

治理

安全风险频被提及 “善治”共识期待 “善智”回音

本报记者 易蓉

AI换脸越发逼真,大模型一本正经胡说八道也难辨真假,新技术带来生产效能的提升却也带来无法忽视的就业冲击……伴随人工智能以越来越快的速度为人类生活与生产方式带来创新和益处,其安全风险也越来越被频繁提起。“以共商促共享 以善治促善智”是2024世界人工智能大会的主题,这两天的各场论坛上,无论是科学家、研究者还是政策制定者、企业家都发出了呼声。这些声音,带着相似的隐忧,传递着“善治”的共识,汇集在一起,形成了“共商”的共振,期待着“善智”的回音。



奇妙AI数字分身展示

本报记者 陶磊 摄



铁氧机器人现场冲泡功夫茶 本报记者 陈梦泽 摄

新理念

将安全作为「全球公共产品」 本报讯(记者 易蓉)昨日,《人工智能安全作为全球公共产品》研究报告在徐汇西岸发布,首次提出将“人工智能安全作为全球公共产品”的理念,提倡各方应致力于建设具有公共性的安全知识、能力、资源,统筹安全与发展双重目标,推动该领域的理论与实践探索。 报告着眼于解决安全知识与资源在全球范围内的共享不足问题,鼓励各国、各机构加强合作,共同提升全球安全治理水平,确保人工智能技术的发展能够兼顾安全与发展,旨在为应对跨国安全威胁、推动安全治理标准的统一与协调、促进人工智能技术的国际合作提供了行动指南。 报告指出,实现“人工智能安全作为全球公共产品”理念的关键原则包括公共原则、科学原则、责任原则、敏捷原则,并结合四个原则发起推动“人工智能安全作为全球公共产品”的行动倡议,以促进多方主体在人工智能安全领域的持续改革与合作。

掠影

智能出行

上午,由隧道股份城建设计集团开发的“智捷易行”无人驾驶巴士亮相2024世界人工智能大会周边。“智捷易行”不仅是上海首个面向重要场景动态接驳并融合人工智能技术的交通创新应用,更在交通关键领域实现了技术创新与服务模式的双重突破。 体验者只需在手机小程序上预约下单,AI算法将自动采集附近所有乘客的出行需求,并根据其起始位置、到达目的地,即时计算接驳线路、调度接驳巴士,让一辆接驳巴士满足周边多人的出行需求,实现公交线路的“拼单定制”,服务市民出行最后一公里。

大模型

昨天,多家科技互联网厂商在2024世界人工智能大会发布大模型产品。 百度智能云宣布文心大模型4.0 Turbo面向企业客户全面开放;百度智能云千帆大模型平台针对大模型工具链和AI原生应用发布十余项全新升级。 阿里云无影云电脑宣布推出首款云原生大模型AI助手——“晓晓”,专为在云电脑上学习、办公等场景设计。 蚂蚁集团百灵大模型已具备能“看”“听”“能”“说”“会”“画”的原生多模态能力,可以直接理解并训练音频、视频、图、文等多模态数据。 商汤科技发布国内首个所见即所得模型“日日新5.0”,交互体验对标GPT-4o,实现全新AI交互模式。 本报记者 杨硕 金志刚 郗阳

专访

上海人工智能实验室领军科学家欧阳万里:

上海竞逐“模都”有优势 “杀手级”应用或“现身”

2024年,人工智能领域风起云涌,模型架构、训练数据、多模态、超长上下文、智能体发展突飞猛进。大模型的演进路在何方?为此,本报记者专访了上海人工智能实验室领军科学家欧阳万里教授。 欧阳万里主要负责AI for Science(科学智能)方向的研究,领衔气象预报大模型“风乌”及“风乌GHR”的研发。在专访中,欧阳万里表示,“通专融合”是建立在过去一段时间大模型研发基础上的思考和总结,是更适合我国大模型的发展路径。

