

大模型爆火这一年

——对ChatGPT在全球引发AI“狂飙”的观察与思考

■本报记者 沈滢莎

“太忙了。”这是清华大学人工智能(AI)国际治理研究院副院长梁正面对记者脱口而出的一句话。

去年11月30日,非营利性研究机构OpenAI推出AI聊天机器人ChatGPT,它所展现出的智能涌现能力在全球范围内引发了一场AI“狂飙”,也在业界点燃了百亿美元大战。在记者询问过的每一个和大模型相关的人中,“忙”是出现频率最高的词。

忙碌的景象不仅在国内,那些大洋彼岸原本早已习惯了“慢慢”打磨产品的AI专家们也进入了“满负荷”状态。“在谷歌、微软或OpenAI的办公室里,你能感受到他们的工作节奏至少比之前快了两倍。”梁正说。

这一切都是因为以ChatGPT为代表的的语言模型向世人展现出了前所未有的能力,一条通往AGI(通用人工智能)的道路被打通,谁也不希望落于人后。

这一年,人工智能领域发生的事件比前十年加起来还要多。据不完全统计,目前国内的AI大模型已超过100个。除了百度、阿里等互联网大厂,创新工场CEO李开复、美团创始人王兴、搜狗创始人王小川等“科技老兵”纷纷重披战甲进军大模型领域。

这一年,大模型改变了公众的关注点。据柯林斯词典统计,今年人们提及“AI”的次数是去年的4倍。而有关AI是否会取代人类的担忧,也比以往任何时候都紧迫而强烈。

前所未有的变革

“这是倾盆大雨前的一场雨,是一个分水岭。”OpenAI首席科学家伊利亚·苏茨克维这样形容ChatGPT的诞生。

一年前的11月30日,ChatGPT上线,短短5天,用户数突破百万。由于太过火爆,OpenAI不得不暂时关闭了用户的测试入口。两个月后,ChatGPT月活跃用户破亿,刷新了App应用用户过亿的最快纪录。半个月前,OpenAI暂停ChatGPT Plus的新用户注册,原因是11月6日的开发者大会后,用户量激增。

一年来,ChatGPT从一个只能处理文字信息的网页端应用,变成了一个可以同时处理文字、语音、图像信息的超级应用,每周吸引上亿人使用。OpenAI也从一个非营利性研究机构变成了估值900亿美元的超级“独角兽”,聚拢了200万开发者。在最新演示中,ChatGPT的表现甚至超越了“神笔马良”——当有人想要做一个创业助手时,只要把概念说给GPT-4Turbo听,两三分钟后便可生成一个应用。

ChatGPT几乎以一己之力改变了AI研究生态。在它爆火之前,谷歌是该领域当之无愧的翘楚,不仅论文发表数量最多,其对思维链路展示、大模型涌现现象的研究也在前列。那时,自然语言处理、计算机视觉等还是众人追捧的AI前沿领域;现在,几乎所有人都涌向OpenAI引领的生成式AI方向。今年2月,研究自然语言处理的复旦大学计算机科学与技术学院邱锡鹏教授团队发布国内首个类ChatGPT模型“MOSS”,也曾一度因用户的大量涌入而“宕机”。

与谷歌技术路线相比,OpenAI是“力大砖飞”的代表。我们的算法非常简单。“伊利亚·苏茨克维直言不讳。他们把从海量数据中获取的标注信息导入模型,不断重复这个过程,最终得到一个超复杂的“黑箱”,即使是OpenAI首席科学家自己也不知道模型中正在发生什么。

ChatGPT上线后,全球大科技公司几乎都发布了自己的大语言模型,比如谷歌的PaLM2,Meta的Llama2,百度的文心一言,华为的盘古等,但只要使用过这些模型,你就会发现,GPT模型是“独一档的存在”。对此,梁正解释说:“这种差距可能无关芯片、算法、数据,纯粹是OpenAI比较幸运。”

幸运是指什么?不妨从凯文·凯利的畅销书《失控》中寻找答案。他在书中预测生命的智慧来自涌现,而科技也是一种生命。伊利亚同样认为,数据在大模型“黑箱”中发生的就是演化,尽管世界上有很多这样的“黑箱”,但发生在GPT身上的涌现与众不同,如同300多万年前开始直立行走的“人类祖母”露西。

