

ThinkBlue
深蓝智库

寻路大模型商业化

从拼参数到讲落地,短短半年时间,大模型经历了一次“神化”,又迅速“接地气”起来。在这轮“百模大战”中,人们越来越笃定,大模型将是比互联网以及移动互联网更大的机会。但在这个过程中,也不可避免地遭遇了技术、安全、商业化模式等方面的诸多挑战,如何跨过这些障碍,找到落地场景,打通商业模式,成为大模型的当务之急。

8月16日,由北京商报社、深蓝媒体智库主办的AI主题沙龙在新闻大厦艺术馆举行,沙龙以“大模型见真章”为主题,邀请业内人士围绕着大模型产品、技术路线、场景寻找、伦理安全等诸多问题,探寻大模型商业化落地的可行之径。会上,北京商报社社长兼总编辑李波涛在致辞中表示,在经济寻求向上突破的当下,几乎唯一没有争议的就是科技,科技可以带来力量。无论是产业还是行业,“破圈”都已形成一个越来越明显的趋势,希望通过深蓝智库的活动,实现产业、行业、人群的“破圈”,将更多前沿的科技成果向更广泛的产业、行业、人群进行传播和输出。

这一波大模型不是泡沫

百度科技与社会研究中心主任余欢



“当所有人都相信这种发展时,就未必是泡沫了”。余欢提到,这一波大模型的兴起不是AI概念第一次被热炒,经过70多年发展,人工智能经历了几次起伏,但不同的是,这一波大模型已经成为当下人工智能的主流方向。

余欢表示,相比起之前的人工智能,这一波大模型的特点是泛化性更强、标准化程度更高、工业化发展趋势更明显。

此外,余欢提到,相比起之前的模型,大语言模型的特点主要分为两部分:情商和智商。当下人们更多解决的是智商问题,即能理解、有逻辑、会推理、能创作,但未来就会向“有感情”去努力,也就是说情商将成为大模型未来一段时间内进一步提升的能力。

今年6月,ChatGPT首次出现流量负增长,一时延伸出业内的各种讨论。对此,余欢表示,ChatGPT流量下滑的逻辑类似于“打江山容易守江山难”,随着时间的推移,剔除一定猎奇心态后,自然会面临有多少用户黏性的问题。但对中国大模型企业而言,流量不仅没有见顶,

甚至可以说是还没开始。

至于商业化,余欢认为,要从付费和真正覆盖成本实现盈利两个方面看。虽然当下尚未出现大规模的批量化落地,但几家大模型厂商已经实现了零星的商业化落地,预计年底可能会出现一些典型的商业化落地案例。

大模型更安全,产业应用才能更深更广

蚂蚁集团大安全AI算法资深专家郑霖



重新定义,分别为大模型的可靠性、可控性、内容安全及合规以及伦理安全。

“总结下来,我们认为可信AI仍是一个非常难的科学命题,它希望达到的是超大规模数据生成学习范式下的安全、可控、可靠。”郑霖称。

除安全问题外,郑霖认为,大模型还需要重视生态的打造,从底层架构,到中间算法,再到上层应用,整体需要各行各业的从业者参与进来,将生态搭建起来。包括芯片短缺和算力等问题,都需要这样一个生态去解决。

在商业化问题上,郑霖提到,数据是一大障碍。他认为,大模型需要用对应企业的数据训练模型,但后者可能会认为这些数据是其核心资产,大模型都有强化学机制,会产生数据合规方面的担忧。

但一些新的场景、新的交互方式可能会延伸出新的发展。郑霖举例提到,之前数字人很火,大模型被认为是给元宇宙的数字人提供了灵魂,“垂类大模型可能会有一些更好的产出机会”。

中小微模式有望成为国内主流大模型

360智脑产品资深专家葛灿辉



在葛灿辉看来,大模型关乎国家生产力发展水平,中国要多方协力发展产业,未来中小微模式有望成为国内大模型发展的主流形式,通过模型压缩和优化,力争用有限资源达到近似性能。

网络安全起家的360是国内少数持续深耕人工智能核心技术的公司。

进入大模型赛道之初,360就确定了“四路齐发”场景布局,即面对消费者推出用户个人AI助力;面向中小微企业提供SaaS(软件即服务)化垂直应用;面对企业、政府和城市类客户支持私有化部署;面对行业客户推出垂直类大模型。

