

大数据预测警务的运作机理、风险与法律规制

陈永生^{*}

内容提要 大数据预测警务的出现使警方侦查破案、预防和打击犯罪的能力获得突破性提升,但同时也会产生一些风险,须对其予以规制。域外预测警务已经过 1.0、2.0、3.0 三个阶段,预测能力不断提升,运作机理逐步优化。大数据预测警务的发展面临双重风险:一是数据的准确性、完整性和新鲜性难以保证;二是算法的错误、歧视难以避免和纠正。应当从三个方面对大数据预测警务进行规范:一是规范数据采集和处理的程序,确保数据的质量;二是建立算法审核机制,对算法的准确性和风险进行监督和评估;三是规制预测警务系统的设置与使用,确保对公民权利的保障。

关键词 大数据 预测警务 算法 法律规制

目 次

- 一、问题的提出
- 二、大数据预测警务的运作机理
- 三、大数据预测警务存在的风险
- 四、大数据预测警务的法律规制

^{*} 北京大学法学院教授。本文系 2022 年度司法部法治建设与法学理论研究部级科研项目“法律援助法实施问题与对策研究”(项目批准号:22SFB3019)的阶段性成果。

一、问题的提出

随着信息、网络技术的迅猛发展,大数据、人工智能越来越广泛地应用于各个行业。刑事诉讼处于国家与犯罪作斗争的最前沿,刑事侦查更应该充分利用科技发展带来的技术红利,提高侦查人员侦查破案、调查收集证据的能力。以美国为代表的域外国家大都重视利用大数据、人工智能的最新成果提高警方侦查和控制犯罪的能力。美国自20世纪末就开始使用犯罪绘图(Crime Mapping)和地理信息系统(Geographic Information System)来测绘和分析犯罪信息的空间分布。执法机构可以使用犯罪地图通知犯罪信息,作出资源分配决定,进行犯罪评估分析,甚至进行犯罪预测,以及对正在进行的系列犯罪进行地理画像(geographic profiling)。^①2019年,荷兰在全国范围内推广一项预测警务系统——“犯罪预测系统”(Crime Anticipation System),该系统可以根据一张犯罪“热度地图”(heat map)预测出不同区域(每个区域的面积为125米×125米)未来两周内的犯罪风险值。荷兰也因此成为世界上第一个在全国范围内部署预测警务系统的国家。^②

近年来,我国也开始注重利用大数据、人工智能提高预防和打击犯罪的能力。2019年10月31日,党的十九届四中全会通过《中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定》,明确要求“提高预测预警预防各类风险能力,增强社会治安防控的整体性、协同性、精准性”。同年,公安部成立了情报指挥中心,承担情报分析、研判预警、指挥调度等职责,并下发了《情指勤舆一体化实战平台建设任务书》,这标志着预测警务正式成为我国公安机关警务建设的重要内容。^③

然而,从总体上看,我国公安机关对大数据的运用主要侧重于已然之罪,即利用大数据查找已经发生的犯罪的实施者,如利用大数据视频技术侦查犯罪、利用人脸图像数据库抓捕犯罪嫌疑人等。而对于未然之罪的预测和防范,也即预测警务的关注严重不足,对作为预测警务制度基石的数据收集与管理、算法设计与审核等问题都缺乏必要的顶层设计。学界也一直没有将预测警务作为重要的研究课题。笔者于2024年6月19日在中国知网上对标题中包含“预测警务”的论文进行检索,一共只能搜索到19篇论文。这些论文基本上都是公安院校或公安系统的学者、专业人士写作的,发表在公安类杂志上,主要是介绍和阐述预测警务的概念、在域外主要国家的应用情况、预测警务的主要模

① See Derek J. Paulsen, *The to Map or Not to Map: Assessing the Impact of Crime Maps on Police Officer Perceptions of Crime*, International Journal of Police Science & Management, Vol.6:234, p.235 (2004).

② See Litska Strikwerda, *Predictive Policing: The Risks Associated with Risk Assessment*, The Police Journal: Theory, Practice and Principles, Vol.94:422, p.424 (2021).

③ 参见张金鹏、王菁怡、孙慧超:《情指行一体化体系中大数据技术的应用》,载《数字通信世界》2024年第3期,第87页。

型、功能、我国对预测警务的探索、存在的不足以及未来完善等问题,^④对预测警务的基本原理、可能面临的风险,尤其是对公民权利可能造成的损害,以及防范风险的举措等深层次问题则关注不足。

为推动预测警务在我国的发展,本文拟对预测警务在域外国家的产生与沿革进行深入考察,剖析其制度机理,研究其可能存在的风险,尤其是对公民权利可能造成的损害,在此基础上,对预测警务在我国的推行与规范提出建言与设想。

二、大数据预测警务的运作机理

警方控制犯罪能力的高低不仅取决于其在犯罪发生后准确查明犯罪事实的能力,在很大程度上也取决于其在案发前准确预测犯罪的能力。然而,在大数据时代到来之前,警方预测犯罪的能力非常有限。有学者将大数据时代之前的犯罪预测方法称为“临床法”(clinical method),这种预测方法建立在个体的直觉、经验等主观判断的基础之上,不受预先设定的已识别变量的控制。^⑤“临床法”预测的可靠性一直饱受质疑,早期的研究认为这种预测方法完全不值一提,几乎跟碰运气差不多。^⑥

