

前沿

在“热辣滚烫”背景下热议大模型的技术与应用——

从“顶天”“立地”到“你中有我，我中有你”

本报记者 邵阳 叶薇



■ 随着人工智能的发展,大模型应用越来越成熟

本报记者 陶磊 摄

大模型,无疑是2024世界人工智能大会的“宠儿”。目前上海已有34款大模型过审,产生了制造、金融等垂类领域应用。全国首个大模型创新生态社区“模速空间”落地徐汇,吸引近80家大模型企业入驻。如果说去年大会,人们还在探讨要让技术“顶天”,要使应用“立地”,那么今年,技术与应用之间,更能感受到“你中有我,我中有你”的紧密。

在本届大会的科学前沿与产业发展两场主论坛,以及多场主题论坛的“未来之声”中,都传递出上海这座创新之城,也正以前所未有的姿态引领全球大模型浪潮,为全球人工智能研究与应用提供了肥沃土壤和无限可能。

“通专融合”是发展路径

如何实现通用人工智能(AGI),让人工智能像人一样可以做得多事,不再局限于某些特定领域?在2024世界人工智能大会科学前沿主论坛上,上海人工智能实验室主任、首席科学家周伯文提出“通专融合是通往AGI的战略路径”。在他看来,建立在AI大模型强大泛化能力上的专业能力,是“人工智能皇冠上的明珠”,与目前的大模型相比,这种系统可以更高效率、更好地适应并解决现实世界中的复杂问题。

不久前,大模型对中国高考“跃跃欲试”。还真实举办了一场“大模型高考”,比试了一番各家能力。结果,大模型“考生”在语文、英语上表现还不错,可数学“集体翻车”,连及格都挺困难。稍感欣慰的是,上海人工智能实验室研发的“书生·浦语”2.0文曲星,在数学成绩上排名第一,超过了GPT-4o。在周伯文看来,“书生·浦语”的“险胜”,显示出“通专融合”研发路径的潜在价值。作为一种新的人工智能发展范式,“通专融合”将大模型的泛化能力与专业能力紧密结合,能打造出完整的技术体系。他认为,这个体系包含三层重点工作:基础模型层、融合协同层、自主进化与交互层——

在基础模型层,科研团队专注于高效地构建大模型的通用泛化能力,尤其是高阶能力,如数学、因果推理等。通过高质量数据的清洗和合成,研发高性能训练框架,高效的模型架构。

在融合协同层,科研人员要将泛化性和专业性有效地结合起来。上海人工智能实验室正在采用多路线协同的算法和技术,构建比肩人类优秀水平的专业能力。通过这些技术,AI能够在复杂环境中做出决策,将复杂任务分解为更易管理的子任务,制定行动计划,并有效地协调多个智能体,以实现群体智能涌现。

在自主进化与交互层,人工智能的自主探索和反馈闭环十分重要。人工智能系统要在真实或仿真世界中自主地收集数据、学习并适应环境,通过与环境的交互获得反馈。这些反

馈对于AI自我进化至关重要,自主进化与交互层使AI能进行“具身”自主学习,最终对“世界模型”有更深刻的理解并与之交互,具备完成开放世界任务的能力。

“在科学发现领域,通专融合也有巨大的潜在价值。”周伯文表示,通过通专融合,AI能掌握科学知识、分析实验结果、预测科学现象,进而在反思的基础上,拥有越来越强的提出科学假设的能力。

“超级能干”其实更重要

百度创始人、董事长兼首席执行官李彦宏是世界人工智能大会的老朋友。在2022世界人工智能大会上,他曾提出,AI的技术发展路线,发生了方向性的改变,从过去的辨别式人工智能,转向了未来的生成式人工智能。当年年末,Chat GPT问世,“恍如隔世,感觉整个世界都变了,人工智能可以说颠覆了绝大多数人的认知。”

在2024世界人工智能大会产业发展论坛上,李彦宏的观点是:大模型的重点还是“卷应用”,“没有应用,光有基础模型,不管是开源还是闭源都一文不值”。

人工智能时代,“超级能干”的应用比只看DAU(日活跃用户量)的“超级应用”更重要,“我们要避免掉入‘超级应用陷阱’,觉得一定要出现一个10亿DAU的App才叫成功,这是移动时代的思维逻辑。”他认为,只要能对产业和应用场景产生大的增益,整体价值就已大于移动互联网。

