

# 大模型评述

孙柏林

(中国自动化学会专家咨询委员会,北京 100000)

**摘要:**大模型是人工智能行业飞速前行的缩影。如今,以大模型、算力等为代表的前沿技术领域不断发展,正在塑造人工智能产业发展的新格局。随着人工智能技术的迅猛发展,AI大模型一直被视为推动人工智能领域提升的关键因素,大模型已经成为了引领技术浪潮研究和应用方向。现从何为大模型,大模型的种类,大模型的优势,大模型案例与大模型的未来等方面对于大模型加以评述。

**关键词:**人工智能;大模型;深度学习

**中图分类号:**TP391.9 **文献标识码:**A

## Review of Large Models

SUN Bo-lin

(Expert Consult Council, Chinese Automation Institute, Beijing 100000, China)

**ABSTRACT:**The big model is a microcosm of the rapid progress of the artificial intelligence industry. Nowadays, cutting-edge fields such as large models and computing power are constantly developing, shaping a new pattern for the development of the artificial intelligence industry. With the rapid development of artificial intelligence technology, AI big models have always been regarded as a key factor driving the improvement of the field of artificial intelligence, and big models have become a leading direction in the research and application of technology waves. This article reviews the definition, types, advantages, cases, and future of large models.

**KEYWORDS:** Artificial Intelligence; Large model; Deep learning

### 1 引言

最近,以 ChatGPT 为代表的人工智能大模型引发社会热议,各行各业都已感受到了人工智能的日益逼近。AI 大模型作为当前最热门的技术领域,得到了广泛的关注。以谷歌的 BERT 和 OpenAI 的 GPT 为代表,模型的参数规模逐步提升至千亿甚至万亿级,训练数据量级也大幅提升,随之带来了模型能力的显著提高,掀起了国内外 AI 大模型的研究热潮。

大模型是人工智能行业飞速前行的缩影。如今,以大模型、算力等为代表的前沿领域不断发展,正在塑造人工智能产业发展的新格局。大模型正助推 AI 发展跨越拐点。

随着人工智能技术的不断发展,深度学习已经成为了目前最为热门的技术之一。而在深度学习领域中,大模型(Large Scale Model)则是一种备受关注的技术。大模型是指由数百万、甚至数十亿个参数组成的神经网络模型,它们可

以处理大规模的数据集,并且具有非常强的学习能力和泛化能力。

2023 年以来,人工智能引发持续关注。在近日举行的 2023 年亚布力中国企业家论坛第十九届夏季高峰会上,与会嘉宾热议人工智能大模型。

2023 年 3 月以来,国内 AI 大模型宛如雨后春笋般地“冒”了出来,“百模大战”就此打响。随着人工智能技术的迅猛发展,AI 大模型一直被视为推动人工智能领域提升的关键因素,大模型已经成为了引领技术浪潮研究和应用方向。

大模型是人工智能领域的热门研究方向。以 ChatGPT 为代表的大模型技术应用掀起了国内外的大模型研究热潮,大模型参数规模和训练数据量级迅速增加,模型性能显著提升。今年上半年,大模型和生成式 AI 可以说是全球最受瞩目的科技突破之一。为此,商业迎来了极其关键的发展时期。

专家认为,人工智能:进入产业级大模型时代。大模型将会是未来十年科技领域里面最重要的事情之一。大模型将开启人工智能的“大一统时代”。

人工智能是引领新一轮科技革命与产业变革的战略性技术,是赋能经济社会数字化、绿色化双转型的主要引擎,具有很强的“头雁”效应。当前,全球很多国家都将数字经济和智能经济作为未来经济增长的重要动力。我国应在数据要素领先、算力、算法跟跑的现状下,高度重视夯实人工智能底层技术与基础软件,营造安全可信的人工智能生态,加快推动应用落地。

## 2 何为大模型?

赋予 AI 大模型的定义是, AI 大模型是通过深度学习算法和人工神经网络训练出的具有庞大规模参数的人工智能模型。

大模型是指具有庞大规模和复杂结构的人工智能模型,它们具有数以亿计的参数和深层次的神经网络架构。这些模型通过学习海量数据和深度神经网络的优化,在各种任务上取得了令人瞩目的成果。

这些模型使用大量的多媒体数据资源作为输入,并通过复杂的数学运算和优化算法来完成大规模的训练,以学习和理解到输入数据的模式和特征。这些模式和特征最终通过大模型中庞大的参数进行表征,以获得与输入数据和模型设计相匹配的能力,最终来实现更复杂、更广泛的任务,如语音识别、自然语言处理、计算机视觉等。

### 大模型的产生过程

追寻大模型的起源,应该从那篇《Attention is All You Need》开始,基于这篇由谷歌机器翻译团队提出的由多组 Encoder、Decoder 构成的机器翻译模型 Transformer 开始,大模型的发展大致走上了两条路:

一条路是舍弃 Decoder 部分,仅仅使用 Encoder 作为编码器的预训练模型,其最出名的代表就是 Bert 家族。这些模型开始尝试“无监督预训练”的方式来更好的利用相较其它数据而言更容易获得的大规模的自然语言数据,而“无监督”的方式就是 Masked Language Model (MLM),通过让 Mask 掉句子中的部分单词,让模型去学习使用上下文去预测被 Mask 丢掉的单词的能力。在 Bert 问世之处,在 NLP 领域也算是一颗炸弹,同时也在许多自然语言处理的常见任务如情感分析、命名实体识别等中都刷到了 SOTA, Bert 家族的出色代表除了谷歌提出的 Bert、ALBert 之外,还有百度的 ERNIE、Meta 的 RoBERTa、微软的 DeBERTa 等等。

