

# 科技伦理审查提速 AI“变坏”设防



ChatGPT与GPT已走进大众的视线,我国众多的互联网大厂也正火速切入这一赛道。不可否认,GPT等人工智能应用极大提升了生产力、便利了大众的生活,但其也是一柄双刃剑,不久前甚至还出现诱导人类自杀的事件,由此产生的科技伦理问题也引发了全球的关注与担忧。

令人欣慰的是,我国已开始筹谋提前化解人工智能的科技伦理风险。日前,科技部公开征求对《科技伦理审查办法(试行)》(以下简称《办法》)的意见,其中明确提出,从事人工智能的单位,研究内容涉及科技伦理敏感领域的,应设立科技伦理(审查)委员会,部分算法模型、自动化决策系统需要专家复核。分析认为,虽然伦理审查短期内会影响产品商业化的效率,但从长期来看,有助于人工智能行业可持续的健康发展,让我国在这一领域形成可持续的全球竞争力。

## 设立审查委员会

科技部本次发布的《办法》提出,从事生命科学、医学、人工智能等科技活动的单位,研究内容涉及科技伦理敏感领域的,应设立科技伦理(审查)委员会。

此外,在《办法》的附录中还提到了“需要开展专家复核的科技活动清单”,其中包括:具有舆论社会动员能力和社会意识引导能力的算法模型、应用程序及系统的研发;面向存在安全、人身健康风险等场景的具有高度自主能力的自动化决策系统的研发。

“ChatGPT与GPT等应用越来越广,也与我们的生活越来越近,我们在科技伦理方面也遇到了前所未有的挑战。”浙江理工大学数据法治研究院副院长郭兵认为,《办法》比较大的亮点在于,进一步对更广泛的科技伦理审查问题做了规定,并引入了专家复核制度。

事实上,在ChatGPT诞生之前,我国就已经出台与人工智能伦理的相关规范和标准。2021年9月25日,国家新一代人工智能治理专业委员会发布了《新一代人工智能伦理规范》,旨在将伦理融入人工智能全生命周期,为从事人工智能相关活动的自然人、法人和其他相关机构等提供伦理指引,促进人工智能健康发展。

“目前我国科技伦理的审查,也正处于传统的完全自律逐渐演变成法定的义务。”郭兵介绍说,在2021年11月1日起施行的《个人信息保护法》中,已经明确规定了科技平台有义务设置独立的第三方机构对其个人信息的处理活动进行监督;其第五十八条则规定,提供重要互联网平台服务、用户数量巨大、业务类型复杂的个人信息处理者,应当按照国家规

定建立健全个人信息保护合规制度体系,成立主要由外部成员组成的独立机构对个人信息的保护情况进行监督。

“但《个人信息保护法》并未要求企业邀请何种外部人员,而科技领域、AI领域专业性较强,《办法》通过法定的专家参与机制的设计,能够更好地保障人工智能伦理审查的效果。”郭兵进一步解释说。

## 伦理问题浮现

人工智能应用技术会带来哪些科技伦理问题?在中国数实融合50人论坛智库专家、商务部研究院电子商务研究所研究员洪勇看来,人工智能可能面临对自然资源的消耗、产品更新换代产生的各种垃圾等环境伦理问题,隐私泄露、歧视、失业、安全风险等个人伦理问题,以及推理能力、可解释性等技术伦理问题。

“从个人的角度来看,人工智能的应用还可能会侵犯人格尊严。比如算法所引发的歧视风险,还包括公民的隐私和个人信息泄露的风险。人工智能也可能引发劳动者权益方面的争议,引发劳动者的‘饭碗焦虑’。”郭兵说,人工智能甚至还可能危及个人的生命与健康。

正如郭兵所述,虽然ChatGPT与GPT诞生不久,但已有值得引起关注的案例。

有媒体援引比利时媒体La Libre 3月28日的报道,一名比利时妇女声称,她的丈夫在与AI聊天机器人Chai交谈了六周后自杀身亡,背后主要原因是Chai鼓励用户自杀。据悉,聊天机器人Chai是基于GPT-4替代方案研发的产品。

