

大模型展现出的通用智能能力具有重要的现实意义和影响力,是人工智能发展中里程碑式的进展。大模型带来的智能革命,将媲美工业革命和电力革命,深刻改变人类社会的生产生活方式。大模型,或许正是开启智能时代的那一台“蒸汽机”。



“百模大战”打响,我们究竟需要什么样的大模型

深瞳工作室出品

采写:本报记者 都芃 刘艳
策划:刘莉 李坤

要问今年最热的科技热点是什么,许多人会脱口而出——大模型。从年初 ChatGPT 引发的大模型浪潮,到相关科技企业上线自家大模型产品,“百模大战”不断升级,它所引发的人工智能“地震”还在持续。

进入9月,国内科技厂商在大模型领域动作频频。9月初,此前一直处于内测阶段的多款大模型产品完成备案,正式上线,面向公众开放。9月21日,在华为全联接大会2023上,因在诸多领域展示了大模型发展的更多可能,盘古大模型成为当之无愧的主角。

“百模大战”也好,“群模起舞”也罢,技术进步和创新喷涌的热潮中,人们关注的焦点集中于,我们究竟需要什么样的大模型?何种大模型是通向人工智能的最终答案?

智能时代的“蒸汽机”

现在到底有多少个大模型产品?似乎没有人能给出准确的答案。

赛迪顾问发布的《2023 大模型现状调查报告》显示,截至2023年7月底,国外累计发布大模型138个,中国则累计有130个大模型问世。

但这股浪潮的源头很清楚,正是 ChatGPT。2022年末,由名不见经传的人工智能公司 Open AI 打造的大语言模型 ChatGPT 横空出世,人们可以使用最日常的语言与其对话,让它回答各种问题、完成各类书面任务,聊天、画图、敲代码……人们第一次如此近距离地感受到人工智能带来的震撼,也让“大模型”这个此前只为少数人工智能领域从业者所知的专业概念,成为了被人们挂在嘴边、反复提及的词汇。

输入内容,进行计算,产出结果,这是

人工智能模型的本质,但大模型的与众不同之处在于其足够“大”。

巨大的参数数量、庞大复杂的算法结构、海量的训练数据以及强大的算力支撑,决定了大模型不同于普通的人工智能模型,其拥有强大的通用智能。

以 ChatGPT 为例,其早期 3.0 版本的参数数量已达到 1750 亿个,4.0 版本的参数量更是指数级增加,达到 1.8 万亿个,训练数据的单位数量更是达到惊人的 13 万亿。相比之下,早期如围棋机器人 AlphaGo 等专用人工智能模型,其参数量往往仅为百万级,与大模型相去甚远。

海量的参数、复杂的算法、强大的训练能力,让大模型成为具备横跨多领域学习能力的通用人工智能,实现了所谓的“智能涌现”,即拥有触类旁通、举一反三的学习能力。如果说此前的人工智能模型只是经过训练后可以完成某些特殊任务的“马戏团猴子”,那么大模型则可以说已经是一个接受义务教育的“中学毕业生”。

在北京智源人工智能研究院院长、北京大学计算机学院教授黄铁军看来,大模型展现出的通用智能能力具有重要的现实意义和影响力,是人工智能发展中里程碑式的进展,放眼人类历史,展望人类未来,大模型带来的智能革命,将媲美工业革命和电力革命,深刻改变人类社会的生产生活方式。大模型,或许正是开启智能时代的那一台“蒸汽机”。

公之于众。

谷歌紧随 ChatGPT 脚步,发布了基于自家大语言模型的聊天机器人“Bard”,互动媒体与服务公司 Meta 也不甘落后,发布了同类大模型“Llama”。国内同样热闹非凡,3月份,百度率先发布研发十余年的知识增强大语言模型“文心一言”。随后,360发布“360智脑”,阿里巴巴上线“通义千问”,科大讯飞推出“讯飞星火”,腾讯推出“混元”,华为推出“盘古”……一众科技公司不甘落后地推出了各自的大模型产品。

彼时,面向公众发布的大模型产品均为内测使用,公众需提出申请,获得内测资格后才能进行体验。7月10日,国家网信办联合国家发改委、教育部、科技部等七部门审议通过并发布《生成式人工智能服务管理暂行办法》(以下简称《办法》),该《办法》8月15日起正式施行。《办法》明确规定,提供具有舆论属性或者社会动员能力的生成式人工智能服务的,应当按照国家有关规定开展安全评估,并履行算法备案等手续。8月底,一批完成算法备案的大模型产品陆续上线,正式面向公众开放。