另外另一条路,则是, Bert 的进路没能突破 Scale Law,而这一点则由当下大模型的主力军,即大模型发展的另一条路,通过舍弃 Encoder 部分而基于 Decoder 部分的 GPT 家族真正做到了。GPT 家族的成功来源于一个研究人员惊异的发现:“扩大语言模型的规模可以显著提高零样本 (zero-shot) 与小样本 (few-shot) 学习的能力”,这一点与基于微调的 Bert 家族有很大的区别,也是当下大规模语言模型神奇能力的来源。而从 GPT-3 开始,当下的 ChatGPT、GPT-4、Bard 以及 PaLM、LLaMA 百花齐放百家争鸣,带来了当下的大模

型盛世。

从模型发展来看,人工智能经历了机器学习模型、深度学习模型、预训练模型和大规模预训练模型四个阶段。

### 数据——大模型的力量源泉

归根结底,大模型的神奇能力是来源于 GPT 么? 人们认为答案是否定的, GPT 家族几乎每一次能力的跃迁,都在预训练数据的数量、质量、多样性等方面做出了重要的提升。大模型的训练数据包括书籍、文章、网站信息、代码信息等等,这些数据输入到大模型中的目的,实质在于全面准确的反应“人类”这个东西,通过告诉大模型单词、语法、句法和语义的信息,让模型获得识别上下文并生成连贯响应的能力,以捕捉人类的知识、语言、文化等等方面。

一般而言,面对许多 NLP 的任务,可以从数据标注信息的角度将其分类为零样本、少样本与多样本。无疑,零样本的任务 LLMs 是最合适的方法,几乎没有例外,大模型在零样本任务上遥遥领先于其它的模型。同时,少样本任务也十分适合大模型的应用,通过为大模型展示“问题-答案”对模式,可以增强大模型的表现性能,这种方式一般也称为上下文学习 (In-Context Learning)。而多样本任务尽管大模型也可以去覆盖,但是微调可能仍然是最好的方法,当然在一些如隐私、计算等约束条件下,大模型可能仍然有用武之地。

同时,微调的模型很有可能面对训练数据与测试数据分布变化的问题,显著的,微调的模型在 OOD 数据上一般表现都非常差。而相应的, LLMs 由于并没有一个显式的拟合过程,因此表现要好许多,典型的 ChatGPT 基于人类反馈的强化学习 (RLHF) 在大部分分布外的分类与翻译任务中都表现优异,在专为 OOD 评估设计的医学诊断数据集 DDXPlus 上也表现出色。

“从目前来看,不可能出现一家 AI 大模型一统天下的情况,因为 AI 大模型是否能够得到广泛使用,一方面要看技术能力,另一方面则要看应用场景和流量格局。”

“AI 大模型百花齐放只是一个阶段性现象。在演变的过程中,各家企业和机构会逐渐找到自己的定位,走向细分市场。”

李彦宏(百度创始人、董事长兼首席执行官)曾在演讲中分析了技术方面的演进:大模型之所以会改变人工智能,原因就在于大算力、大模型、大数据,导致了智能涌现。他解释了何为“智能涌现”:过去的人工智能是想让机器学会什么技能,就教它什么技能。大模型导致智能涌现后,以前没教过的技能,机器也会了。同时他还表示,人工智能发生了方向性改变,从辨别式 AI 走向生成式 AI。回到大模型本身的应用上,李彦宏在演讲中表示,大模型会重新定义营销和客服。“道理很简单,就是谁拥有最佳的跟客户沟通的方式,谁就会拥有这个客户。”

科技创新总是在发现问题、解决问题中前行,每一项颠覆性创新成果从来都不是水到渠成的自然发展,而是由量的积累带来质的提升。对于人工智能大模型而言,参数规模的

倍增带来了智能涌现。当下,大模型密集发布,各大厂商纷纷加码人工智能,这样的大模型“狂”潮还会持续多久,发展方向是什么?

针对业界关注的热点问题,《人民邮电》报记者采访了北京邮电大学经济管理学院教授曾剑秋,他认为:

AI大模型兼具“大规模”和“预训练”两种属性,在海量数据上进行训练后能够完成一系列下游任务。

曾剑秋认为,大模型带来的颠覆性创新可以总结为:规模大、速度快、场景全。

首先,参数规模上,从 GPT-1 的 1.17 亿参数到 GPT-3 的 1750 亿参数,大模型的参数规模动辄上千亿,能够从海量数据和知识中学习,在不同领域和场景中“轻松”完成任务。

其次,得益于信息基础设施稳步推进带来的算力提升,大模型能够在短时间内响应用户并提供归纳整理后的内容,从聊天机器人转变为数字助手。

最后,在应用广度方面,大模型似乎是“全才”,写诗、编程、绘画、解题样样全能,具有广阔的应用场景。

曾剑秋还认为,大模型是建立在网络能力、数据能力和计算能力三者之上的创新应用。大模型需要在多个设备之间进行数据传输和参数同步,而坚实的网络基础为大模型提供了高效、稳定、可扩展的训练和部署环境。数据要素作为数字经济的核心生产要素,已成为推动产业升级、优化经济结构和打造经济增长点的战略性资源。