他以快递行业举例,利用大模型能力处理订单,快递公司做到了“一张图、一句话寄快递”,不再需要其他烦琐流程,时间从3分多钟缩短到19秒。“而且,90%以上的售后问题,也都由大模型来解决,效率提升非常显著。”

在AI应用的发展方向上,李彦宏最看好智能体。他表示,随着基础模型的日益强大,开发应用也越来越简单,其中,最简单的就是智能体,“只要用‘人话’把工作流说清楚,再配以专有知识库,即可以做出一个很有价值的智能体,‘比互联网时代制作一个网页还简单’。他认为,医疗、教

育、金融、制造、交通、农业等各行业领域,未来都会依据自己的场景和特有经验、规则、数据等等,做出各种智能体,将会出现数百万量级的智能体,形成庞大生态。

“搜索是智能体分发的最大入口。”李彦宏举例说,高考后,有大量考生填报志愿,需要选学校和专业,会遇到各种问题,这时,百度的高考智能体就可回答考生的各种疑问,“高峰时期,百度的高考智能体每天要回答超过200万个考生问题,总共只有1000万考生,说明一天当中有大比例的人在利用这个智能体”。

应用助推AI“超级时刻”

商汤科技董事长兼首席执行官徐立一直觉得,人工智能行业“热辣滚烫”,像GPT带来的聊天式的应用,Sora带来的视频上的震撼,但它还没有到“超级时刻”——因为没有真正走进一个行业的垂直应用中,引起广泛的变化。

不过这几天,徐立的想法有点儿变化。他的中学老师不停在群里询问,说怎么样用人工智能去写文案、生成祝福的图片,发到他的退休群里。徐立忽然间意识到,“超级时刻”和应用是互相成就的,“只有‘超级时刻’带来的认知变化,才能推动这样一个应用。所以倒推回来,如果用应用来支撑,那么我们现在这个时刻就是‘超级时刻’。”

徐立告诉大家,应用是决定这个时代是不是人工智能“超级时刻”的一个关键。

他打趣说,现在大模型某种程度上只是记忆库——只是背下了所有的知识点,仅有的一点智能实际上来自于互联网数据背后带有的高阶逻辑思维链。“如果要推动人工智能‘超级时刻’的到来,需要大模型可以展现出卓越的深度思考能力。那么合成的人工数据,特别是高级思维的数据往往是非常重要的。所以越是有应用的场景,才能形成高质量数据的核心。”徐立分析。

他还认为,实时的交互性能带来流畅的体验,亦是推动“超级时刻”以及应用变化的核心之一——在这一点上,端侧是非常重要的突破点。

商汤科技今年以来着重推动模型的端侧优化,使实时交互更加流畅,“用好端和云两部分的计算资源,才能够成为一种全自然的交互模式。”

此外,推动“超级时刻”很重要的一点是可控。因为大模型如果不具备可控性,那么它作为工具,本身能够带来的效能提升就非常有限,“你不需要做得很好,但你需要知道你哪里做得不好,有了这样一个边界,实际上才能做到真正的可控技术以及可持续的技术发展。”

开放行业场景占先机

华为常务董事、华为云CEO张平安表示,我国人工智能发展离不开算力基础设施的创新,并且要敢于开放行业场景,让人工智能在行业应用上领先。

“我们不能把AI基础设施依赖于是不是有最先进支撑的AI芯片。如果没有了,我们就没办法在AI上领先,这个观点必须要摒弃掉。”张平安介绍,华为创新的方向是将端侧的AI算力需求通过光纤和无线网络释放到云上,通过端云协同获得无缝的AI算力。通过云侧的算力,让端侧既保持了丰富的功能,又极大地降低了功耗和对芯片的依赖。

中国拥有千行万业的业务场景,也拥有全球最大的软件创新人群。张平安指出,中国的人工智能发展道路,追求的应该是在行业领域构筑大模型的全球领先地位。如果各行各业都积极拥抱AI,积极地开放行业的业务场景,中国很有机会在toB(面向企业客户)领域构筑起全球的领先优势。