借助于世界上最大的算力和最强的GPU芯片,这一年,GPT正以人类难以想象的速度继续进化。事实上,在基于GPT-3.5模型的ChatGPT上线时,OpenAI已经完成了GPT-4的训练。而据微软研究院估算,GPT-4已经是一个早期的、仍不完整的AGI系统。业界认为,800亿参数是大模型得以涌



一年前,OpenAI推出的大语言模型ChatGPT掀起了生成式人工智能的热潮,奥特曼(左)也成为科技界风云人物。均视觉中国

现的“门槛”,这意味着全球有多个大模型可能涌现智能。对于大模型的未来,梁正借用菲尔茨奖得主陶哲轩的说法表达自己的预见:如果使用得当,到2026年,AI将成为数学研究和许多其他领域值得信赖的合著者。而数学,一直是象征人类智力巅峰的学科领域。

前所未有的繁荣

在人工智能投资领域,美国与中国是全球投入最多的两个国家。随着ChatGPT的横空出世,各国对AI的投资继续加码,中美之间的AI投资差距正在缩小。

早在ChatGPT发布之初,中国计算机学会术语工委副主任、同济大学特聘研究员王昊奋就意识到,这是一个改变人与计算机交互模式的“接口”。这也解释了为何整个科技界对它如此兴奋,以至于任何可能“搭边”的人都动了心。

浙江大学上海高等研究院常务副院长、人工智能研究所所长吴飞对大模型的态度有过一次明显改变。“今年3月前,我对它还是旁观待变的态度,现在我已经深刻感受到它的‘通用之力’。”

大模型是比互联网门槛更低的“接口”。吴飞解释说,互联网的能力是赋能特定领域、特定任务。比如用互联网订餐,可以让程序员做一个美食推荐应用,而要赋能体育,则可以做一个体育新闻推送应用。因此,现在我们每个人的手机里都装着各种不同功能的应用。大模型则不同,基础大模型可以解答我们提出的任何问题,但如果想要在专业上更进一步,则需要垂直领域大模型。

过去一年,吴飞团队打造了多个垂直领域大模型。只用45TB(万亿字节)法律方面的专业数据,他们就训练出了“智海-录问”,一个为法律从业者提供法律问答、知识检索增强问答、案情分析、意图识别等多项法律辅助服务的工具。过去一年,该模型已应用于浙江省高级人民法院,大大加快了法官办案流程。

科大讯飞今年则一口气发布了12款行业大模型,涉及金融、汽车、运营商、工业、住建、法律等多个领域。科大讯飞总裁刘庆峰援引知名咨询机构Gartner的数据表示,目前全球已有5%的企业在使用大模型,到2026年,这一数字将达到80%。

在邱锡鹏看来,大模型在应用端的发力速度会比预计的更快,覆盖面也会更广。因为大部分软硬件服务公司都开始接入大模型,在感叹效率提高的同时,用户可能不知道自己其实已经在使用大模型了。

梁正用30年前的互联网“诞生时刻”来类比当下,且相较于前者,大模型能够覆盖的行业更多,更加无可避免。他甚至认为,30年的互联网繁荣似乎就是为大模型准备的,因为驱动大模型涌现的主要数据来自互联网。

这一年,大模型已经在教育、医疗等诸多领域展现出过人本领。不久前,ChatGPT根据病情描述和检查报告,帮助一位母亲为孩子找到了病因——新生儿发病率最高只有0.025%的脊髓栓系综合征。此前,这位母亲已经带着孩子求医3年,求助了17位医生,

都没有诊断出原因。

前所未有的分裂

如果说ChatGPT的问世让全球科技界沉浸在在大模型的狂欢之中,那么关于AI伦理之争则是新技术“狂飙”之下的灵魂拷问。

从3月底的千人联名信到5月底的AI风险声明,在要不要继续推动大模型发展,如何防范AI风险上,即使是AI阵营内部也呈现出前所未有的分裂。5月初,有“AI教父”之称的计算机科学家杰弗里·辛顿离开谷歌,加深了人们对AI是否已经失控的担忧。他坦言,之所以离开谷歌,就是为了完全自由地说出AI潜在的风险,向世人提出警示。