数据能力对于大模型的训练和应用至关重要,因为大模型需要从海量数据中学习和捕获知识以提升泛化能力。在计算能力方面,大模型需要海量的计算资源来执行复杂运算和存储巨量参数,需要使用高性能、低功耗、高度并行的计算设备和系统(如 GPU),可以说,算力是大模型训练、人工智能发展的基础设施。

曾剑秋谈到,这三种能力是大模型发展、人工智能产业进步、信息技术创新应用的“沃土”。入局大模型的门槛主要包括基础门槛、训练门槛和应用门槛。曾剑秋谈到,基础门槛包括网络能力和数据处理能力,需要具备大模型训练计算所需的高性能服务器、GPU 设备、高速网络等。训练门槛是指大模型的“成长”迭代需要不断训练,需要提供符合一定标准的高质量训练数据,因而需要花费时间和计算资源来对数据进行清洗和预处理,从而保障数据的质量和准确性。应用门槛是指大模型的落地需要与垂直行业需求深度融合,同时,对于企业来说需要权衡收益与研发成本。基于公有数据的大模型是通用底座,很难满足专业场景的特定需求。

曾剑秋谈到,如何打通大模型技术和产业应用的“最后一公里”,找到计算资源和产业应用效果之间的平衡,是需要关注的重点课题。大模型“热”需要“冷”思考,要超前布局也要理性发展。曾剑秋表示,盲目跟风研发大模型不可取,当热潮退去,留下的注定是推动人工智能技术革新、深耕垂直行业应用的数智实干家。同时,曾剑秋建议建立国家级的基础大模型平台,加强智算中心、超算中心等算力资源统筹,

加强跨数据中心算力协同能力,从战略高度着手,培育大模型发展基础,构筑人工智能发展优势。

众所周知,一个 AI 由模型、数据、算力三要素构成。相比于传统的识别型 AI(小模型),以 GPT 为代表大模型在数据、模型等方面均有不同程度的革新,赋予了大模型更强的通用性。这恰好改善了小模型时代的产业化痛点,AI 实现即插即用。AI 企业终于能摆脱手工作坊般的生产模式,有机会变成一门好生意。

大模型的这一价值,几乎吸引了全球所有的目光。连早已退休、专心搞慈善的比尔盖茨,都为此兴奋不已。他在一篇文章中写道,自己有幸亲历了人类可能最重要的两场革命的开端:第一次发生在 Windows 萌芽、PC 市场刚刚兴起的 80 年代,而第二次正是去年——大模型刚刚开始涌现的时候。

当前通用与行业大模型研发呈爆发式增长。大模型对当前产业发展将起到怎样的作用?

冯俊兰(中国移动集团级首席科学家,IEEE Fellow)认为:

1 大模型带来了人工智能技术范式的变革。正如牛顿定律之于物理学的意义一样,大模型开启了人工智能的“大一统时代”。

二是大模型带来了人工智能研发模式的转变。包括三个趋势:第一,研发团队由小到大,需要以企业为主体的创新模式。大模型的打造是一个集大算力、大数据、算法和应用于一体的系统工程,需要算法和工程兼备的专家型人才,具有高投入、高门槛和长周期特性,企业在大模型研发上更具优势,这将催生以企业为主体的产学研合作的创新模式。第二,研发路径由分方向、分领域到分层。传统模型是面向特定场景、特定任务训练得到的,而大模型是面向多种任务的通用模型。第三,大模型训练推理需要将各要素统筹考虑,包括底层计算芯片、框架、算法、理论等在内,进行端到端优化。

三是大模型还将催生人工智能新的业务赋能模式。

四是大模型也将引发产业范式的变化,形成新的产业生态结构,推动人工智能产业链分工更清晰,整个生态的研发效率得到显著提升,产业链更加成熟。

事实上,大模型的本质是数据、算力与算法的组合。大模型爆火的背后,是对算力基础设施建设的更高要求。算力作为整个数字经济时代的“水电煤”,为 ChatGPT 等人工智能大模型的快速发展提供了关键的技术与基础。

### 3 大模型的种类

大模型可以应用于各种领域,例如:

1) 自然语言处理:大模型可以用于自然语言处理任务,例如机器翻译、语音识别和文本生成等。

2) 图像处理:大模型可以用于图像处理任务,例如图像分类、目标检测和人脸识别等。

3) 推荐系统:大模型可以用于推荐系统任务,例如电影

推荐、商品推荐和音乐推荐等。

4) 游戏 AI: 大模型可以用于游戏 AI 任务, 例如围棋、扑克和象棋等。

大模型的分类主要分为结构分类和用途分类。具体如下:

#### 结构分类

深度模型包含多层神经网络的模型, 例如深度卷积神经网络、深度循环神经网络等; 宽度模型每一层有多个神经元的模型, 例如宽度卷积神经网络、宽度循环神经网络等; 深度宽度均衡模型深度和宽度都较大的模型, 例如残差网络、深度分……更多。

#### AI 大模型的分类

AI 大模型根据不同维度有不同的分类方式, 如根据任务类型可分为监督学习模型、无监督学习模型和强化学习模型; 根据模型结构可分为 DNN 深度神经网络、CNN 卷积神经网络、RNN 循环神经网络; 根据模型规模可以分为大规模模型和中小规模模型。而重点从模型数据类型、模型工作模式和模型开发模式来分析 AI 大模型的分类。

1) 按模型数据的媒体类型, AI 大模型可以分为语言模型、图像模型以及多模态模型:

