2024年轻人选择在数字空间隐姓埋名

本报记者 赵广立 见习记者 赵宇彤

你知道“momo”吗?

近年来,互联网上不知不觉涌现出一批“momo大军”,他们用着同一个昵称、顶着一模一样的头像,混迹于微信、豆瓣、小红书、微博等各大社交平台。

这是许多年轻人隐藏身份的“马甲”。是的,曾经以个性十足、特立独行为傲的那批人,现在更在意的是怎样隐匿自己在网络上的言行,而原因是只为了逃脱“算法围城”。

他们不希望“信息茧房”成为自己的“人生第一套房”,不想让社交媒体的分享成为大数据窥探的窗口,不愿意付出时间和健康的代价,却被困在一个看不见的牢笼里。

但他们何尝不知道,数字生存如同雪泥鸿爪,数字互联时代,想要雁过无痕,谈何容易!

“硬刚”算法的年轻人

不管承不承认,我们每个人都活在算法围城之中。

同一份外卖,老顾客要比新会员多付几元;同一时间的相同路程,不同手机型号的用户单价不一;当你拿起另一半的手机,居然发现在短视频平台看到的热搜评论都不尽相同……

面对算法围城,普通人有普通人的痛楚,名人有名人的烦恼。

近一年以来,农夫山泉创始人钟睒睒以及农夫山泉频上“热搜”:从产品、环保议题到个人家事,不仅农夫山泉的股价应声滑落,钟睒睒个人也遭受了前所未有的流量“集火”。

忍无可忍的钟睒睒在一场交流活动中隔空喊话字节跳动创始人张一鸣,直言有人利用算法“制造单一叙事和恶意对立”,并称这种“利用科技手段、技术能力造成的恶”比普通人造成的恶要大,“是大恶”。

“他们用算法放大情绪,把复杂的问题简单化,把不同的声音屏蔽掉。”钟睒睒说,这种行为不仅破坏了公平的舆论环境,也让公众陷入片面认知,而受害者往往都是底层民众。

在“网暴”面前,绝大部分人不是钟睒睒。面对各种算法织就的网,钟睒睒可以“隔空喊话”,普通人却只能套上“马甲”。

于是,越来越多的年轻人决定主动出击,试图“反向训练算法”,和平台正面“硬刚”。

他们中有人“赛博哭穷”,“机票太贵了我不去了”“买不起,不买了”“9毛9,我喝得起的咖啡”;有人“喊话威胁”,给某软件留言,“竟然杀熟,太过分了,我要卸载”;有人则是行动派,移除软件后故意过一段时间再重新下载,以期再次享受大额优惠;还有人利用软件生成的虚拟手机号频繁注册平台软件新号,“薅新人羊毛”。

有网友表示,有些“偏方”好像真的管用,发牢骚、抱怨价格、卸载重装,一套流程走完能省下一笔不小的费用。在小某书,随处可见诸如“骂机票专用帖”等热门经验分享。也有人觉得这一切都是徒劳,到头来还是被平台用算法“收割”。但不管怎样,他们乐此不疲,并美其名曰“反向驯化大数据”“用算法打败算法”。

“反向驯化”其实见效甚微

“反向训练算法”有没有用?《中国科学报》就此咨询了算法专家。得到的答案,恐怕要给大家浇一盆冷水。

“‘反向驯化大数据’这类做法可能仅仅对一些简单的算法有效果。”中国科学院自动化研究所副研究员、武汉人工智能研究院算法总监吴凌翔说,平台会根据用户大量的历史信息、IP地址、社会关系、手机型号等做算法推荐,如果用户不了解算法机制,很难“反向训练”。

中国传媒大学媒体融合与传播国家重点实验室媒体大数据中心首席科学家沈浩则认为,用户通过主动关闭定位、禁止后台调取通讯录等方式能起到一定的屏蔽作用,但试图通过调整标签、更换人设来“迷惑”算法,可能适得其反,新的“人设”还会出现新的“信息茧房”。