在6月底举办的华为开发者大会2024上,华为盘古大模型5.0正式发布,在“全系列、多模态、强思维”三个方面进行了全新升级,在自动驾驶、工业设计、建筑设计、具身智能、数字内容生产、高铁、钢铁、气象、医药等领域,提供了更加丰富的创新应用和落地实践。以钢铁行业为例,目前盘古大模型已经在宝钢的一条热轧生产线上线,预测钢板精度提高5%,每年有望多生产2万吨钢板,增收9000多万元。



■ 参展者在合合信息展位尝试敦煌遗书合成样本的文字修复,亲身体验AI古籍修复模型
本报记者 陈梦泽 摄

■ 商汤科技Vimi可控人物视频生成大模型现场演示
本报记者 陶磊 摄

面向生命科学、物质科学、大气科学三大领域——

“上智院”合作构建科学数据平台亮相

本报讯(记者 张炯强)昨天,在2024世界人工智能大会“人工智能:科研范式变革与产业发展”论坛上,上海科学智能研究院(简称“上智院”)合作构建的、面向三大领域的科学数据平台亮相。

在科学大模型中,数据的质量和规模是决定模型效果的关键。当前,海外的科学数据库非常重视科学数据的整合范围、数据质量以及对历史数据的整理与汇集。然而,由于建设时间短、基础设施缺乏,现有科学数据库仍普遍存在规模小、质量参差不齐、数据不互通等问题。为此,上智院主导建设了一套涵盖多模态科学数据的科学数据平台,具备从数据采集、加工到管理和建模的全链路能力,保证数据的高效加工、可信和安全互通。该平台基于分布式并行计算将数据处理效率提升了10倍,并通过制定安全互通规范、开发数据安全工具以及全流程记录和溯源数据操作行为,实现了数据互通的有规可循、有器可用、有迹可查。

基于这一科学数据平台,上智院与合作伙伴构建了面向生命科学、物质科学、大气科学三大领域的高质量科学数据集。

面向生命科学的siRNAOD3数据集,是首个面向药物设计的siRNA修饰序列数据集,具有规模大、信息全、时效强的特点,能够加速创新药物的研发。

面向物质科学的QO2Mol数据集,收录了12万个来自真实化合物数据库的有机分子和超过2000万个构象的数据集,涵盖10种不同元素,将推动物质科学基础模型的研发,促进新材料、新能源和新药物产业的发展。

面向大气科学的FMWF数据集,由高精度大模型中期天气预报数据构成,具有精度高、预报时间长等特点,将为交通运输、能源调度、防灾减灾等行业的发展提供支持,提升社会和经济效益。该数据集由伏羲中期天气预报大模型产生,包含地面和高空多个气象要素,总数据量超过2TB。

随着科学数据平台的发布,全球科学数据生态联盟也正式启动。首批联盟成员单位包括中国电信股份有限公司、中远海运财产保险自保有限公司、上海临港新片区跨境数据科技有限公司等单位。该联盟旨在通过政府、企业、高校、研究机构等多方合作,构建全球化、多领域的科研大数据资源开放与共享平台。

“上智院”和复旦大学共同主办的第二届世界科学智能大赛,也同时宣布启动。大赛设置百万奖金,在上智院·天池平台上向全球发起选手招募,共同探索科学智能前沿领域。今年大赛新增社会科学、逻辑推理两大赛道,推进更广泛的科学智能创新发展。

当天文遇上AI——

仰望星空,寻找“另一个地球”更有谱了

“搜寻系外地球是回答‘地球是否孤独’的下一步,这也是天文学研究者的共同梦想。”昨天下午,中国科学院上海天文台研究员、讲座教授葛健,在2024世界人工智能大会“人工智能赋能中国空间天文专题”论坛上介绍了中国天文学家的AI+探索。

一直以来,全球天文研究者始终在仰望星空寻找“另一个地球”。国际上已运行的空间搜寻系外行星的卫星全都采用成熟的凌星法,已经发现了约70%的已知系外行星(5000多颗)和60%的地球大小的行星,是目前最有效和最精确的搜寻系外行星的方法。但是,目前仍未找到“地球2.0”。中国科学院空间先导背景型号研究支持的地球2.0空间巡天(简称ET)也是使用技术相对成熟的凌星法来搜寻地球2.0的空间卫星项目。葛健是“地球2.0”研究项目创始人和首席科学家,他介绍,这项太空任务的科学目标是监测超过120万颗类似于