大数据时代的到来使这一切发生了改观。由于收集和处理数据的能力大幅提升,大数据时代的预测更多地建立在客观数据和科学方法的基础之上,这种预测又被称为“精准预测”(actuarial prediction),其机理是“考察客观的、机制性的、可重复的预测因素,通过实证研究方法对这些预测因素进行选择 and 验证,并将其运用于被量化的已知结果”^⑦。对犯罪的精准预测也遵循基本相同的逻辑:通过理论和实证研究选择和验证客观的、机制性的、可重复的预测因素,将这些预测因素运用于通过大数据技术收集和处理的的大量数据,在分析技术(算法)的支持下获得犯罪的概率或者预期值,这一概率或者预期值将成为指导警察行动的重要依据。由于大数据预测警务技术能够较为精准地预测未来可能发生的犯罪的时间、地点甚至人员,因而警方可以据此合理地配置警力资源,将主要警力配置于最可能发生犯罪的地点或者目标人群,这能够极大地提高警方控制犯罪的能力和效率。在司法资源捉襟见肘的今天,大数据预测警务的出现给控制犯罪带来了新的希望和曙光,因而各国警察机构都对其青睐有加。

作为一种精准预测的大数据预测警务大约发端于20世纪90年代。虽然起源时间相对较晚,但迄今为止已经历几次较大的代际更迭。美国学者安德鲁·格思里·弗格森

④ 参见严佳灵:《预测警务技术嵌入社会风险治理:综述与展望》,载《公安研究》2024年第3期;唐鸿宇:《预测警务模式下合法控制潜在犯罪人之分析》,载《河北公安警察职业学院学报》2023年第1期;李国军:《论大数据驱动下的预测警务创新》,载《中国人民公安大学学报(社会科学版)》2015年第6期。

⑤ See Andrew Guthrie Ferguson, *Policing Predictive Policing*, Washington University Law Review, Vol.94:1109, p.1119 (2017).

⑥ See Alexander Scherr, *Daubert & Danger: The "Fit" of Expert Predictions in Civil Commitments*, Hastings Law Journal, Vol.55:1, p.17 (2003).

⑦ Jordan M. Hyatt et al., *Reform in Motion: The Promise and Perils of Incorporating Risk Assessments and Cost-Benefit Analysis into Pennsylvania Sentencing*, Duquesne Law Review, Vol.49:707, p.726 (2011).

将大数据预测警务的发展概括为三个阶段：预测警务 1.0 阶段（predictive policing 1.0）、预测警务 2.0 阶段（predictive policing 2.0）和预测警务 3.0 阶段（predictive policing 3.0）。^⑧支持这三个阶段的理论基础有所不同，这三个阶段的预测能力存在明显区别，与此同时，这三个阶段所引发的问题也不尽相同。

（一）大数据预测警务 1.0 阶段

有学者认为，最早的大数据预测警务实验开始于 20 世纪 90 年代纽约市警察局使用的统计数据比较系统（Compare Statistics System）。^⑨但大多数学者认为，真正开启大数据预测警务 1.0 时代的是美国洛杉矶警方。洛杉矶警方与几所主要大学的研究机构进行了一项实验：使用一种预测算法对三种财产犯罪进行预测：入室盗窃、汽车盗窃和车内财物盗窃。实验取得了成功：在实验开展的最初 6 个月内，入室盗窃罪下降了 25%。^⑩

大数据预测警务 1.0 与后文将要考察的大数据预测警务 2.0 之所以能够取得较为普遍的成功，关键在于这两个阶段的预测警务都有较为深厚的犯罪学理论作为支撑。“服务于任何目的的预测分析和犯罪绘图都应当有建立在犯罪学、社会学或者其他学科理论之上的强有力的理论基础。如果与理论基础的联系贯穿于预测分析使用的全过程，那么这种分析将会更为可靠。”^⑪犯罪学早期一个较为成熟和著名的理论是社会解组理论（social disorganization theory）。该理论认为，“集中于特定地点的经济劣势与其他社区层面的分裂变量交织在一起，减损了诸如家庭、教堂和学校等社会组织对社区成员，特别是年轻人施加影响的能力”。基于这一认知，社会解组理论认为：“‘地点’和‘方位’，也就是空间位置比其他任何个体特征与犯罪的相关性都要强”。^⑫此后，学者们在社会解组理论的基础上发展出生态学理论（ecological theory）。该理论被认为是与犯罪绘图和预测警务关系最为密切的理论，该理论重点关注防御空间，并主张通过改变环境设计来预防犯罪。^⑬

上述犯罪学理论揭示出一个重要事实：犯罪与环境因素有着非常紧密的联系，特定的环境漏洞是导致犯罪的重要原因，这成为大数据预测警务 1.0 和大数据预测警务 2.0 的重要理论基础。大数据预测警务 1.0 最典型的一个模型“近重复模型”（near repeat model）就建立在这一认知的基础上。“近重复模型”认为，一旦某一地点受到某一犯罪的侵犯，那么从统计学角度来说，在初次犯罪后的很短一段时间内，该地点很有可能再次受到同种犯罪的侵犯。^⑭学者们认为，之所以出现“近重复”现象，部分原因在于相同的

^⑧ See Andrew Guthrie Ferguson, *supra* note 5, 1115-1116.