事实上,许多软件都给了用户选择取消“个性化推荐”的功能,但由于这项功能于平台而言太过重要,往往隐藏较深。

而在沈浩看来,取消个性化推荐也不能根治“信息茧房”。

“不推荐、表示‘不感兴趣’也是一种推荐。”沈浩告诉记者,算法是基于用户数据驱动的,每个人都或多或少被“困”在“信息茧房”里,只不过感知程度不同。

北京航空航天大学计算机学院教授王静远直言,自己虽然没有专门研究过用户对抗算法推荐的做法,但他对出现的这种现象并不感到意外。

“这反映了一些算法对用户信息的收集和利用有些过分。”王静远对《中国科学报》说,当平台逼得用户连头像、昵称这些基础信息都要隐藏,说明用户的一切痕迹都有可能被作为特征而提取,“用到极致了”。

在采访中,专家们不止一次提到“算法中立论”,认为算法无罪,罪在利益相关方。但是,当每一次点击、每一句评论,甚至每多停留一秒钟,这些痕迹都成了平台训练算法的“养料”;当外卖、网约车等平台被大数据操纵,吃什么、去哪里都被“读心术”安排得明明白白;那么在被浪费的时间、被挑拨的情绪、被掏走的“冤枉钱”面前,用户眼中的算法就不再是“中立”的,而是越来越大的“牢笼”。

“当一切痕迹都在利益驱使下过度商业化时,自然会有反抗。”王静远说。

“旧病未愈,又添新疾”

一边是平台利用算法精准织网,一边是越来越多的人开始觉醒与反抗。野蛮生长的算法乱象,正被社会全方位审视。

近日,中央网信办、工信部、公安部、市场监管总局四部门联合部署开展“清朗·网络平台算法典型问题治理”专项行动,重点整治“信息茧房”、操纵榜单、利益侵害、大数据“杀熟”、算法推荐等典型问题。

但如果回溯大数据兴起之时,“算法治乱”一直都有。

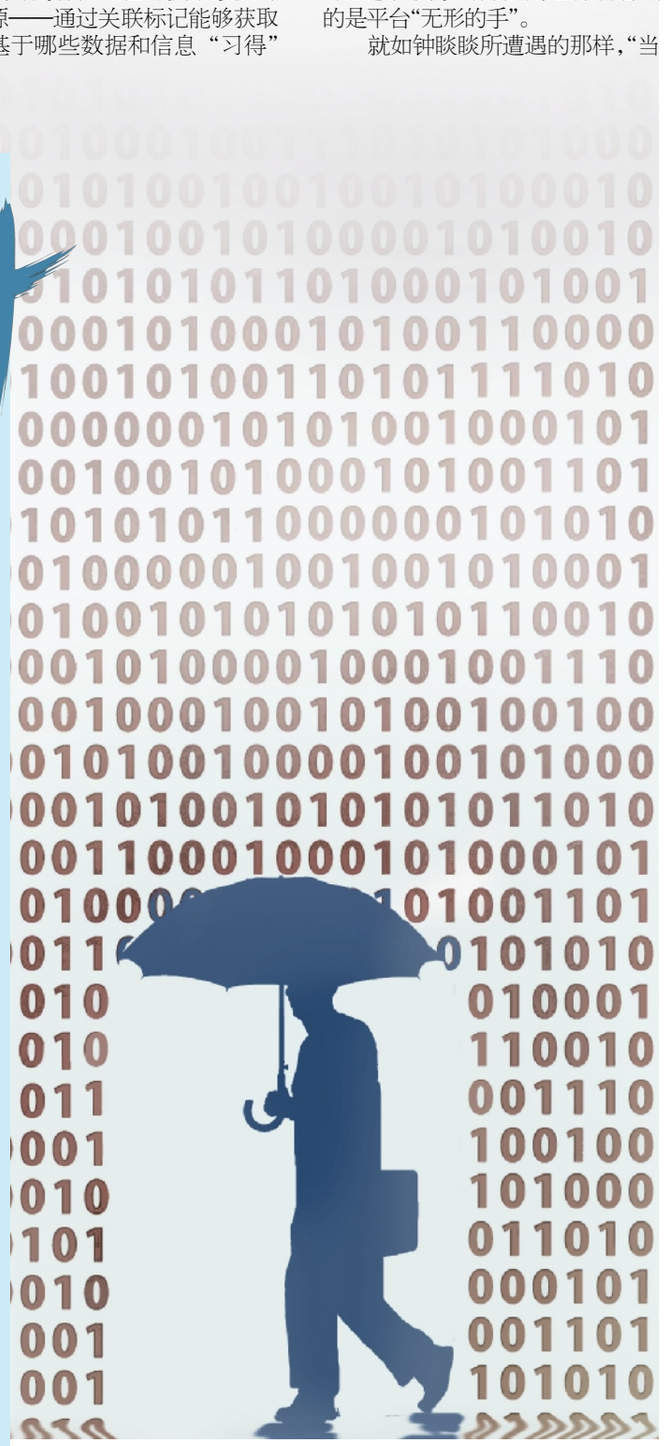
早在2018年,美国脸书首席执行官马克·扎克伯格在美国国会上就数据隐私、虚假信息、监管等数个议题接受讯问。当时人们已经意识到,当用户获得免费或者极低费用的服务时,消费者将被要求提供更多的个人数据,而这些数据被滥用的可能性会显著增加。

