈 纯 陈松灿 邓伟文 董振江
杜军平 付宜利 古天龙 桂卫华 何 清 胡国平
黄河燕 季向阳 贾英民 焦李成 李 斌 刘 民
刘庆峰 刘增良 鲁华祥 马华东 苗夺谦 潘 纲
朴松昊 钱 锋 乔俊飞 孙长银 孙茂松 陶建华
王卫宁 王熙照 王 轩 王蕴红 吾守尔·斯拉木
吴晓蓓 杨放春 于 剑 岳 东 张小川 张学工
张 毅 章 毅 周国栋 周鸿祎 周建设 周 杰
祝烈煌 庄越挺

《中国人工智能系列白皮书----大模型技术》编写组

陶建华 吴 飞 黄民烈 文继荣 王海峰 刘知远
刘 静 杨小康 聂 帅

目录

第 1 章 大模型技术概述	1
1.1 大模型技术的发展历程	1
1.2 大模型技术的生态发展	5
1.3 大模型技术的风险与挑战	7
第 2 章 语言大模型技术	9
2.1 Transformer 架构	9
2.2 语言大模型架构	13
2.2.1 掩码语言建模	13
2.2.2 自回归语言建模	14
2.2.3 序列到序列建模	14
2.3 语言大模型关键技术	15
2.3.1 语言大模型的预训练	15
2.3.2 语言大模型的适配微调	17
2.3.3 语言大模型的提示学习	20
2.3.4 语言大模型的知识增强	22
2.3.5 语言大模型的工具学习	23
第 3 章 多模态大模型技术	25
3.1 多模态大模型的技术体系	25
3.1.1 面向理解任务的多模态大模型	25
3.1.2 面向生成任务的多模态大模型	27
3.1.3 兼顾理解和生成任务的多模态大模型	29
3.1.4 知识增强的多模态大模型	31
3.2 多模态大模型的关键技术	32
3.2.1 多模态大模型的网络结构设计	32

3.2.2 多模态大模型的自监督学习优化	33
3.2.3 多模态大模型的下游任务微调适配	35
第 4 章 大模型技术生态	37
4.1 典型大模型平台	37
4.2 典型开源大模型	40
4.2.1 典型开源语言大模型	40
4.2.2 典型开源多模态大模型	49
4.3 典型开源框架与工具	53
4.4 大模型的训练数据	56
4.4.1 大模型的训练数据处理流程和特点	56
4.4.2 大模型常用的公开数据集	59
第 5 章 大模型的开发训练与推理部署	62
5.1 大模型开发与训练	62
5.2 大模型推理部署	64
5.2.1 大模型压缩	65
5.2.2 大模型推理与服务部署	66
5.3 软硬件适配与协同优化	67
5.3.1 大模型的软硬件适配	68
5.3.2 大模型的软硬件协同优化	68
第 6 章 大模型应用	70
6.1 信息检索	70
6.2 新闻媒体	71
6.3 智慧城市	72
6.4 生物科技	72
6.5 智慧办公	73
6.6 影视制作	74
6.7 智能教育	74

6.8 智慧金融	75
6.9 智慧医疗	75
6.10 智慧工厂	75
6.11 生活服务	76
6.12 智能机器人	76
6.13 其他应用	76
第 7 章 大模型的安全性	78
7.1 大模型安全风险引发全球广泛关注	78
7.2 大模型安全治理的政策法规和标准规范	79
7.3 大模型安全风险的具体表现	81
7.3.1 大模型自身的安全风险	81
7.3.2 大模型在应用中衍生的安全风险	82
7.4 大模型安全研究关键技术	84
7.4.1 大模型的安全对齐技术	84
7.4.2 大模型安全性评测技术	87
第 8 章 总结与思考	90
8.1 协同多方合作，共同推动大模型发展	91
8.2 建立大模型合规标准和评测平台	92
8.3 应对大模型带来的安全性挑战	93
8.4 开展大模型广泛适配，推动大模型技术栈自主可控	94
名词索引	95
参考文献	97
编写人员贡献	116

第 1 章 大模型技术概述

1.1 大模型技术的发展历程

2006 年 Geoffrey Hinton 提出通过逐层无监督预训练的方式来缓解由于梯度消失而导致的深层网络难以训练的问题[1]，为神经网络的有效学习提供了重要的优化途径。此后，深度学习在计算机视觉[2]、语音[3]、自然语言处理[4]等众多领域取得了突破性的研究进展，开启了新一轮深度学习的发展浪潮。总结过去十多年的技术发展，基于深度学习的人工智能技术主要经历了如下的研究范式转变：从早期的“标注数据监督学习”的任务特定模型，到“无标注数据预训练+标注数据微调”的预训练模型，再到如今的“大规模无标注数据预训练+指令微调+人类对齐”的大模型，经历了从小数据到大数据，从小模型到大模型，从专用到通用的发展历程，人工智能技术正逐步进入大模型时代。

2022 年底，由 OpenAI 发布的语言大模型 ChatGPT 引发了社会的广泛关注。在“大模型+大数据+大算力”的加持下，ChatGPT 能够通过自然语言交互完成多种任务，具备了多场景、多用途、跨学科的任务处理能力。以 ChatGPT 为代表的大模型技术可以在经济、法律、社会等众多领域发挥重要作用。大模型被认为很可能像 PC 时代的操作系统一样，成为未来人工智能领域的关键基础设施，引发了大模型的发展热潮。

本次大模型热潮主要由语言大模型（亦称为大语言模型）引领。语言大模型通过在海量无标注数据上进行大规模预训练，能够学习到大量的语言知识与世界知识，并且通过指令微调、人类对齐等关键技术拥有面向多任务的通用求解能力。在原理上，语言大模型旨在构建面向文本序列的概率生成模型，其发展过程主要经历了四个主要阶段[5]：