#### 语言模型(Language Models)

语言模型是一个单纯的、统一的、抽象的形式系统, 语言客观事实经过语言模型描述, 比较适合于电子计算机进行自动处理, 因而语言模型对于自然语言的信息处理具有重大的意义。

#### 主要类型:

语言模型主要有三种类型: ①生成性模型, ②分析性模型, ③辨识性模型。

生成性模型从一个形式语言系统出发, 生成语言的某一集合, 如 N. 乔姆斯基的形式语言理论和转换语法。

分析性模型从语言的某一集合开始, 根据对这个集合中各个元素的性质的分析, 阐明这些元素之间的关系, 并在此基础上用演绎的方法建立语言的规则系统, 如苏联数学家 O. C. 库拉金娜和罗马尼亚数学家 S. 马尔库斯用集合论方法提出的语言模型。

在生成性模型和分析性模型的基础上, 把二者结合起来, 便产生了一种很有实用价值的模型, 即辨识性模型。辨识性模型可以从语言元素的某一集合及规则系统出发, 通过有限步骤的运算, 确定这些元素是一堆乱七八糟的词还是语言中合格的句子。如 Y. 巴尔-希列尔用数理逻辑方法提出的句法类型演算模型。

#### 图像模型(Image Models)

又分为: 图像描述模型、图像的视觉模型、图像运动模型、图像数据模型、

图像检索模型、纹理图像模型、图像差值模型等等。

#### 多模态大模型(Multimodal Models)

多模态大模型指的是将文本、图像、视频、音频等多模态

信息联合起来进行训练的模型。多模态大语言模型(MLLM)是近年来兴起的一个新的研究热点, 它利用强大的大语言模型作为大脑来执行多模态任务。MLLM 令人惊讶的新兴能力, 如基于图像写故事和无 OCR 的数学推理, 在传统方法中是罕见的, 这表明了一条通往人工通用智能的潜在道路。

2) 根据模型工作方式, AI 大模型可以分为生成模型和强化学习模型:

#### 生成模型(Generative Models)

概率生成模型, 简称生成模型(Generative Model), 是概率统计和机器学习中的一类重要模型, 指一系列用于随机生成可观测数据的模型<sup>[2]</sup>。生成模型的应用十分广泛, 可以用来不同的数据进行建模, 比如图像、文本、声音等。

#### 强化学习模型(Reinforcement Learning Models):

强化学习(Reinforcement Learning, RL), 又称再励学习、评价学习或增强学习, 是机器学习的范式和方法论之一, 用于描述和解决智能体(agent)在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题。

强化学习的常见模型是标准的马尔可夫决策过程(Markov Decision Process, MDP)。按给定条件, 强化学习可分为基于模式的强化学习(model-based RL)和无模式强化学习(model-free RL), 以及主动强化学习(active RL)和被动强化学习(passive RL)。强化学习的变体包括逆向强化学习、阶层强化学习和部分可观测系统的强化学习。求解强化学习问题所使用的算法可分为策略搜索算法和值函数(value function)算法两类。深度学习模型可以在强化学习中得到使用, 形成深度强化学习。

3) 根据模型开发模式, AI 大模型可以分为开源大模型和在线大模型:

#### 开源大模型(Open Source Models)

开源大模型是指由社区或组织共同维护的, 具有较高品质的、免费的 AI 模型库。这些模型库包含了各种领域的模型, 涵盖了图像分类、自然语言处理、语音识别等多个领域。在这些开源模型库中, 用户可以找到适合自己需求的模型, 并且可以在自己的数据集上进行训练和部署。

目前, 开源大模型已经成为 AI 领域中的热门话题。这些模型具有以下优点:

1) 易于使用: 开源大模型可以帮助用户简化模型训练和部署的过程。用户不需要从头开始训练模型, 只需要下载预训练好的模型并进行微调即可。

2) 开放性: 开源大模型是由社区或组织共同维护的, 可以保证透明、公正、开放的特点。用户可以自由地使用、修改和分享这些模型, 促进了 AI 行业的发展。

3) 高质量: 大多数开源大模型都是由专业的团队开发和维护的, 质量有保证。同时, 这些模型也经过了大量的测试和优化, 可以达到较高的准确率和效果。

4) 免费: 开源大模型是免费的, 用户可以无需支付任何

费用就可以使用这些模型。这对于中小企业和个人用户来说非常有利。

#### 在线大模型(Online Models)

在线大模型是指部署在云端或网络上的大型机器学习或深度学习模型。这些模型通常由云服务提供商或在线平台提供,并通过接口或 API 的方式进行访问和使用。在线大模型的优点是用户无需关注底层的硬件和软件架构,只需通过网络请求即可获得模型的预测结果。在线大模型还可以实现实时或按需的模型调用,适用于各种应用场景,如语音识别、图像处理和自然语言处理等。

上述对 AI 大模型的分类只是一些常见的示例,实际上,大模型的分类可以更加细分,根据具体任务和应用领域的需求而定。

专家认为,行业大数据的利用与管理值得高度关注,它是行业大模型的基础,行业大模型是中国千行百业升级发展的必然选择。这是由于下列原因:

第一,大模型是由算力驱动的系统性工程。

第二,包括智算场景在内的算力应用是未来云计算业务增长的一个新的热点。

第三,大模型算力是落实国家‘东数西算’的一个最佳场景,也是一个很好的应用范式。”