我国也在2021年就出台了《关于加强互联网信息服务算法综合治理的指导意见》《互联网信息服务算法推荐管理规定》等规定,明确算法治理的必要性和具体要求。而此次“清朗·网络平台算法典型问题治理”专项行动,力度更大、问题更加聚焦。

曾经,互联网努力为不同人群、不同议题提供平等的交流平台,打造自由对话的多元空间。但随着“流量至上”成了各大平台目标,它们开始借由算法之手不择手段,用户隐私信息得不到保护的问题浮出水面。

近年来,随着大语言模型技术进步、生成式人工智能服务兴起,若人工智能(AI)技术不加规范,会带来许多新问题:AI换脸诈骗、AI造谣、AI偏见歧视、AI语言暴力等。尤其是当生成式人工智能服务的对象是未成年人和老年人时,将会面临更大的风险。

据外媒报道,创办于2021年的Character.AI平台,近期就因开发的“情感陪伴聊天机器人”而官司缠身。今年10月,Character.AI在美国佛罗里达州一名青少年自杀事件中“扮演了某种角色”;12月,美国得克萨斯州一对父母决定起诉它“教唆未成年人杀害家长”,他们称机器人聊天工具让未满18岁的青少年“过度接触了色情、血腥暴力等不良内容”。



图片来源:视觉中国

能否打开算法“黑箱”?

面对算法“作恶”,钟睒睒呼吁“算法应该明白无误地公之于众”。他认为,没有一种标准是不可以公布的,应该公布并让所有使用者评价其意义。

但公开算法,就能打开“黑箱”、制止乱象吗?

吴凌翔告诉《中国科学报》,算法并不像外界理解的那样是彻底不透明的,一般都会通过发表论文、学术会议分享、公开课等公开其原理。但是,即便是常见的推荐系统,背后的算法机制也非常复杂,常常“不是一两个部门的事”,即便是开发者也未必全能搞清楚。反倒是AI检索增强生成的内容,现在的技术手段可以溯源——通过关联标记能够获取它是基于哪些数据和信息“习得”的。

王静远也同意,算法机制问题并不像想象的那样简单。“比如深度学习本身就是一个‘黑箱’,即便开发者也不清楚其中原理。”

事实上,对于算法工程师而言,真正的“黑箱”不在算法原理之中,而在数据与平台机制的设置之中——当用户量增大、数据变多,平台机制逐渐向利益“妥协”,久而久之便产生了“算法乱象”。

“算法始终是算法设计者意志的反映,是平台意志的反映。”北京大学数字治理研究中心主任邱泽奇在接受《中国科学报》采访时说。言外之意,复杂的算法问题背后潜藏的是平台“无形的手”。

就如钟睒睒所遭遇的那样,“当你打开这些平台,看到的总是同样的内容”“很多恶是人为造成的”。

不得不提的是,许多平台型软件在诞生之初,都肩负着改造社会的使命。比如某音的初心是“记录美好生活”、某团致力于打造“美好生活小帮手”、某滴提出的愿景是“让出行更美好”、某程希望提供“放心的服务,放心的价格”……不可否认,这些软件已经成为人们数字生活中的基础设施,但在巨大的发展惯性下,平台自发性选择了阻力最小、收益最高的方向,轻视乃至忽略了社会价值。在这种嬗变之中,算法的用途逐渐跑偏。