或许是 ChatGPT 带来的震撼过于强烈,以自然语言对话为主要功能体验的大语言模型产品成为许多科技厂商的首选,甚至一度成为大模型的代名词。首批通过备案上线的大模型产品中有百度的“文心一言”、字节跳动的“豆包”、智谱华章的“智谱清言”等。截至9月底,已有十余家大语言模型产品官宣上线,大部分均提供聊天对话、创意写作、代码生成等服务。

但在大语言模型之外,也有厂商选择了另一条道路,不“恋战”大语言模型,而是直面行业具体问题。

在9月21日举办的华为全联接大会2023上,华为常务董事、华为云 CEO 张平安表示,盘古大模型要帮助各行各业的客户解最难的题。例如,在矿山领域,精煤产率是困扰业界多年的难题,山东能源集团已将“盘古”矿山大模型全面应用到采、掘、机、运、通等9大业务系统,21个场景中。在“盘古”的助力下,济宁二号煤矿每年能多产出8000吨精煤,精煤产率提升千分之二,增收约2000万元。

正在举办的亚运会上,也不少了大模型的身影。由浙江移动打造的亚运保障网络运维大模型在幕后为亚运会保驾护航。它让保障人员“无论在哪,无论何时”都能以“对话”方式获取保障情况,全面提升了亚运会保障的工作效率,降低了保障工作的技术门槛。

浙江移动网管中心副总经理方伟告诉科技日报记者,无论是面向具体行业的垂直大模型,还是直面公众的大语言模型,从底层技术来看,二者并不冲突。“这两种选择就像是此前的互联网发展,既有改变人们生活的移动互联网,也有改变社会、改变行业的产业互联网。大模型也是基于通用的底层技术,衍生出了不同用途。比如浙江移动的网络运维大模型就是使用通用大模型通过微调和提示工程来满足亚运网络保障的场景。”

垂直大模型和通用大模型,究竟哪条赛道能够通向终极人工智能的“罗马城”,答案或许是一条条大路通罗马。

“万模群舞”或在不久的将来

无论是哪种类型的大模型,在“百模大战”的背景下,其功能、用途、场景的重叠都无法避免。但在业内人士看来,大模型的发展还远未触及天花板,不仅“百模大战”不是终点,“万模群舞”或许就在不久的将来。

“大模型的研发是一个拼细节的过程。从技术路线上看,目前各家基本上都是基于 Transformer 架构来做,方法很类似,但效果确实不一样,决定成败的是细节。”腾讯有关负责人接受采访时说,由于资源投入程度、细节把握程度的不同,最后不同产品的差异会逐渐显现。“从应用领域来说,会迎来一个‘百花齐放’的场景,比如有的人专注于自然语言处理,有的人专注于医疗领域,有的人专注于教育领域等。就目前而言,大模型的天花板还远远没有触到,技术体系和应用场景都

在不断演进。”腾讯有关负责人说。

无论是“百模大战”还是“万模群舞”,要避免低水平的雷同复制,杜绝“重复造轮子”带来的资源浪费,关键仍在于不断丰富大模型的应用生态。今年5月,国际期刊《自然》发表了一项百度在生物计算领域的突破性研究成果,其提出的 mRNA 序列优化算法 LinearDesign,对生物医学领域创新具有重要价值,这也是中国互联网科技企业首次以第一完成单位的身份在《自然》正刊发表成果。而在这背后,是百度在2022年5月便推出的“文心”生物计算大模型。借助大模型,人类能够快速找到并设计出活性更好、性质更优的候选物分子,从而大幅提升新药研发和疫苗设计效率,让新药研发从单点突破的“手工作坊”阶段进入到规模化开发阶段。目前,基于“文心”生物计算大模型和飞桨深度学习框架搭建的面向小分子、大分子和 RNA 的药物设计平台——飞桨螺旋桨 PaddleHelix,已经应用于超过30家医药企业。

“无论是从技术层面还是产品层面,百花齐放、百家争鸣对于当下大模型发展都更加有利。算法研究阶段可能五花八门,但到工程选型阶段,可能会逐渐聚焦到一个或少数几个类型。而到了具体产品选型层面,面向不同市场、不同行业,产品形态可能又是多种多样的。”清华大学计算机系教授唐杰认为,无论哪种大模型都有各自的局限性,各种模型互相学习、竞争,在市场中大浪淘沙是必经之路。