最具戏剧性的事件要属最近发生在OpenAI内部的“宫斗”:首席执行官山姆·奥特曼被董事会开除,几天后,他又宣布重返OpenAI。业内普遍推测,如何规制AI发展是公司内部矛盾爆发的导火索。作为掌握最“聪明”大模型的公司,OpenAI的态度很大程度上代表了AI发展的最真实状态,而据其判断,AGI有很大几率会在10年内到来。

人类已经致力于发展AI超过60年,为什么这一次如此警铃大作?伊利亚·苏茨克维表示,AGI与迄今为止人类创造的任何东西都不同,它是有自己目标的完全自主的生物。如果这一天到来,人与AGI的关系相当于人与动物,只不过这一次站在动物位置上的是人类。

奥特曼则表示,AGI只是发展AI道路上的一个小节点,他们的远景目标是创造出远超人类智能的超级AI。不过即便如此,他仍向美国国会表达了希望加强对OpenAI监管的愿望。今年7月,OpenAI首次提出“超级对齐”概念,宣布将投入20%的计算资源,花费4年时间全力打造一个超级对齐系统。

如何保证AI向善?梁正说,目前业界观点主要分为两派:一派认为应该先在AI中普及人类的道德规范再允许其发展;另一派则认为AI在发展中会形成自己的道德规范,就像人类在群体行为和人类一样的外形。前提是AI需要拥有和人一样的外形。

一方面人们对可能到来的AGI争论不休,而另一方面,充分的竞争则会加速AGI的到来。上海交通大学教授江晓原对人类能否对一路“狂飙”的AI适时“踩刹车”表示悲观。他说,企业是推动AI“狂飙”的决定性力量,它无法依靠自己的道德慢下来,因为如果你不做,别人就会赶你。

仍有许多科学家坚信AI只是人类发明的一件好用的工具,即便如此,它仍然有毁灭人类的危险。清华瑞莱智慧科技有限公司AI治理研究院院长张伟强提出了这样一个假设:人类总是在不断试错的过程中走向成功,而大模型提供了前往正确答案的直通车,人们将由此减少很多试错机会。正如刘慈欣在科幻作品《镜子》中描写的一种人类“结局”,人类因为从不犯错而走向灭亡。

还有更多问题

“与上半年的烈火烹油相比,下半年大

家对大模型的态度务实了许多,它还有很多问题没有解决,比如模型复杂度、多模态融合、模型评测等,我们都在全力解决。”邱锡鹏说。

吴飞也表示,继续通过某些任务对大模型进行特定训练,以期在某个指标上超过ChatGPT,意义可能并不大,中国在大模型上的机会与当年的互联网一样,即从实际应用倒逼模型迭代。

许多专家都表示,国内大模型的挑战主要来自两方面——算力和数据。大模型与其他软件应用不同,不仅训练起来费钱,运行起来更烧钱。想要处理用户输入的问题,基本上每个字都要在大模型上跑一遍,参数上万亿的大模型,每跑一遍都要调用多个A100 GPU,每个价值一万美元,十分“烧钱”。

国内使用大模型的用户画像也和国外有所区别。简单来说,国外有更多高端用户倾向于使用GPT,这些相对高质量的海量数据使GPT的迭代比国内大模型要快。

如何破局?当务之急是提高算力。据浪潮信息、国际数据公司(IDC)和清华大学联合发布《2021-2022年全球算力指数评估报告》,15个重点国家的算力指数平均每提高1个百分点,国家的数字经济和GDP将分别增长3.5%和1.8%,预计该趋势在2021-2025年将保持持续。吴飞认为,随着“算力网”的落地,国内大模型会在未来几年不断带来好消息。

大模型是AI发展唯一正确的路吗?Meta首席AI科学家杨乐昆对此并不认同。在他看来,大模型是“一条下坡路”,远离了通往更强大的AI的道路。他在《经济学人》撰文说:“大型语言模型能有多聪明、能有多准确,都是有局限的,因为它们没有现实世界的经验。”