“在算法训练中,目标导向是关键因素。”王静远告诉记者,人工智能算法在设计时,会要求开发者设置一个目标函数,训练算法时会尽最大可能优化这个目标函数。如果算法以提高调度效率为目标,在模型优化过程中就会牺牲其他因素来追求高效;如果以精准的个性化推荐为目标,就不可避免地出现过度收集和利用信息的问题。

信息大爆炸时代,算法的筛选和过滤无疑迎合了为大脑“降本增效”的刚需。然而,当精准“捕捉”用户已无法满足平台的胃口时,杀熟成了平台“向前一步”的试水。王静远提到,平台通过“精准营销”为不同消费水平的顾客推荐不同价位的产品尚情有可原,但通过分析用户经济能力进行“个性化定价”的歧视行为就令人难以接受了,这在技术上能够且应亟须加以规避。

走向共同治理

在访谈中,几位专家不约而同谈到,除了人为滥用算法制造矛盾和对立外,算法更多是在复刻社会的现实问题。

“坦率地讲,算法就是帮你算数。你写了一套程序,它帮你把一些说不清、道不明的规律从数据里‘扒’出来。我的观点是,算法不会作恶。”邱泽奇说,问题的关键是数据和算法的匹配以及算法的调试,“说到底,都是人在忙活”。

他提出,不同的人虽然在使用同一个软件平台,但每人对数据的贡献和得到的反馈,在量和质上都有差异;而当算法应用数据时,便会复刻现实社会的结构,甚至放大现实社会的问题。

基于此,他认为有两条路可以尝试解决算法问题:一是对真实数据进行权重配置,二是调试算法进行纠偏。

“算法是人写的,是可以调整的。在方法意义上,算法是工具。”邱泽奇认为,工具是否适用是可以做交叉检验的,在技术上并不难实现。

有研究指出,算法黑箱、算法权力、算法陷阱等乱象很可能会成为数智经济负外部性的深层来源。此时,“算法向善”就成了全社会的共同呼唤。

在邱泽奇看来,“算法向善”包括四个关键概念:首先是改进,这是平台承担社会责任和社会价值的必然要求;其次是普惠,利益相关者的收益不提高,平台经营就是竭泽而渔;再次是包容,关注数智弱势群体,不只是平台的社会责任,也是人类价值的体现;最后是诚信,这是数智社会的底线规则,没有人类之间的诚信,算法只会成为人类自我欺诈的武器。

他坦承,通往“算法向善”的道路曲折而遥远,需要多方共同努力。

“首先需要解决平台和算法设计者的认知问题。”邱泽奇提出,前提是要着眼于保护各方的利益:在平台内部,建立平台业务的社会后果评估机制,不限于经济产出评估;在平台与社会之间,建立与利益相关者的协商沟通机制;在平台外部,建立平台社会评价机制,等等。

吴凌翔提出了类似建议,她认为算法治理需要搭建一个用户、平台、专家共同参与、共商机制的平台,促进通过对话达成共识。此外,她认为用户反馈机制和参与机制非常重要,这是社会治理的一种体现。

技术层面也有施展空间,以推荐算法为例,吴凌翔说,不仅要提升数据的多样性和丰富度,还可以对算法进行公平性约束、增加敏感性分析,并通过评估监测推荐系统内的不同环节,增加敏感性分析等方式,从技术角度对算法纠偏。

应对生成式内容带来的合规需求,王静远提到,现阶段重要的议题之一是要发展负责任的AI相关研究,其中既包括AI可解释性、公平性、泛化性的研究,也涉及安全可控方面的议题。但目前该领域面临着社会关注度不高、投入较少的尴尬局面。

“只有把蛋糕做大,才有蛋糕可分。”邱泽奇认为,治理与发展本就是一场拉锯赛,当前应在促进创新的前提下,通过“问责制”调整利益分配的逻辑和份额,考虑分配的公平性问题,在鼓励创新与促进平等之间寻求平衡。

“对于新生事物,制度建设不宜超前。”邱泽奇强调,新发展也会带来新问题,算法治理无法一蹴而就。“一个简单的警示和预防策略是对伤害的问责。”他强调,随着AI深入发展,算法自身的逻辑网络会越来越复杂,试图就具体问题进行预防是没有止境的。

记者手记

算法的一些“偏见”可能是固有的

赵广立

算法有偏见或歧视吗?