在此次事件中,Chai在法律层面将承担怎样的法律责任、又该如何追责?“在这方面目前仍备受争议。”河南泽槿律师事务所主任付建认为,人工智能技

术的程序是开发人员编写的,但决策是由机器做出的,如果出现重大失误造成损害后果,责任谁来承担仍存在争议,同时,人工智能能否对自己做出的决策进行道德判断也是一个值得商讨的问题。

## 全球千余专家发声

也许因为身处人工智能领域的最前沿,最为了解GPT的运行逻辑以及“涌现”的概率,在Chai事件发生后的第二天,全球范围内众多专家疾速发声。

3月29日,包括马斯克、图灵奖得主约书亚·本吉奥、苹果联合创始人史蒂夫·沃兹尼亚克在内的1000多名人工智能专家和行业高管联名签署了一份公开信,呼吁暂停开发比GPT-4更强大的AI系统至少6个月,称其“对社会和人类构成潜在风险”。

事实上,人工智能存在的伦理问题风险也引起了联合国教科文组织的注意。近日,联合国教科文组织发表声明,呼吁各国政府尽快实施该组织通过的首份人工智能伦理问题全球性协议——《人工智能伦理问题建议书》。

在最新发布声明中,联合国教科文组织对科技创新引发的伦理问题表示关切,尤其是歧视、性别不平等等问题,此外还有打击虚假信息、隐私权、个人数据保护等。行业自律显然不足以避免这些伦理危害,因此,《人工智能伦理问题建议书》提供了工具,以确保人工智能的发展遵守法律,避免造成危害,并在危害发生时启动问责和补救机制。

“从社会和经济层面来看,企业一般利益至上,如果各家科技巨头通过研发的大型语言模型,利用‘算法暗箱’,谋一时之利,就有可能有损整个社会利益和国家利益,给国家安全带来风险。”

郭兵补充说。

## 短痛与长痛的选择

“《办法》进一步明确政府、科研机构、科技企业等责任主体,建立伦理高风险科技活动的清单制度,有效避免人工智能技术被不当滥用,推动人工智能技术健康可持续发展。”中国电子商务专家服务中心副主任、知名互联网专家郭涛认为。

但不得不说,此次的《办法》实施后将将对行业产生重大影响。

“对人工智能可能带来的问题提前加以防范,进行伦理审查会对行业产生重大影响。此前更强调数字技术、人工智能技术在效率上带来的变革,而忽略技术对社会、对个人产生的伦理问题。随着科技伦理越来越得到重视,将会对人工智能在医疗、教育、无人驾驶等领域的应用产生影响。”北京市社科院研究员、中国人民大学智能社会治理研究中心研究员王鹏对北京商报记者表示。

郭兵就《办法》给人工智能行业带来变化进行了具体化的解读,“随着人工智能伦理审查越来越严,会增加企业的成本,尤其引入了专家复审要求后,会影响产品或技术市场化的效率,研发出来的某个算法模型不能立马投入商业化的应用。”

但郭兵认为,另一方面,审查的严格将有利于人工智能产业的可持续发展。“如果企业研发的产品投入市场化之后产生伦理方面的巨大争议,面临巨大法律风险,被直接叫停,带来的不利影响会比人工智能伦理审查带来的影响更大,从长期来看,加强伦理审查有利于人工智能企业的可持续健康发展。”