走“通专融合”发展之路

Q:如何看待中城大模型发展的生态?上海在“模都”竞争中,有哪些优势? 欧阳万里:上海将人工智能作为三大先导产业之一,着力建设人工智能“上海高地”,有着非常好的基础。第一,上海有坚实的科研基础优势,上海人工智能实验室和顶尖高校,从科研基础上支撑起人工智能的进步;第二,上海有诸多头部人工智能企业聚集,他们间的彼此合作与竞争,会带动上海人工智能

调查

提升人工智能在教育场景的应用 《中小学教育数字化转型区域进展报告》发布

教育数字化是未来发展的方向,绝大多数校长和教师持积极欢迎的态度,但目前在向数字化转型过程中,也相继出现了一些“水土不服”的现象。今天上午发布的《中小学教育数字化转型区域进展报告》提出,要着力提升数字化技术在具体教育场景中的有效应用。 此份报告主要是基于华东师大上海智能教育研究院对沪苏浙三省十个地区、上百所学校、近万名教师的问卷调查结果。绝大多数教育工作者认为,教育数字化的主要成效体现在优化课堂教学效率,促进教育公平、拉近师生距离等方面。但在实际应用中,数字化进校园也给日常教学带来了不少新问

题。华东师大教育学部主任袁振国教授说,在教育以外的几乎所有领域,数字化程度越高,人的参与程度就越低。但是,教育不是这样的,教育活动是人与人的联系,数字技术越发达,对人运用机器的能力要求就越高,人的能力越高,机器的作用发挥得就越好。 在调研过程中,长三角地区的校长们普遍认为,随着数字化教育的普及,教师可以在实体和虚拟两个空间组织学生开展教学活动,学生也可以随时随地通过虚拟技术参与到教学活动之中,等于拆除了学校的围墙,突破了实体空间的限制。为此,报告提出在人机协同的智能化阶段,“机”不应仅限于计算

治理之“变”

“与传统技术革命不同,人工智能带来的治理冲击具有特殊性。”上海交通大学国际与公共事务学院特聘副教授贾开解释,人工智能引致影响的速度可能远快于传统技术革命,这在很大程度上缘于人工智能技术与整个社会数字化转型进程的互促效应,技术迭代速度快,其嵌入社会、影响社会的过程也极大缩短;人工智能产生影响的结果可能迥异于传统技术革命,就业替代效应或将大于就业创造效应;此外,人工智能产生影响的领域与机制可能是高度不确定的,“尽管我们大致能框定人工智能安全风险类型,但风险产生的时间地点、对象范围、因果过程等都是未知的,甚至不能被提前预知。” “人工智能技术和应用的不确定性与复杂性,使得从业者和监管者不得不面临‘共同无知’,治理问题的界定与治理方案的探索始终处于动态变化之中,”清华大学文科资深教授、苏世民书院院长薛澜更是冷静地强调了人工智能治理的新挑战,“同时,治理主体在文化、伦理、价值观等各个维度都存在较大异质性,风险偏好的差异使得达成治理共识较为困难;治理与发展要求同步演化,以避免出现‘监管洼地’,但治理更新的速度却往往落后于技术创新和应用步伐。” 清华大学公共管理学院院长特聘教授赵静也观察到AI治理议题正快速变化,她认为这与人工智能技术本身的快速成长性和应用不确定性密切相关,尤其是随着人工智能技术在经济社会中的全面应用提速,除了对各类潜在风险的关注外,技术发展和应用对人类普遍生存的安全影响被提上日程,“现在大家更关心如何构建一个安全可信的人工智能发展框架,基础理论、治理思路,从而应对和解决前沿人工智能发展背后的基本安全问题。”