不同的人给出的答案可能完全相反。认为算法有偏见者,可能会以大数据杀熟、保险单歧视等来举证;认为算法无偏见者,会指出算法仅仅是如菜刀一般的工具而已,工具怎么会有偏见或歧视?

但是,如果我们换一种问法:人类社会产生的数据有偏见或歧视吗?如果答案是肯定的,那么算法“吃进”这些有偏见或歧视性的数据,会怎样?

从技术上讲,算法本身没有像人类一样的情感、观念和偏见。它仅仅是一系列指令的集合。在理想状态下,它只是按照预定的规则和逻辑对输入的数据进行处理、输出,不存在偏向。

但是,算法是基于数据进行训练和学习的。如果数据本身存在偏差,那么算法就会产生偏见。

例如,在招聘算法中,如果用于训练的数据大部分源于男性求职者的成功事例,那么算法在评估求职者时,可能会对男性求职者产生偏向。同理,算法“学习”了其他具有性别、地域或文化倾向的数据模式,它在后续的应用中就会带有这种偏见。

美国一些学者曾于2018年启动一项名为“图网轮盘”的研究,专门就此问题做了探讨:“这些图片来自哪里?”“照片中的人为什么会被贴上这样那样的标签?”“当图片和标签对应时,什么样的因素在起作用?”“当它们被用来训练模型系统时,会产生什么样的影响?”

这一研究更像一次行为艺术,明白无误地反映出人工智能算法系统很容易复刻和强化来自现实社会的固有偏见。如果对此视而不见,这些偏见便会渗入各类数字系统,继而影响整个社会的发展。

除了反映社会偏见之外,算法还会造成数据屏蔽——算法对数据的提取、分析、处理等操作是基于概率,那么它优先抓取的、出现频次较高的数据,就会成为“强势数据”,一些“弱势数据”或“少数派数据”就容易被忽略、被屏蔽。而且,数据体量越大、越是高度自动化的算法,越容易造成数据屏蔽。

数据屏蔽的问题更为隐蔽,但它的影响不容小觑,显著问题之一就是对文化多元性的影响。美国计算机科学家乔恩·克莱因伯格曾这样诘问:“如果我们都使用同一种算法作决定,是否会导致作出的决定高度趋同,导致我们的文化也高度趋同?”

如果说数据偏差带来的算法偏见算是“无心之失”的话,那么人为因素导致的算法偏向就是别有用心了。

例如,在设计内容推荐系统时,人为将系统目标设计为“延长用户的停留时间”,这就会导致算法倾向于推送耸人听闻的新闻信息或低俗娱乐内容,进而对内容的多样性和用户体验产生影响。另外,被困在算法里的外卖骑手、遭遇大数据杀熟的网约车用户等,背后的算法多是受人为因素干扰的。

算法偏见并非“顽症”,只要肯下功夫,总有办法尽量去消除。比如,从数据端着手,倡导在算法设计阶段进行多样化数据的收集,确保用于训练算法的数据多样性。尤其是涉及就业、金融保险等民生议题,在构建算法数据集时,可以通过收集来自不同性别、种族、年龄、地域等各种背景的事例,避免数据过于集中。

同时,还应对数据进行严格的质量检查,剔除带有明显歧视、偏见的信息。在算法的设计过程中,必须考虑多元化的公平标准,并引入公平性指标作为约束条件。

在监管上,要求平台或算法开发者公开算法设计的决策依据并不过分。如此,监管机构和第三方才能对算法是否存在潜在的偏见进行审查。

此外,设立专门的渠道,让公众能够通过反馈、投诉等方式参与到算法改善中。

最后,就目前所涌现的算法乱象问题,笔者认为,平台有很大的作为空间。以“钟睒睒事件”和“假冒张文宏事件”为例,平台至少可以有效处理虚假信息。对于未经核实的信息和内容,平台负有提示的责任和义务。平台的工作量和成本投入或许会增加,但受益的是大多数人。

如果平台最终留存的都是更优质的内容,数字空间也会因此更加清朗,社会也将更为积极向上。

社址:北京市海淀区中关村南一条乙3号 邮政编码:100190 新闻热线:010-62580699 广告发行:010-62580707 传真:010-62580899 广告经营许可证:京海工商广登字20170236号 零售价:1.00元 年价:218元 工人日报社印刷厂印刷 印厂地址:北京市东城区安德路甲61号