邱锡鹏表示,在使用工程性的方法确保自己“不掉队”的前提下,还要想办法研究下一代大模型,即在算力不够的现有条件下,有无新途径去实现生成式AI。“前者我们只能跟随,后者则有几率超越。”他透露,MOSS正朝着这一目标努力。这一年来,它在架构上有所创新,规避了现有系统的一些缺陷;在“对齐”上采用了更人性化、更多维度的幻觉数据集,使内容生成更安全;在学习能力上也更加主动。

面对大模型高昂的成本和能耗,整个行业也在寻找解决办法。一些科技巨头亲自下场研发芯片,比如谷歌、亚马逊、微软,甚至OpenAI都在考虑针对AI运算自研芯片,或是让消费者购买更适合大模型计算的手机和电脑,分担算力成本。

“开源会让创新速度更快。”上海对外经贸大学人工智能与变革管理研究院副院长、上海开源信息技术协会秘书长张国锋说,目前国内大模型的开源方式都以企业为主导,并且只开源了整个大模型生态的一小部分,从而出现了算力分散、算力资源浪费等问题。他建议整合政府、高校、企业等优质的大模型生产要素数据、算法和算力,采用开源项目的协作模式和治理方式降低成本,建立大模型时代的“Linux”生态位。



发展时间线

- 2015年** 12月11日 OpenAI由埃隆·马斯克、山姆·奥特曼等人创办。
- 2016年** 6月16日 OpenAI发表生成式模型的研究成果。
- 2018年** 2月21日 马斯克从OpenAI辞职。
- 6月11日 GPT-1推出。
- 2019年** 2月14日 GPT-2推出。
- 7月22日 微软宣布与OpenAI合作,投资10亿美元开发AI工具。
- 9月19日 OpenAI根据使用偏好和反馈数据微调GPT-2。
- 2020年** 6月11日 GPT-3推出。
- 2022年** 1月27日 InstructGPT推出,提升遵循指令的能力。
- 11月30日 ChatGPT问世。
- 2023年** 1月23日 微软对OpenAI追加投资100亿美元。
- 1月30日 ChatGPT月活跃用户破亿。
- 2月1日 ChatGPT Plus推出。
- 2月7日 微软必应搜索引擎接入ChatGPT功能。
- 3月1日 ChatGPT应用程序编程接口推出。
- 3月14日 GPT-4推出。
- 3月22日 AI专家及行业高管发表联名信,呼吁6个月内暂停开发比GPT-4更强大的系统。
- 3月31日 意大利限制使用ChatGPT。
- 4月25日 ChatGPT用户可以控制模型训练的数据。
- 4月28日 ChatGPT在意大利恢复服务。
- 5月18日 iOS系统的ChatGPT应用程序推出。
- 5月23日 必应成为ChatGPT的默认搜索引擎。
- 5月30日 奥特曼与众多专家签署联合声明,呼吁降低AI带来的灭绝风险。
- 8月28日 OpenAI推出ChatGPT企业版。
- 9月25日 ChatGPT推出人机语音对话功能。
- 9月27日 ChatGPT可实时联网。
- 10月19日 ChatGPT应用程序内可运行DALL-E 3模型。
- 11月6日 OpenAI举行首届开发者大会。
- 11月15日 ChatGPT Plus暂停新用户注册。
- 11月17日 奥特曼被OpenAI解雇。
- 11月19日 OpenAI任命临时首席执行官,微软宣布奥特曼加入。
- 11月21日 奥特曼回归OpenAI,出任首席执行官。

(注:时间均为美国当地时间) 本报记者 孙钦祺编译整理



制图:张继

(上接第一版)选择美国这一全球最大的高端医学影像市场作为正式出海的首发站,源于联影医疗秉持的“高举高打”市场策略,即率先进入高端市场的高端医院,塑造行业影响力。

困难也显而易见。业内皆知,高端医疗装备行业是一个集多学科交叉、人才密集、知识密集和创新密集等特点于一体的高新技术产业,行业壁垒高。高端市场不仅对产品技术质量以及服务的要求极高,也已有盘踞多年的行业巨头。学术和科研合作先行,成为联影医疗打开美国市场的第一步。早在2017年,凭借业界领先的分子影像技术,联影医疗敲开了由加州大学戴维斯分校等美国顶尖分子影像专家组成的“探索者联盟”的大门——“探索者联盟”正式宣布双方将共同把全身PET扫描的PET设备从构想变为现实。在2018年的RSNA上,世界首款全身PET-CT uEXPLORER探索者

