



读+Neo-reading周刊



扫一扫发现更多

2024年12月31日 星期二 主编:王永芳 责编:李煦 美编:职文胜 版式:陈笑宇 责校:江洋

音乐节难得初心不改 专栏

群医闪耀:同济迁汉70年 读书

同济大学法学院副教授、上海市人工智能社会治理协同创新中心研究员陈吉栋:

算法向善,终归是人的问题

□ 长江日报记者马梦娅



陈吉栋。

算法变“坏”了? 监管治理不够完善

反向驯服大数据? 你很难骗得了计算机

“手机窃听我说话,我和同事聊到过年旅行的行程,于是机票就涨价了。”“平台用得越多越贵,点了半个月的外卖店铺,这次居然涨十块钱。”大数据杀熟一直是人们无法证实,也同样无法证伪的赛博传说。

怀揣着对杀熟行为的抵制心理,许多网友开始了对数据“反向驯服算法”的测试。

有网友复制N遍“机票太贵了,买不起了”发在评论区,在平台立“穷鬼人设”,从而期待影响算法的推送机制。很快,这股反向驯服的风就吹到了各个平台,网友开始线上“许愿池行为”。

网友们希望打下来的价格还有酒店、奶茶、火锅券等,他们通过故意让大数据“听见”一些混淆信息来反向戏弄算法。

网友们为什么这么做?陈吉栋表示,用户跟APP的互动,总是一个信息被记录处理的过程。至于APP是否搞算法推荐,可能在多大范围内搞推荐,取决于这些APP的信息处理规则。

最坏的结果是,在一个APP收集的数据,可以在某个生态的数个APP内部流通,甚至超越一个生态进行信息共享利用。不过,整体上看,目前各个APP之间存在割裂,特定APP可能会基于控制和利用数据的考量,控制数据流通利用。“虽然法律规定了个人信息诸种权益内容,但是我们在使用APP时仍然存在潦草且盲目授权的情况,鉴于信息流通的特征,我们目前也缺乏对信息利用过程的有效监管手段。我们的监管仍是事后的,你无法查知这些数据使用了多少次,是否已经提供给别人,或者是否已经被他们共同加工处理。”此外,“我们也没有大范围地使用分布式数字身份(认证)”。

技术发展 宗旨是“用到合适之处”

至于网友“欺骗计算机”以获取更优惠的价格,陈吉栋表示,如果特定用户表现出了频繁的价格敏感行为,大数据和算法确实可能根据用户的行为和偏好来调整推送内容(也就是“欺骗”成功)。不过,长远来看,这样的可能性不大——多数情况下,平台已经掌握了我们的“大”数据,描绘出了个人精准画像。例如,你在购物平台上一年的购买了30万元数额的商品,你也许被数据定位为优质客户,比如你在使用打车软件时,一直是专车用户,那么拼车单就不大可能会推送给你。

陈吉栋表示,退一步想,算法针对不同的人群推荐不同价格的商品,并不都是坏事,关键在于当事人是否知情且同意。

技术和法律一样,都是为了人民的福祉。技术发展宗旨是“用到合适之处”。例如你的出行很迫切,但购票平台的机票已经卖空,你可以选择去第三方平台加价购买,这难道不是一种进步吗?但是他特别强调,“算法推荐系统本质上是一个定价系统,根据用户的不同标签显示出的购买实力进行定价。但是,用户要有知晓并自由选择的权利”。

在采访的最后,陈吉栋举了一个例子,小区里有位老奶奶让他很敬佩。每次丢快递包装时,她都会拿一支马克笔仔细把名字和地址抹去,非常注重自己的隐私权。我们的权利要自己去捍卫和保护,在数字时代,这点对于增强信息素养和避免破解隐私理论尤为重要。我们不能一方面谴责个人信息滥用的行为,一方面又不假思索地盲目同意,不加保护地任由我们的信息泄露。”

读+:在您看来,“以人为本,智能向善”这一理念诞生的时代背景是什么?