整个场景布局中的核心是API,这一平台的任务是赋能千行百业、打造开发者业务。“API主要提供各种能力,比如翻译、对话能力等,简单说就是API接人大模型可以做新应用。”葛灿辉这样介绍基于大模型开发行业应用的过程,“底层是360多个基础大模型,基于底层能力、API开发者工具和360智脑API开放平台,客户可以开发各种行业应用,比如知识管理、医疗问答、智能客服等。”

在葛灿辉看来,API的开放和兼容对快速推进上述

流程至关重要。

谈到未来大模型技术趋势,他认为,中小微模式有望成为国内主流。怎么理解中小微模式?葛灿辉这样解释,即在保持较高预测准确性的前提下,大幅度减小模型的大小和计算成本,优化计算效率,以实现高效的模型预测在较小的设备上运行。

大模型流量见顶后还能继续加速推进

元语智能联合创始人兼COO朱雷



“伪需求”,大量粗浅的、简单的伪需求会导致客户认为,向其兜售这种模型方案的厂商不够落地,进而呈现出劣币驱逐良币的情况。

“当下人们认为大模型难落地,也是因为没有发现有价值的应用。”朱雷称,面向企业客户做商业化落地,需要有真实的需求牵引,从用户出发,从需求出发,借用AI技术让业务快速增长。这其中涵盖了三个关键词,快速、业务和增长。

朱雷解释称,“快速”是因为很多企业等不及,受限于预算、经营环境以及客户需求等问题,企业通常不能花费一两年的时间等待大模型达到GPT4的效果才考虑落地。

“业务”是因为很多模型厂商兜售方案时,其实与其主营业务没有太大关系,只是泛泛而谈的大模型场景,而无法解决具体的业务问题。

“增长”则是因为当下很多方案都以“降本增效”为导向,但商业化公司的重点很多时候不在降本,而在增效。毕竟省钱的空间是有限的,只有挣钱才是最重要的事情。

资金不足导致大模型“外热内冷”

文渊智库创始人王超



“每一个科技浪潮都经历了爆发、泡沫、瓶颈和沉淀期”,王超例数了国内外科技圈发展的著名节点,“但以后冒出点水花的智能硬件、VR(虚拟现实)都没有找到新的方向,一直到2023年大模型的出现。很多人认为大模型标志着一个新的科技浪潮的到来。”

与其他几个经典风口不同的是,“大模型浪潮在爆发、泡沫、瓶颈和沉淀期的时间会缩短,它的商业化探索期也一样”,王超认为。至于所谓大模型流量见顶论,在他看来,“可能是舆论、普通用户参与度方面的观感,大模型的发展趋势和整个科技史没有区别,会在泡沫、瓶颈、产业爆发的路上继续”。

大模型商业化的困难来自于芯片,也来自于资金。“由于种种原因,资本投入大模型行业的‘弹药’相比智能手机等时代要小得多,资金量不大而且资金来源没那么多多元,这导致大模型行业即使在5、6月份最火的时期,整体也是‘外热内冷’,也让新入局企业的商业化面临问题”,王超说。

但他同时认为,“中国一定会找到一个新办法。因为

中国的创业者,包括新兴的和垂直类创业者,还在不停地进入大模型行业,中国创业者的基数优势依然明显。我们只需要找到一些应对芯片和投资的办法。很多政府类基金正在大踏步进入大模型领域,国内大模型企业的商业化一定会找到一个实现路径”。

AI颠覆性创新需要现有业务重构

新壹科技CTO张华伟



去重构,目前国内大模型混战还在起步阶段”。

7月27日,新壹科技发布了聚焦视频领域的垂直大模型,以及基于大模型的AIGC(人工智能生成内容)视频生成平台“秒创视频”、基于数字人的“新壹数字人”、明星数字人祝福视频制作产品“秒祝”。

之所以推出垂直类大模型,新壹科技的想法是,如果不从实际场景出发,会导致大部分通用大模型的实用性并不强,难以真正重塑实际的生产。新壹科技需要建立一个连接AI技术能力和垂直场景需求的大模型,能够提供释放生产力的能力,并创造更多实际价值。

在张华伟看来,AI是辅助工具,“像AI生成视频可以快速辅助创作者、达人以及媒体用户,快速提高生产力”。

不过,“这种效率提升需要一个过程,而且很漫长”,张华伟说,“大模型的能力是要在商业化场景落地的,需要数据、用户和场景。过程之所以漫长是因为企业有业务发展惯性,一个颠覆性的创新需要去做业务重构,这种重构会牺牲一部分人的利益,或者需要企业进行转型,要慢慢来。”

北京商报记者 杨月涵 魏蔚