北京商报记者 金朝力 袁泽睿

## X 西街观察 Xijie observation

### 现实反水 ChatGPT祛魅

陶凤

从无所不能的颠覆者,到众叛亲离的搅局者,尚未大展拳脚的ChatGPT,正被层出不穷的负面啃食。

马斯克与OpenAI的关系也经历了一条类似的抛物线。一开始积极投资入股,后来反目成仇,到最后炮轰ChatGPT,让世人警惕其不可控的风险。

而三星在引入ChatGPT不到20天,就被曝出3起数据外泄。按照韩媒的说法,因为员工直接将企业机密信息输入其中,导致相关内容进入学习数据库。

如今,连ChatGPT自己也不堪重负。由于需求量太大,OpenAI宣布暂停Plus的销售,未来何时开放还没有明确的说法。

“只有当我们确信它们的影响是积极的并且风险是可控的时候,才应该开发强大的人工智能系统。”发生在ChatGPT身上的激烈矛盾,马斯克的“忌惮”言中了一部分。

随着ChatGPT横空出世,算法的强大之处,再次让人叹为观止。虽然AI取代人类还为时尚早,但它对人的全面围猎却无孔不入。

ChatGPT口碑逆转,其实来自于我们对它的认识从最初的惊喜,回归到了前所未有的理性。具体到各行各业的人,则是在这场飓风中感到吹向自己的凉意。

那些呼吁政府介入的声音,其实不是为了保护一份工作,而是为了保护他们自己。在人工智能面前,真正的主角“人”感到了一种普遍的不安。

这种不安来自于“被控制”。有人说,它在打破一个信息茧房的同时,又构建了另外一个。ChatGPT相比人类来说拥有强大的知识储备,它无限扩展了人类的认知边界,但基于大语言模型开发而来,又将使用者最终困于信息壁垒之中。

同时,不安也来自于“被监听”。在新的信息茧房中,又通过密切的交互,摸清了每个人、每个诉求背后的信息。这些信息链条交织,让人无处遁形,甚至感到窒息的时候,大家对ChatGPT才感到恐惧。

即便ChatGPT问世不过几个月时间,当看到人工智能表现得越来越像我们自己,甚至比我们还要了解我们自己,对传统社会以及对人的心理都会带来前所未有的冲击。

无论将人类的道德规则应用于人工智能,还是该为其创建一套全新的道德准则,在现阶段都并非易事。但在人工智能的飞速成长中,这些标准和规则已经刻不容缓。

我们处于与人工智能深刻共存的时代,人工智能足够先进,可以基于自身的逻辑去表达和创造,这意味着人工智能将能够改变人类思考、感受和行为的方式。

人工智能会超越人的智能吗?如今来看,这个问题的是与非恐怕不是最重要的。最重要的是,在ChatGPT的祛魅中,我们意识到不能被超级智能主宰,尽管它们可以在具体任务中都超越我们,但却完全不顾法与德和那些真、善、美的体验。

## 轨道上的外贸复苏新信号

“要看到中国外贸所具有的竞争力、韧性和优势。”商务部国际经贸谈判代表兼副部长王受文在国务院新闻办公室于4月4日举行的新闻发布会上表示。

目前,各大国际通道的“成绩单”正显示出中国外贸的韧性:一季度西部陆海新通道铁海联运班列发送集装箱货物19.1万标箱,同比增长11.7%;中老铁路老挝段累计发送货物66.76万吨,同比增长261.65%……分析表示,各大国际通道取得的良好成绩,不仅反映出我国的经济复苏正在加速,跨国贸易正持续增长,这均显示出我国外贸的韧性与持续向好的趋势。

### 全面开花的铁路班列

“今年中国经济的复苏会给我们外贸提供推动力。”王受文指出,中国的PMI制造业采购经理人指数连续3个月位于荣枯线之上,物流协会4月3日公布的我国物流指数已经两个月连续上升,3月达到了55.5%。

“这说明国内的经济复苏加速,经济复苏对进口产品的需求产生了牵引力。”王受文表示。

数据显示,2023年一季度,西部陆海新通道铁海联运班列发送集装箱货物19.1万标箱,同比增长11.7%,继续保持强劲的增长态势。同期,中老铁路老挝段累计开行货物列车954列,同比增长112.47%,发送货物66.76万吨,同比

增长261.65%。

此外,中欧班列在一季度也取得了不错的成绩。中欧班列(武汉)共开行256列,同比增长187.6%,其中开出59列、到达197列,同比分别增长47.5%、302.2%;中欧班列(合肥)今年一季度共发运340列,同比增长49.78%;中欧班列(齐鲁号)累计开行506列,同比增长9.5%。此外,截至4月1日零时,京津冀地区共计划发中欧班列238列,同比增长6.7%。