## 4 大模型的优势

大模型的优势:

相比于传统的小型模型,大模型具有以下优势:

1) 更强的学习能力:由于大模型拥有更多的参数和更复杂的结构,因此它们可以更好地拟合数据,并且具有更强的学习能力。

2) 更好的泛化能力:由于大模型可以处理更大规模的数据集,因此它们可以更好地发现数据中的规律和模式,并且具有更好的泛化能力。

3) 更高效的计算能力:虽然大模型的参数数量很多,但是由于它们采用了分层设计和分布式训练等技术,因此可以在现有的硬件设备上高效地训练。

专家们分析, AI 大模型的特点诸如:

一是大量的参数:大模型通常具有庞大的参数规模,拥有数以亿计的参数,这些参数可以存储模型的知识 and 经验,更多的参数意味着模型具有更强大的学习能力和表示能力,能够更好地捕捉数据中的复杂模式和特征,以便进行推理和预测。

二是上下文理解和生成:大模型能够理解和生成更具上下文和语义的内容,通过注意力机制、上下文编码器等关键技术来学习和训练大量的语言、图像等输入数据,可以从复杂的真实场景中提取有用的信息。

三是强大的泛化能力:大模型通过在大规模数据上进行训练,具有强大的泛化能力。它们从大量的数据中学习得到广泛的特征和模式,并且能够在未学习过、未见过的数据上也

同样表现良好。对未学知识的泛化能力也是评估大模型的重要指标。

四是计算资源需求大:大模型对于数据和计算资源的需求非常大。需要强大的计算资源来进行参数优化和推理,这需要具备出色的并行计算能力的 GPU、TPU 处理器集群,这使得训练和使用这些模型成为一项具有挑战性的任务。

五是迁移学习能力:大模型在一个或多个领域上进行预训练,并能够将学到的知识迁移到新任务或新领域中。这种迁移学习能力使得模型在新任务上的学习速度更快,同时也提高了模型在未知领域中的性能。

六是预训练与微调:大模型可以采用预训练和微调两阶段策略。在预训练阶段,模型通过大规模无标签数据进行学习,学习到一种通用表示。在微调阶段,模型使用有标签数据对模型进行细化训练,以适应具体的任务和领域。这种在大规模数据上进行预训练,再在具体任务上进行微调,能够让大模型适应不同的应用场景。

七是多领域应用:大模型应用领域广泛,可应用于多个领域,并解决多种任务,如自然语言处理、计算机视觉、语音识别等。大模型不仅在单一模态领域中有很强的表现,更能够进行跨模态的任务处理。

人工智能通用大模型可以依靠人工“喂养”的大量数据,极大地扩充知识边界。

## 5 大模型案例

人工智能产业风头正劲,迄今已有 80 余个大型模型公开发布。

当前人工智能领域涌现出了许多强大的 AI 大模型,下面列举出一些目前备受瞩目的 AI 大模型:

#### 案例 1: OpenAI GPT 大模型组

ChatGPT 是 OpenAI 于 2022 年 11 月发布,其在自然语言的理解和生成上的卓越表现使得在短短两个月的时间用户突破 1 亿大关,其是基于 GPT (Generative Pre-trained Transformer) 架构开发大型语言模型,其为对话式交互提供更好的支持和响应,并在社交对话、问题回答和一般性对话等场景中展现出优秀的表现。

#### 案例 2: Google PaLM & PaLM 2 大模型组

PaLM (Pretraining and Fine-tuning Language Model) 是在 2020 年由 Google Research 团队发布的一种用于自然语言处理任务的预训练和微调模型,它的第二个版本,最新的大型语言模型 PaLM 2 于 2023 年 5 月在 Google I/O 开发者大会上推出,其将支持谷歌 AI 聊天机器人 Bard,其凭借改进的数学、逻辑和推理技能,可以帮助生成、解释和调试 20 多种编程语言的代码。且为了满足更多的使用场景, PaLM2 提供了 4 个模型: Gecko、Otter、Bison、Unicorn,其中最小的 Gecko 模型可以在移动端运行,并计划在下一代 Android 系统中集成。

#### 案例 3: 百度文心大模型组

百度于 2023 年 3 月正式发布了 AI 大模型文心一言。

基于百度智能云技术构建的大模型,文心一言被广泛集成到百度的所有业务中。并且推出了文心 NLP 大模型、文心 CV 大模型、文心跨模态大模型、文心生物计算大模型、文心行业大模型。且提供了多样化的大模型 API 服务,可通过零代码调用大模型能力,自由探索大模型技术如何满足用户需求;

#### 案例 4:讯飞星火认知大模型

科大讯飞于 2023 年 5 月正式发布了星火认知大模型,其具有 7 大核心能力,即文本生成、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。

#### 案例 5:阿里通义大模型

阿里通义大模型覆盖语言、听觉、多模态等领域,致力于实现接近人类智慧的通用智能,让 AI 从“单一感官”到“五官全开”,分别在 2023 年 4 月和 6 月推出了通义千问和通义听悟。

#### 案例 6:清华开源大模型 ChatGLM

GLM-130B 是清华智谱 AI 开源项目,其目的是训练出开源开放的高精度千亿中英双语模型,能够让更多研发者用上千亿参数模型。并且在 2023 年 3 月开源了更精简的低门槛大模型 ChatGLM-6B,这是一个具有 62 亿参数的中英文双语语言模型,在 6 月份,推出了二代开源模型 ChatGLM2-6B,具有更强大的性能、更长的上下文、更高效的推理(推理速度提升 42%)、更开放的开源协议。