应变之策

阿里研究院执行院长袁媛认为,科学家、企业、政府已形成了一种自觉的共识,逐渐形成了

磨合共治的协作模式,大家共同面对问题,合作形成公共知识,再各自改善。 这种共识并非与生俱来。袁媛提到,在电商时代应用判例式人工智能技术时,无论平台、商家、用户、社会都需要安全交易,需要建立“生态信赖”,阿里巴巴就认识到用算法参与治理的重要性,从企业内部启动了治理进程。“那时算是治理初期,市场上的企业没有太多共识,治理模式也相对简单,企业处于‘被规制’的角色。而现在,随着公众认知提升,产业接受度提高,政府期待日渐成熟,各方将逐渐对‘善智’形成更为健全的共识。” 升级治理模式并非易事。上海国际问题研究院研究员、网络空间国际治理研究中心秘书长鲁传颖表示,20世纪80年代美国学者曾提出科林格里奇困境——政策制定者面对快速发展的技术存在“太早治理因噎废食,太晚治理脱缰失控”的困境,“所以建立沟通模式,寻找恰到好处的时机,以敏捷治理等最小成本最大收益的方式开展良好治理是很重要的,目前中国正在寻找好的模式、机制、节点。” 贾开也认为,人工智能治理不能遵循传统技术治理范式按照预先确定规则来加以管控,而更加需要依赖政、产、学、研的共同探索,并在此过程中积累治理知识、能力、资源。此次会上,上海交大联合、徐汇区政府及研究机构、媒体等21家单位共同发起“西岸对话:人工智能安全与治理对话网络”,旨在打破壁垒,在共商共建共治过程中共同探索、共同进步,通过对话机制,建设平台渠道、素养提升等常态工作,促进各方在人工智能安全与治理领域的紧密合作,以期在快速发展的时代中更好地统筹发展与安全。 贾开认为,类似“西岸对话”的共商共建共治机制才能实现AI治理升级的价值,“在单独一方都不知道人工智能治理风险内涵及其应对机制的前提下,只有保持开放对话的治理关系,才可能形成科学共识,找到合适的治理方法。”

中国呼声

人工智能治理是全球议题,当前国际社会

对人工智能治理既有建立实体多边监管的联合国模式,也有部分欧美国家开展的小范围联盟式治理探索。鲁传颖认为,这场实践中,中国正全方位积极参与人工智能国际治理,并设计了一条不同于西方国家的治理思路。去年中国政府发布《全球人工智能治理倡议》,围绕人工智能发展、安全、治理阐述人工智能治理中国方案。鲁传颖指出,中国方案强调人工智能发展必须坚持“以人为本”理念,提倡更为和谐、平衡的人类关系视角,倡导构建一个平衡各方利益、共同应对人类挑战的全球人工智能治理新模式。 “不同模式在当前非常需要对话与互补,因为人工智能的影响是‘生态性’的,任何一方都不可能完全排除他方,所以存在形成合作共识的基础。”上海交通大学政治经济学院名誉院长、广州粤港澳大湾区研究院理事长郑永年谈到,“当前尽管仍然存在分歧,中美、中法之间都在人工智能领域展开积极对话,只有在坚持开放、对话、合作基础上,我们才有可能提升人工智能国际治理水平,在未来能形成具有约束力、执行力的合作共识。” 大会期间,科学家从技术视角提出了一些新的治理理念和路径。“应将人工智能治理视为一类‘知识’和‘行动’并进的公共产品。”薛澜提出,这类公共产品具有非排他性,即所有从业者都应服从相关规制;同时,这类公共产品也具有非竞争性,即治理方案的扩散和运用不影响他者对该方案的享用。 斯坦福发布的《人工智能指数报告》中,中国连续两年成为AI信任度最高的区域。袁媛认为,这说明目前中国的AI治理和企业自律都有较好的表现。她期待接下来,企业、政府、研究机构等第三方共同制定安全规范,媒体公众参与监督,大家一起把事情做好。“市场是最好的大考场,君子之过如日月之食,谁出错都在聚光灯下。”她还建议,为营造更为向善的生态,应采用激励与监管相融的方式,对安全治理方面投入更多的企业予以激励。



“玄武”上肢操控作业展示 本报记者 陶磊 摄