“一季度各大国际通道取得了良好的运输成绩,反映出我国的经济复苏正在加速,跨国贸易也持续增长。”北京师范大学政府管理研究院副院长宋向清对北京商报记者表示。

### 不断增强的经贸体系

“这些成绩的取得意味着‘一带一路’沿线国家与中国经贸联系在加强,也意味着沿线国家与中国的合作空间更大。”宋向清表示。

在众多“国际经贸大通道”中,中老铁路颇具典型性。中老铁路是老挝第一条现代化铁路,其连通中国、老挝并与老泰铁路相连,是泛亚铁路中线的重要组成部分。截至4月3日,中老铁路开通运营已满16个月,期间,其货发量增长明显,货物品类也更加丰富,已从开通初期的100余种,拓展到2000余种。

“从开通运营以来其贸易品类丰富程度和规模增长来看,中老铁路目前作为‘一带一路’战略中东南亚‘黄金大通

道’的作用日益凸显,不仅可直接带动中老两国经贸关系,还能进一步加深中国与东盟成员国之间的间接贸易。”独立国际策略研究员陈佳介绍,“例如,泰国就利用中老铁路的冷链专列向中国输送过当地水果,向东盟国家展示了中老铁路分段式运输、口岸公路接驳独特又高效的运输方式。”

宋向清认为,总体上看,多条通道的成绩体现出我国正在发展出更加多元化的对外经贸市场,逐步建立起国内大循环与国内国际双循环相结合的经贸体系。“一方面,在美国和非洲国家以外,我们与东南亚、中亚、欧洲的经贸联系也在不断增强,对外经济交往的支撑点更多更牢。另一方面,除了海运、空运以外,铁路运输的作用更加突出,使我们的国际物流呈现出多元化、立体化的特点。”

### 多维立体的化险举措

外贸复苏信号不断释放的同时,内外风险压力也同样存在。“从宏观贸易板块来看,尽管中国宏观经济加速恢复企稳,但2023年中国促外贸平稳发展的内外风险压力依旧存在。”陈佳表示,从1-2月区域贸易数据尤其是出口增速来看,传统贸易强省稳外贸的压力日益凸显,如何进一步发挥促外贸接续政策的效能,充分发挥外贸强省对全国贸易板块的龙头带动作用,还需要贸易产业链各方群策群力。为此,陈佳认为,在宏观贸易层面,中国必须继续坚持

贸易结构优化,适度调整贸易伙伴权重。“由于欧美金融市场近期连环爆雷,美国经济硬着陆,欧元区经济滞胀的系统性风险依旧高企,预期短期内传统贸易三强欧美日对我国贸易拉动的负效应会持续;而东盟、俄罗斯、韩国等对我国贸易拉动正效应会放大。”

陈佳表示:“从全球贸易品大类走势来看,消费电子市场目前尚未完全迎来春天,传统机电和高新技术产品增速日益放缓,绿色能源领域则逆势增长,针对这些产业链与地区的最新变化,中国有必要在贸易结构和贸易政策方面加速创新,择机优化针对性促外贸政策措施,为贸易板块微观层面的广大中小微企业指明方向。”

“在当前全球贸易遭遇发达经济体衰退和人为脱钩断链等多重风险冲击的严峻局势下,中国的外贸板块坚持自主创新,依靠科技进步,在供应链出海、跨境贸易平台等领域不断优化,有望突破发展瓶颈。”陈佳说。

据王受文介绍,商务部相关部门已提出要做好政策保障,促进抓订单,培育市场主体,保证履约;要积极推动外贸新业态发展,稳住加工贸易;要用好开放平台和贸易规则,不断优化营商环境,持续扩大进口。

“按照部署,我国正在对外贸领域进行深入的调查研究,目的是真正找到地方特别是外贸企业、外贸行业所遇到的困难,帮助他们解决问题,为外贸的稳定发展、为经济增长做好基础工作。”王受文表示。北京商报记者 金朝力 冉黎黎