以上这些大模型只是当前众多 AI 大模型中的一小部分,随着技术的不断进步和研究的不断推进,可以期待更多更强大的 AI 大模型的涌现。

伴随着大模型热度的不断攀升,其在教育领域的应用及优势也开始显现。记者看到,当前已成为学生家庭刚需的智能硬件,在大模型的加持下呈现出巨大变化——越来越“聪明”的使用体验已经让原本的学习工具,逐渐具备了更多惊人的功能和附加价值。

人工智能的迅速发展为人类社会带来了巨大的变革。经过多年的发展,人工智能已经进入了大规模产业应用阶段,在金融、安防、医疗等多个领域实现应用。

在沈抖看来,大模型创业带来的众多机会,将带动数字化经济更进一步,有机会在全球范围内掀起一股产业再造的浪潮。

他认为,随着大模型深入数字经济,更多产业会出现新的改变。智能化不仅让整个生产流程大幅提效,也改变了很多原有的生产关系,包括人和人、人和设备、人和系统的关系。以前没有智能时,很多工作实际上是靠人来操作;有了智能以后,机器、设备、系统都可以按照人的思想去学习。

很多时候,“大模型很好!”这个断言后紧跟着的问题就是“大模型怎么用,什么时候用?”,面对一个具体任务时,应该选择微调,还是不假思索的上手大模型?

而在美国,已经诞生了依托于大模型开展业务的独角兽。在德克萨斯州,一家初创企业凭借微调后的 AI 应用,在短短 18 个月时间做到了 15 亿美元估值,年收入已有 3000 万

美元,比直接做大模型的企业还赚钱<sup>[10]</sup>。

不过,产业化潜力仅仅是大模型价值的一个方面。不同于过去两年流行的元宇宙、Web3 等概念,大模型带来了实打实的生产力提升。

2023 年 6 月 27 日,北京市科委、中关村管委会举办北京市人工智能策源地引领推介活动的首期活动——北京市首批人工智能行业大模型应用案例发布会,发布北京市首批 10 个行业大模型典型应用案例。

车企或自研 AI 大模型,或与通用大模型企业合作,围绕智能交互和自动驾驶方面进行探索……近日,AI 大模型落地汽车行业的消息不断传出。不过,有业内人士认为,其商业化前景依旧不明朗,未来发展任重道远。

AI 大模型扎堆“上车”当前,AI 大模型正在各行各业快速渗透,汽车也成为其目标落地场景。

大模型产业应用将成未来焦点。

据悉,华为发布全面智能化战略,赋能千行万业。2023 年 9 月 20 日,在华为全联接大会 2023 上,华为提出全面智能化战略,将加速推动千行万业智能化转型。业内人士表示,华为发布了多个行业智能化解决方案,服务越来越多的业务场景。为抓住 AI 战略机遇,华为提出全面智能化战略,让所有对象可联接,让所有应用可模型,让所有决策可计算。

华为将通过算力底座、AI 平台、开发工具的开放,支持大模型在智能化时代“百花齐放”。

当前,我国大模型发展从“百模大战”向产业应用转变。大模型如何赋能实体经济?我国存在哪些优势和挑战?未来大模型产业将呈现何种趋势?针对以上问题,《人民邮电》报记者日前采访了中国移动集团级首席科学家、人工智能与智慧运营中心总经理冯俊兰,国内大模型新锐公司衍远科技市场部负责人石晓锐以及瑞莱智慧(Real AI)合伙人、高级副总裁朱萌。

联储证券在研报中指出,2022 年 AI+教育的市场规模已经达到 3000 亿元,未来的市场规模有望达到 7 万亿。与此同时,7 月 14 日,国家网信办联合发改委、教育部、科技部、工信部、公安部、广电总局公布了《生成式人工智能服务管理暂行办法》,鼓励生成式人工智能技术在各行业、各领域的创新应用。而此前,相关政策利好也在接连落地。

以上这些大模型只是当前众多 AI 大模型中的一小部分,随着技术的不断进步和研究的不断推进,可以期待更多更强大的 AI 大模型的涌现。

据悉,在军事领域,生成式 AI 的迅猛发展及其在认知攻防对抗领域中的广泛应用,是未来认知战发展的大趋势。这就要求要深入研究生成式 AI 的认知应用,把握其作用发挥的方法路径。AI 大模型和生成式 AI 是人工智能领域的两大热门话题,它们在各个领域展现出了巨大的潜力和价值。随着技术的不断进步,相信这两大技术将会在未来带来更多的惊喜和创新。让我们一起期待这场智能科技浪潮的席卷吧!