陈吉栋:智能向善,起源于数据向善,与之类似的还有科技向善,“xx向善”这类的倡导,一开始仅是提倡者向人们昭示未来的“xx”的独特性质,也就是“xx”这一工具,有助于向善。至于善是什么,如何向善,这类的问题,最终需要一定的社会共识、资源的投入与场景化的落地执行等,是一个需要不停反思的漫长过程。

这个理念的提出在国内外都有章可循。早在2014年,彭博社就举办了关于数据向善的对话,2017年国际电信联盟(ITU)举办了首次人工智能向善的全球峰会。2018年谷歌启动了人工智能向善的项目。后来这一口号就成为各国政府、国际组织广泛接受的流行语。

2023年,我国提出《全球人工智能治理倡议》,围绕人工智能发展、安全、治理三方面系统阐述了人工智能治理中国方案,倡议的核心理念就是“发展人工智能应坚持以人为本”。2024世界人工智能大会暨人工智能全球治理高级别会议在上海举行,智能向善、安全发展、公平普惠、协同共治,这些理念是讨论的热点问题。

与以往的技术不同,人工智能技术具有更突出的技术属性与社会属性,一方面,人工智能研发与应用,极大地改变了人类的生产、生活,人们可以享受前所未有的便利化服务。另一方面,从法律上观察,人工智能也改变了物理世界中主体对其权益的支配格局。人成

读+:“以人为本,智能向善”的核心突破点在哪?在算法设计环节,如何精准掌握“技术效率”与“人性关怀”的平衡?

陈吉栋:现阶段的任务之急是尽早进行纵向的、真实世界的风险评估,特别是在招聘、医疗保健和司法等敏感场景应用中。探索将人类价值注入算法逻辑是未来的方向。但是在目前这个阶段,我们还没有很好的技术方案将价值观、红线、道德护栏和法规整合并嵌入AI产品的编程之中,基本的方案还是在模型训练、数据管理与使用中的调教价值对齐的问题。

平衡点,可能在于寻回“主体性”尤其是信息主体(用户)在设计环节中的地位,最基本的信息主体(用户)在生命周期管理。在数据收集阶段,平衡隐私与数据丰富性。在搜集阶段遵守知情同意规则,明确告知用户数据的用途、收集范围和保护措施,获得用户的同意。在数据的存储阶段,安全与访问性的权衡格外重要。及时清理数据,妥善评估并保护用户权益。

读+:作为普通民众,如何保护自身权益,享有有温度的科技服务?

陈吉栋:关于智能向善,现阶段应该注意对于老年人、未成年人、劳动者等群体,进行特别的保护。

随着我们数字化生存的进程加速,虚实交互更加频繁,我们的权益日益数字化,相关的侵权变得更加隐蔽。但客观来说,我们对于这些权益边界的认识还不太清晰,传统的侵权规则变化还在调整中,但尚未形成明确的规则。如果从更宏观的角度看,更为庞大的民事、行政、经济法制的调整则将更为持久。

此外,要抛弃传统的责任救济的被动观

为被计算的数据,人与人之间关系的建立在人工智能系统的使用之上,当然就直接或者间接受人工智能的影响,也能发生人身与财产权益的侵害。这就使得人类有史以来在物理世界建立的秩序规则被挑战,基于此建立的信任也发生了变化,这个就是基本的背景。

人工智能研发决定了它的“性”是否是“善”的,而应用则决定了它是否可以“扬善除恶”。随着人工智能的不断增强,尤其是大模型的发展,其导致的损害,可能是灾难性的、存在性的。“以人为本,智能向善”的理念提倡,具有重要的意义。

读+:算法了解我们,给我们推送个性化的信息,展示有性价比的商品。然而,我们使用网约车、网上订酒店、订外卖……有时反而会被“杀熟”,感觉被“算计”了。这背后的原因是什么?

陈吉栋:在您说的这个语境里,“杀熟”其实是“因人定价”或者差异化定价。差异化定价可以构成价格歧视,侵害消费者的利益,也可能构成不同竞争主体之间的不正当竞争行为。现在法律上已经对这些行为设有明文规则。

需要讨论的是,就技术面而言,经营者基于识别消费者支付意愿,实现“大数据杀熟”,本质上还是一个数据利用的问题。根据已有的研究,大数据杀熟需要具备三个条件,即经营者有差异化定价的能力、商家能找出“杀熟”的目标用户以及用户能接受较高的价格。因

此外,应开发并赋予用户数字身份,供给更多的数据匿名化和加密技术工具或服务。

读+:您能列举一些典型的科技向善的案例吗?