太阳的恒星,以探测“地球2.0”的存在,预计发现近三万颗系外行星,是目前已知数目的6倍;ET的高精度测光数据将是开普勒数据的至少10倍,极大地拓展了系外行星的参数空间。

“利用人工智能技术我们已经取得许多可喜进展。”他举例说,利用GFPC全新算法,团队在开普勒数据中发现5颗超短周期亚地球系外行星,而且搜寻速度提高了一千倍,精度和完备性也有所提升;近期利用AI在其他数据中发现宇宙早期破纪录的中性碳吸收,这项研究有助于探秘行星的演化,“我们发现行星演化比我们想象的要复杂和丰富得多。”未来,人工智能赋能的空天研究将对系外行星的各种种群的完备样本,在尝试“地球是否孤独”的全球问题回答中,发出中国天文学家的声音。

围绕AI驱动天文研究、AI与系外行星发现、AI与空间高能天文、AI与空间天文数据、AI引领卫星

制造新范式等多个维度,海内外专家在论坛开展了主题演讲和圆桌讨论。麻省理工学院卡夫利天体物理与空间研究所物理助理教授安德鲁·范德堡认为,随着数字探测器和强大计算技术的出现,宇宙探索将变得更加高效,科学家得以通过更详细、更数字的方式识别系外行星,这些发展改变了人们对宇宙的理解,扩大了发现太阳系以外潜在宜居世界的可能性。紫金山天文台副台长、暗物质间接探测的相关物理研究组首席研究员范一中也在论坛上发出理性呼声,他以暗物质粒子探测卫星“悟空”为例,在“悟空”八年半的稳定运行过程中积累了庞大数据,AI技术与传统方法各有千秋,他认为在空天研究领域,应同步考量效率、成本、精确性等多维要素,要与AI结合探寻最佳研究工具。

本报记者 易蓉

相关链接

世界人工智能大会 掠影

SAIL大奖

4日,“卓越人工智能引领者奖”(SAIL奖)揭晓。上海人工智能实验室的DeepLink人工智能开放计算体系;中国南方电网有限责任公司的新型电力系统智能仿真器“取电”;墨奇科技(北京)有限公司的MyScale AI数据库;施耐德电气(中国)有限公司的能源管理和工业自动化领域的EcoStruxure AI引擎;清华大学卢志的虚拟扫描光场显微镜实现鲁棒的高时空分辨率三维成像这五个项目斩获这一世界人工智能大会最高奖项。

镇馆之宝

比起高深的算法对垒、高端的大咖对话,世界人工智能大会每届“镇馆之宝”都是离普通市民最近的应用。4日,2024世界人工智能大会八大“镇馆之宝”揭晓:阿里云通义灵码、联想YOGA Book 9i AI元启、基于蚂蚁百灵大模型的“支付宝智能助理”、商汤科技Vimi可控人物视频生成大模型、特斯拉克越野旅行车、智谱AI基座大模型、智能飞行器展示区,以及一个神秘惊喜宝藏——“十八金刚”人形机器人先锋阵列。

创新指数

4日,《2023全球人工智能创新指数》发布,这是该指数第四次在世界人工智能大会上揭晓。指数显示,美国和中国稳居第一梯队,得分明显高于其他国家,二、三、四梯队整体格局比较稳定。2023年美国以74.71的总分大幅领先,中国总分为52.69分,排名第二,比排名第三的英国高出近15分。指数分析认为,我国在人才培养、科研产出和产业发展等方面取得积极成效,但数据开发利用、原始创新等方面仍存在不足。

古籍修复模型

以AIGC技术为代表的数字化修复手段为延续古籍生命力提供了新的思路。合合信息旗下扫描全能王携手华南理工大学团队,将AIGC技术应用于敦煌遗书残卷的图像数字化修复上,共同打造了AI古籍修复模型。大会期间,合合信息展位开放了敦煌遗书合成样本的文字修复体验项目,公众可在不同位置扫描样本卷轴,见证AIGC技术通过字形修补、褪色修复、背景补全等方式,让残损的古籍焕发新的生命力。本报记者 叶薇 邵阳 金志刚