## 6 大模型的未来

AI 产业能够二度迎来春天,背后其实是大模型技术迈向成熟。

大模型必然是未来很长一段时间人们工作生活的一部分,而对于这样一个与人们生活高度同频互动的“大家伙”,除了性能、效率、成本等问题外,大规模语言模型的安全问题几乎是大模型所面对的所有挑战之中的重中之重,机器幻觉是大模型目前还没有极佳解决方案的主要问题,大模型输出的有偏差或有害的幻觉将会对使用者造成严重后果。同时,随着 LLMs 的“公信度”越来越高,用户可能会过度依赖 LLMs 并相信它们能够提供准确的信息,这点可以预见的趋势增加了大模型的安全风险。

除了误导性信息外,由于 LLMs 生成文本的高质量 and 低成本,LLMs 有可能被利用为进行仇恨、歧视、暴力、造谣等攻击的工具,LLMs 也有可能被攻击以未恶意攻击者提供非法信息或者窃取隐私,据报道,三星员工使用 ChatGPT 处理工作时意外泄漏了最新程序的源代码属性、与硬件有关的内部会议记录等绝密数据。

展望一下大模型研究的未来,目前大模型主要面临的挑战可以被归类如下:

实践验证:当前针对大模型的评估数据集往往是更像“玩具”的学术数据集,但是这些学术数据集无法完全反应现实中形形色色的问题与挑战,因此亟需实际的数据集在多样化、复杂的现实问题上对模型进行评估,确保模型可以应对现实世界的挑战;

模型对齐:大模型的强大也引出了另一个问题,模型应该与人类的价值观选择进行对齐,确保模型行为符合预期,不会“强化”不良结果,作为一个高级的复杂系统,如果不认真处理这种道德问题,有可能会为人类酝酿一场灾难;

安全隐患:大模型的研究要进一步强调安全问题,消除安全隐患,需要具体的研究确保大模型的安全研发,需要更多的做好模型的可解释性、监督管理工作,安全问题应该是模型开发的重要组成部分,而非锦上添花可有可无的装饰;

模型未来:模型的性能还会随着模型规模的增加而增长了,这个问题估计 OpenAI 也难以回答,针对大模型的神奇现象的了解仍然十分有限,针对大模型原理性的见解仍然十分珍贵

AI 大模型具有诸多优点的同时也存在一些挑战和限制,如训练时间长、计算资源需求大、模型复杂度高、通用泛化能力受限等等。此外,由于其庞大的参数规模,大模型可能面临可解释性和隐私等方面的诸多挑战。

百度首席技术官(CTO)王海峰说,“大模型发展到今天,已经表现出了能理解、有逻辑、会推理、能创作、有情感的能力。同时,大模型也出现了一些问题,比如信息时效性相对差。”这是因为在进行模型训练时,不会把每天最新的数据和信息及时训练到底座模型中,导致模型在生成过程中出现准

确性较低问题。

王海峰进一步表示,大模型研发依赖算法、算力和数据综合支撑,在应用层面上也面临一系列技术挑战:首先是数据规模大,数据质量参差不齐;其次是模型体积大,算法难度高;第三是算力规模大,性能要求高。

王海峰说,尽管大模型涉及非常复杂的算法和算力数据,但王海峰对其应用前景十分看好。他表示,大模型打造出来之后,应用方只需基于大模型平台进行操作和使用。

“大模型的产业化路径可以简单理解为,生产大模型的厂商基于复杂的算法和算力数据把模型生产‘铺装’出来,然后通过推理部署支持千行百业,让大模型得到很好的应用。”

## 7 小结

尽管存在着诸多问题和挑战,但 AI 大模型的蓬勃发展已经势不可挡。最近几年,随着深度学习和硬件技术的快速发展,出现了一系列强大的大模型,其中最著名的就是以 Transformer 架构为基础的 BERT、GPT 和 T5 等模型。以 GPT-3 为例,它具有 1750 亿个参数。该模型在自然语言处理任务中表现出色,能够生成高质量的文本、回答问题和进行对话。而这仅仅是大模型的开始...

AI 大模型作为人工智能领域的重要进展,具有庞大的参数数量和强大的计算能力。其发展历程经历了多年的技术突破和硬件进步。通过不断地研究和改进,性能更优,功能更强的 AI 大模型在不断的涌现,也在各行业领域中发挥更广泛、更深远的影响。

目前,AI 企业开始在“大模型的划时代”意义上达成共识。

在新一轮 AI 竞争中,我国能否快速跟上?有专家认为,构建一个大模型至少需要几千颗芯片,大算力需要芯片之间互联,而芯片互联需要超高速带宽。目前,我国大模型在整体综合实力和创新发展上,较头部企业仍有较大差距。

受限于生产制造的设备、工艺,我国许多企业仍在追赶阶段。在推理领域则蕴藏着诸多机遇,由于场景碎片化,还没有任何国际标准和垄断性的企业。我国仍在追赶阶段。

高文(中国工程院院士,北京大学教授)说,“现在算力最强的是美国,算力第二是中国,但中国的算力增长速度是全世界第一。算力在整个经济活动中、在高质量发展中是非常关键的,不仅在改变目前人工智能的格局,也在改变科学研究的格局。”

然而,国内算力布局存在供需对接失衡的情况:中西部地区算力过剩,而用于产业互联网的算力不足。在高文看来,这一问题的解决途径是像建设电力网一样建成算力网络,即把算力通过网络连接起来,形成随处可见的一张网。

“算力要上网需要解决两个挑战性问题,一是这些资源怎么封装、调度,二是怎么让这些节点有高速、超宽带的技术连接。这是现在亟须解决的技术问题。”高文介绍说,为了让用户像用电一样方便使用算力,

(下转第 24 页)



- 响[J]. 物流技术与应用,2020,25(4):134-136.
- [13] 曹琦,陈闻轩. 应急设施选址问题研究综述[J]. 计算机工程,2019:1-14.
- [14] C K Y Lin, R C W Kwok. Multi-objective metaheuristics for a location-routing problem with multiple use of vehicles on real data and simulated data [J]. European Journal of Operational Research, 2006, 175(3): 1833-1849.
- [15] Zheng Zeyu, Wang Jianping, Wang Xiumin. ONU Placement in Fiber-Wireless (FiWi) Networks Considering Peer-to-Peer Communications [C]. IEEE: IEEE, 2009.
- [16] 许骏,许晓东. 一种群体智能融合算法及其在应急设施选址的应用[J]. 计算机工程与科学,2014,36(4): 667-673.
- [17] A Sinha, P Malo, K Deb. A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications [J]. Arxiv, 2017.
- [18] W Zhao, Z Zhang, L Wang. Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications [J]. Engineering Applications of Artificial Intelligence, 2020, 87 (Jan.): 103300. 1-103300. 25.