陈吉栋:目前,这样的例子报道的还是挺多的。

例如阿里巴巴和公安部合作的“团圆”系统(找回失踪儿童),主要应用人脸识别、物联网及云计算、大数据等技术。利用人脸识别技术来识别和比对失踪儿童的照片与现有的图像数据,同时基于大数据分析和云计算技术快速收集、分析和处理各种与失踪儿童相关的个人信息、照片、案件进展等,及时提取出有用信息。与物联网技术相结合,扩展“团圆”系统的覆盖范围,将人脸识别技术应用于公共交通、商场等公共场所的监控系统,实时捕捉失踪儿童的信息,有效提高了找回失踪儿童的成功率。

在医疗资源紧张、早期筛查和疾病管理水平有待提高的情况下,腾讯将AI技术运用

念。传统的法律注重事后救济,当我们自身权益受到侵害时再寻求救济,但在智能服务提供商面前,我们无力与其背后的技术体系和能力对抗,无法证明所遭受的损害,维权的成本增高。我相信,未来,我们可能会有更多的救助、保险等制度,分担这些损失。另外,要注意利用技术抵抗技术,善用加密、数字签名等手段保护自己。

在监管上,在民众接受智能服务的过程中,我们不能期待一般民众熟悉法律规则,但最起码应当了解相关的参与者,妥善管控自己的身份及数字财产,这要求法律给这些主体设置相关的义务。

只有民众相信智能向善,我们才可能有

为价格歧视的危害更大,所以应该考虑遏制价格歧视的规则设计。

由于算法价格歧视是由大数据推动的,对于这种行为的监管,政策制定者可以瞄准算法价格歧视的基础——数据处理。比如,有效审查限制商家的信息处理规则、支持消费者个人信息可以携权、落实信息主体(即消费者)的算法解释和来遏制这种极端形式的价格歧视。

读+:“算法黑箱”有时让大众心存疑虑,引发信息茧房、隐私泄露担忧。我们应该如何精准约束算法,既不阻碍技术创新,又能让算法“更懂人心”、契合大众需求?

陈吉栋:“黑箱”无处不在。实际上,你我的大脑也是黑箱。清代王永彬的《围炉夜话》里提到:“百善孝为先,论心不论迹,论迹寒门无孝子。万恶淫为首,论迹不论心,论心世上无完人。”人心惟危,不过人的思想或内心想法,不一定转化为外在行为,这是与人工智能系统的不同之处。

现有研究已经证明,固有偏见的模型可能会延续甚至加剧其训练数据中存在的现有社会偏见,如性别偏见、种族偏见和上下文偏见,当模型在现实场景中部署时,可能导致不公平或歧视性的结果。模型中持续存在的偏见还可能引发道德和操作问题。例如,这些模型的偏差输出可能会对边缘化群体产生负面影响,导致错误信息,并破坏用户信任。

约束算法(算法治理)是一个系统工程,

于医疗领域,特别是癌症的早期筛查,以帮助基层医院提升诊疗水平。腾讯与全国200多家医院合作,将AI影像识别技术应用于胃癌、食道癌等多种癌症的早期筛查。通过AI对医疗影像的快速分析,医院可以更早期地发现潜在的病变,提高诊断的准确性和效率,助力基层医疗的智能化转型。

北京高德软件有限公司基于基础的全球卫星导航系统、地理信息系统、智能算法、大数据分析、云计算和人工智能等多种科技的应用开发了救护车“一路护航”功能。它攻克了定位模块存在的偏差和漂移这一难题,实现“被保障车辆”自动精准匹配到路线,实时自动校准定位,并通过定位模块直接向高德地图实时回传“被保障车辆”位置信息,确保救护车获取精准的位置信息,实现“生命通道”一路畅行。

读+:对于平台或企业而言,践行“科技向善”理念,在内部数据管理、算法研发流程上,要遵循哪些原则?

智能社会建设的根本基础,而实现这一数字信任,关键是激励民众获得保护,且积极参与沟通,最好的沟通是让民众参与到人工智能发展与治理中。

读+:您是如何理解“技术是否中立”的?