给大模型发展更多耐心

“百模大战”虽然激发了更多可能,但也带来了不少负面影响。当众多厂商蜂拥而上大模型时,参差不齐的产品让人眼花缭乱,其中暗藏的法律、道德风险也应引起人们的警惕。

例如,在面向公众的大语言模型产品中,“一本正经胡说八道”现象已经多次引发质疑。“我们常说大模型会产生‘幻觉’,就是因为大模型的内容准确性还不高,专业性能力不足,如果大模型产生的内容直接公开在互联网,会使当前良莠不齐的互联网信息质量更加低下。”方伟认为,大

记者手记

不要等大模型“无所不能”时才入局

◎ 刘艳

ChatGPT 让公众再一次体验到现代人工智能的强大功能,它背后的大模型技术及应用范式,将极大冲击现有的人工智能产业研发路线、发展方式,并有望成为新一轮科技革命与产业变革的核心驱动力和新入口。

这一范式的变化,引发了全球大模型竞速。中国工程院院士郑纬民认为,大模型是新型基础设施的关键底座之一,大模型的竞争也是国家科技战略的竞争。

谁都不想错过这一机遇,全球科技企业纷纷下场投身这一科技新赛道。在这一轮 AI 竞赛中,中国不能掉队。

《中国人工智能大模型地图研究报告》显示,从全球已发布的大模型分布看,中国和美国大幅领先,超过全球总数的80%,中国已形成了紧跟世界前沿的大模型技术群。

随着我国数字科技领军企业大模型悉数登场,各方思辨与争论不断,“百模大战”是否存在重复建设?垂直类大模型是否有建设的必要?

在百度创始人、董事长兼首席执行官李彦宏看来,新的国际竞争战略关键点,不是一个国家有多少个大模型,而是大模型上有多少原生的 AI 应用,这些应用在多大程度上提升了生产效率。他点出了一个浅显的道理,科技作用于社会、造福于人类,必会向不同行业延展。

先发者不一定制人。在大模型角逐中,全球基本形成了各有优势的中美两大集群,中国企业

型目前在内容安全、版权、主体责任等方面还存在多种风险。

而当深入到具体行业时,不同行业的特点也对大模型在准确性、安全性、专业性等方面提出更高要求。360创始人周鸿祎接受采访时表示,虽然公开的大模型是通用的,甚至是万能的,但是当深入到具体行业时仍然面临着缺乏行业深度、不懂企业内部知识、易导致企业内部数据泄露、无法控制成本等问题。方伟也同意这种说法:“例如大模型应用在通信网络保障方面时,就不允许有任何差错,目前准确性还有待提升。”

除了目前大模型自身仍然存在的缺陷,随着竞争的逐渐激烈,我国在相关数据、算力、算法等方面的困境也逐渐暴露。

例如,在训练数据方面,唐杰指出,目前我国的数据开放态势并不强,“大多数机构还是想把数据握在自己手里”。但唐杰也坦承,即使机构有数据开源的意愿,在目前环境下,仍然存在多重风险。他建议,我国关于大模型训练数据应用、开源的相关法律法规还需要进一步细化、健全,让有开源意愿的机构能够放心地进行数据开源,充分激发互联网精神。

而在算力方面,“百模大战”对算力的高需求也让我国本就紧张的算力资源更加捉襟见肘。在“东数西算”基础上建立起的算力互联网或有望成为解决这一问题的答案。通过对不同地区的算力资源进行调度,实现算力资源的均衡分配。但目前仍然需要在高带宽、低延迟的算力网络建设方面突破更多技术瓶颈。

如果说数据和算力是大模型的硬件“基础设施”,算法则体现着更多“人”的因素。小冰公司首席执行官李笛认为,就训练大模型而言,参数的数量并不是最重要的,工程化过程中的调优才是真正的考验。“研发、训练一个大模型可能只需要一个月,但调优可能需要一年的时间。这个过程格外需要谨慎、细致、耐心,有足够的定力和专注力,也是一种“工匠精神”的体现。”

耐心,是谈及大模型发展时多位业内专家反复提及的关键词。“大模型不是把数据‘喂’进去,算力一跑就有了。这其中有许多复杂细致的工作要做。”唐杰表示,人工智能自20世纪50年代被提出至今,其发展已经历过多次起伏,任何技术的发展都不是一蹴而就的,要对其有充分耐心。



①基于大模型的诗歌生成。
②华为盘古大模型语义理解系统。
③基于大模型的现代农业模型。
④中国移动“九天”AI酷玩体验馆。
⑤基于大模型的住院医生站管理系统。
本版图片由视觉中国提供