- [19] I Zelinka, L Skanderov. Simulated Annealing in Research and Applications [J]. Chapters, 2012.
- [20] 张得志,潘立红,李双艳. 考虑供应商选择的选址—库存—路径的联合优化[J]. 计算机应用研究,2019,36(8): 2338-2341.

#### [作者简介]



张煜(1985-),男(汉族),福建省福州市人,硕士研究生,工程师,主要研究方向为网络智能运维、多属性决策。

王磊(1983-),男(汉族),江苏省丹阳市人,博士,副教授,硕士研究生导师,主要研究方向为系统

仿真、优化理论。

姚昌华(1982-),男(汉族),重庆市人,博士,教授,硕士研究生导师,主要研究方向为智能无线网络、电磁频谱对抗。

林春盛(1984-),男(汉族),福建省泉州市人,硕士,工程师,主要研究方向为指控自动化。

#### (上接第7页)

今年将组织全国做算力、做网络的人一起推进项目。

“未来大模型产品的发展趋势可能是通用大模型与垂直领域细分模型的结合。”中国人民大学数字经济研究中心主任李三希认为,这不仅需要具备坚实的技术基础,如大规模、高质量、多样化的语料库,创新的大模型算法,自研的机器学习框架和强大的算力基础设施等,也需要大模型产品具有坚实的基于场景的应用。未来,从实践中来,到实践中去的“实用级”大模型将成为趋势。

我们相信,AI大模型将会沿着“在发展中前进,在前进中发展”的道路奋勇向前!

#### 参考文献:

- [1] 百度研究院. 2023年十大科技趋势预测[R]. [https://mp.weixin.qq.com/s/jEiqFTLh5E3h\\_2h7RP7ClA](https://mp.weixin.qq.com/s/jEiqFTLh5E3h_2h7RP7ClA), 2023-01-05.
- [2] 阿里巴巴达摩院. 十大科技趋势[R]. <https://files.alicdn.com/tpsservice/9b8b05280c05294f158bdb0dee61f78b.pdf>, 2023-02-24.
- [3] 张依依. 大模型正在塑造人工智能产业新格局[N]. 中国电子报, 2023-4-11.
- [4] 张乾君. AI大模型发展综述[J]. 2023-3.
- [5] 齐旭. 大模型叩响工业大门[N]. 中国电子报, 2023-8-4.
- [6] 赵熠如. “百模大战”来了 AI大模型市场百花齐放[N]. 中国商报, 2023-8-1.
- [7] 沈怡然. AI大模型要革AI的命[N]. 经济观察报, 2023-8-14.
- [8] 卢岳. AI教育大模型落地词典典[N]. 消费日报, 2023-8-17.
- [9] 杨子怡, 实习生, 谭心琪. 大模型产业应用将成未来焦点[N]. 人民邮电, 2023-8-18.
- [10] 杨子怡, 实习生, 谭心琪. 大模型产业应用将成未来焦点[N].

人民邮电, 2023-8-18.

- [11] 杨阳腾. 人工智能应用前景广阔[N]. 经济日报, 2023-9-14.
- [12] 罗锦钊, 孙玉龙, 钱增志, 周鲁, 王金桥. 人工智能大模型综述及展望[N]. 无线电工程, 2023-08-29.
- [13] 李洋. 智能算力已成AI大模型主要动力[N]. 中国高新技术产业导报, 2023-8-28.
- [14] 罗云鹏. 国产大模型: 创新为道落地为王[N]. 科技日报, 2023-9-18.
- [15] 车万翔, 等. 大模型时代的自然语言处理: 挑战、机遇与发展[J]. 中国科学: 信息科学, 2023, (9).
- [16] 刘佳. 创业者涌入AI战场大模型将迎应用层之战[N]. 第一财经日报, 2023-9-21.
- [17] 黄婉银. 行业首倡大模型建筑新未来中国价值地产年会今日深圳举行[N]. 每日经济新闻, 2023-09-25.
- [18] 余惠敏. 人工智能大模型布局千行百业[N]. 经济日报, 2023-9-27.
- [19] 栾娜. 大模型在产业链纵深处实现价值[N]. 河南日报, 2023-10-4.
- [20] 曹渝. 戍边退役士兵地位获得及政策仿真研究. 国家社会科学基金重点项目资助(2021ASH003);
- [21] 焦勇勤. 大模型媒介: ChatGPT引发的智能传播革命及其社会影响[J]. 海南大学学报(人文社会科学版), 2023-10-08.
- [22] 何喜军. 锻造国产大模型核心竞争力[N]. 经济日报, 2023-10-10.

#### [作者简介]



孙柏林(1936.11-),男(汉族),北京人,少将,研究员,《计算机仿真》杂志顾问,主要研究方向: 军事科学。