陈吉栋:技术的中立性是个很经典但有争议的话题,在西方有特定的文化背景和内涵。在我国,我认为技术中立仍是值得我们追求的目标。不过,技术中立从来不是纯粹的技术问题。从网络中立到人工智能的中立发生了一系列变化,都要求我们穿透技术,看见背后隐藏的经济、政治和文化的博弈,认识其问题的复杂性,规制、治理、权益

护栏,或者站在危险处发呆,等等,以此来训练AI模型。只要看到上述行为,系统就会自动向当地水务部门发出警报,马上就有人介入。

去年6月,这个防溺水AI识别出一个想跳河轻生的小伙子。从发出警报到把人救下来,前后不到15分钟。8月,又有一个女孩想跳河,从发现开始不到10分钟,人就被救下来了。

你看,科技向善一点都不抽象,它甚至也不是一句道德主张。技术的本质之一,是把人的愿望用工程化的方式实现。而科技向善的本质,是人用新技术应对现实中一个个真实的挑战,落到实处才行。

王永芳

随着算法融入生活,这个系统可能会越发的复杂。探索对算法进行约束——主要还是约束人的开发、生产与应用行为。我们在这里仅可以抽象地讨论几个要点。

第一,最难的还是理念。我们的理念是“既要、又要、还要”的理念。要平衡发展与安全,要兼顾公平与效率,要共治共享。发展、公平与共享等是目标,目标伟大崇高而可贵,但难的是,安全底线、权益保护与协同治理。

第二,治理应该是精准的。精准的意思是“采取什么样标准”,精准抓到需要治理的,不需要治理的则支持其发展。一个趋势是,算法治理要考虑到短期的经济利益和长期的社会发展,避免因追求短期利益而牺牲长期可持续发展。为此,我建议还是要从最核心、最重要的场景先开展治理实验或者风险分级,避免盲目地追求大而全的人工智能立法。

第三,无论哪种模式,都要求有公益的低廉的基础设施,尤其是公益性的风险评估数据库、评估工具、算法透明和可解释的工具,包括用户、监管机构和研究人员。这些基础设施是监管与规则实施的基础支撑,也可以有效降低企业的合规成本。

最后可能也是最重要的,在数字社会,一个科学的监管系统需要多元主体的参与。我称之为“合作监管”,或者叫监管3.0。合作的起点是赋予信息主体监管的能力,比如算法解释或者参与反馈等;在政府之外,监管的主体还包括服务提供商及其竞争者,这些主体具有规制的资源,构成了监管的可力量。

陈吉栋:企业内部怎么管理数据、怎么研发算法,本身是企业的私事,当然前提是没有相关的法律规定。因此,一般来说,规则的红线或者说底线取决于企业自身的判断。实践中,企业往往会基于成本收益法则,决定其底线到底有多低,当然前提是不违法。整体来看,现阶段大企业的合规能力较强,其红线和底线可能也严格一些,但合规的基本原则是自适应的,满足需求是第一位。因此,对于企业而言,合规没有最好,只有更好。

说几个要点。首先,企业得有数字化转型的基本需求和能力,没有这样的需求,数据不能解决其问题,这是注定不能持久的工作;第二,要想利用数据,要有一个数据的治理体系,这包括相关主体和资源的分配;第三,要有数据需求交互的程序与使用的监管激励规则;第四,有专门的人工智能的风险识别与实验小组,做一些内部的对抗实验;第五,可以考虑在公司内部设有相关的伦理委员会。

最后说一句,我们要警惕合规的形式化及其背后的风险。

分配等问题。

目前,技术中立的观点已经进入法律规则,但也遇到越来越多的挑战。科学与技术归根到底是人的活动,技术的功能和性能由人确定,技术的潜在用途和使用方式由人规划,因此技术可能包含着一种预设的偏见。

人类创造技术是为了改善生活、促进社会进步,但如果技术的应用偏离了这个初衷,就需要进行监督和纠正。

随着大数据、人工智能的发展,数据生产、处理及其后续的价值利用越发在一个密闭的技术体系中进行。从这个意义上说,我们不能被技术数量上新进式增减所蒙蔽,而应该时刻警惕整体的生态变革。