及其生成的人体全身动态PET-CT成像“电影”亮相,成为分子影像领域跨时代的创新技术。联影医疗也借此迎来美国市场的开门红。伴随更多的uEXPLORER入驻中国、美国、欧洲等地的顶尖研究机构,上百篇科研成果在顶尖期刊发表,更多合作的科研机构纷至沓来。

今年的美国核医学与分子影像学会年会上,耶鲁大学分子影像中心名誉主任理查德·卡森教授发布基于全球首台数字化脑专用PET设备NeuroEXPLORER拍摄的首张大脑影像图。“这一脑专用PET实现了令人难以置信的灵敏度。我们将它应用于帕金森病的研究,可以观测到疾病早期黑质细胞的病变;在非早期的痴呆症病变区域,我们也可以观察到常规设备难以达到的微小程度。未来,更多前沿的脑疾病研究将借助这一领先的设备开展。”他说。

这项世界级的创新成果来自2020

中国医疗科技以创新造福更广阔人群

年联影医疗美国子公司与耶鲁大学、加州大学戴维斯分校联合承担的美国国立卫生研究院脑计划重大项目。联影医疗美国子公司正是作为产业界唯一伙伴,开发了这台世界首创的设备。

更多的科研合作正基于联影医疗的创新技术不断开展。在拥有20多位诺奖得主的美籍华裔科学家领衔的联影医疗正与科研人员携手推动下一代磁共振智能应用开发,为肿瘤心脏病学、心脏电生理学等研究提供更有力的依据。

成功唯有依靠创新、依靠产品质量

回顾出海五年,张强说,他们始终

坚信,成功唯有依靠创新、依靠产品质量,而非低端价格竞争。当然,这绝对不是一条容易走的路。驰骋在竞争激烈的医疗器械赛道,联影医疗是凭借什么做到的?

世界首创全身临床超高频5T磁共振、世界首台75cm超大孔径3.0T磁共振、世界首款全身PET-CT uEXPLORER等一系列极具竞争力的“人无我有”创新技术与产品,正逐步敲开美国更多医院的大门。当前,美国超过30个州已安装联影医疗的设备。

“融合人工智能等先进技术,我们打造了一系列跨产品线的数智化超级平台。这不仅能帮助当地医院推动精准诊疗、提升就诊效率,还能弥补当地医疗机构经验不足、人才不足等短板。”张

强说。

在美国达拉斯,一家医院购买了联影医疗高端CT设备uCT Atlas。在对一名插满管道的患者的紧急救治中,该医院第一次使用了这台智能化CT。只需按下,这台CT就会自动完成摆位、扫描,且图像质量非常高。技师感叹,从没用过这么好操作的CT,极大提升了工作效率。

服务全球1万多家医疗机构

目前,联影医疗的创新产品正在服务全球1万多家医疗机构。根据联影医疗公布的财报,2023前三季度,联影医疗海外营收占总营收17.8%,同比增长近40%,远高于中国市场增速。

在波兰,超过200台高端影像设备正服务当地多家医疗机构;在意大利,2023年联影医疗全真数字PET-CT入

意大利顶尖综合性大学附属医院;在印度,联影医疗数字PET-CT新增占有率2021年已居首位;在拉美地区,2023年阿根廷迎来该国第一台数字PET-CT就是联影“智造”……截至今年9月底,联影医疗已在全球成立30个区域总部与子公司,销售网覆盖全球主要发达市场及新兴市场。

值得一提的是,联影医疗专利申请广泛覆盖全线产品。截至2023年10月底,联影医疗海内外累计提交专利申请数为7900余项,其中发明专利6500余项。

“出海过程中,我们能借助的最重要力量就是创新。依靠创新,让先进医疗技术造福更广阔人群。”张强说,联影成立之初的目标就是成为世界级医疗科技创新引领者,“我们希望逐步培育一个全球化的创新生态,聚焦创新源头,做世界级的创新”。