^⑨ See Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, Washington Law Review, Vol.89:35, p.43 (2014).

^⑩ See Andrew Guthrie Ferguson, *supra* note 5, 1126-1130.

^⑪ Laura Myers, Allen Parrish & Alexis Williams, *Big Data and the Fourth Amendment: Reducing Overreliance on the Objectivity of Predictive Policing*, Federal Courts Law Review, Vol.8:231, p.240 (2015).

^⑫ *Id.*, 240.

^⑬ *Id.*, 240-241.

^⑭ See Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, Emory Law Journal, Vol.62:259, p.277 (2012).

犯罪分子返回来再次实施了犯罪,另一重要原因在于特定的环境漏洞并没有消弭,例如,可能是因为该地点的房屋建筑结构存在缺陷,也可能是因为该地缺乏警力。总之,导致“近重复”现象的部分原因在于特定的环境漏洞诱发了犯罪。除“近重复”理论之外,其他相关理论,如“日常活动”(routine activity)理论、“理性选择”(rational choice)理论、“犯罪模式”(crime pattern)理论等也都揭示了环境因素与犯罪的紧密联系。^⑮

大数据预测警务 1.0 的实验将犯罪学的研究成果简化为数据点(data points),并大量收集与之相关的历史数据,然后运用实验性算法预测出某一特定种类的犯罪可能发生的地点。这一地点是非常精确的,通常在 500 平方英尺以内。巡警收到显示预测结果的地图后,在巡逻的过程中会尽可能去造访那些地点,从而防范犯罪的发生。实验的结果证实了研究人员的预测。继洛杉矶警察局的实验取得成功之后,加利福尼亚州的其他一些地方也进行了类似的实验,结果也都非常令人振奋。例如,在圣克鲁兹的实验中,财产犯罪率下降了 4% 至 11%;在莫德斯托的实验中,犯罪率呈两位数下降;阿罕布拉警局报告称,使用预测算法一年之后,车内财物盗窃罪的犯罪率下降了 21%,汽车盗窃罪的犯罪率下降了 8%。^⑯

(二) 大数据预测警务 2.0 阶段

如果说大数据预测警务 1.0 还停留在对财产犯罪的预测上,那么大数据预测警务 2.0 将对犯罪的预测扩展到暴力犯罪。一般认为,与财产犯罪不同,暴力犯罪的发生在地理学上更为随机,^⑰因而对暴力犯罪的预测会更为困难。然而,犯罪学的研究表明,暴力犯罪同样会受到地理环境因素的影响,因而对暴力犯罪的精准预测也是完全有可能的。例如,犯罪学理论认为,特定的环境漏洞可能是暴力犯罪的重要诱因:有些小巷可能因为灯光昏暗、有便捷的逃跑路线以及更为接近受害群体等原因更加吸引抢劫罪犯;有些酒吧可能因为酒精混合物、毒品以及夜间判断错误等原因更容易诱发暴力斗殴。^⑱

预测警务 2.0 阶段最具代表性的预测模型是“风险地形模型”(Risk Terrain Modeling)。“风险地形模型”更多地将犯罪视为某一地区的社交、物质以及行为因素之间动态交互的结果,而不是将犯罪视为先前事件决定的结果。因而,“风险地形模型”并没有将关注的重点放在过去的犯罪之上,而是将重点放在可能增加犯罪风险的当前环境风险因素之上。^⑲基于这一理念,“风险地形模型”的运作机制在于寻找导致不同犯罪的风险因素,并运用多层计算机绘图系统进行绘图。

寻找并确定导致不同犯罪的风险因素对于“风险地形模型”来说至关重要,大量的理论和实证研究已经探寻出诸如杀人、伤害、强奸、抢劫、枪击等诸多暴力犯罪的风险因素。例如,有研究认为,导致谋杀和非过失杀人的风险因素包括帮派活动、枪支的可获得

^⑮ See Andrew Guthrie Ferguson, *supra* note 5, 1129.

^⑯ See Andrew Guthrie Ferguson, *supra* note 5, 1130.

^⑰ See Andrew Guthrie Ferguson, *supra* note 14, 283.

^⑱ See Andrew Guthrie Ferguson, *supra* note 5, 1132-1133.

^⑲ See Andrew Guthrie Ferguson, *supra* note 14, 281.

性、毒品交易、毒品和药品的使用、失业;^{②①}导致枪击犯罪的风险因素包括毒品交易、有风险的住宅开发项目、^{②②}帮派活动、枪支的可获得性、犯罪历史^{②③}。

大数据预测警务 2.0 阻遏犯罪的逻辑与 1.0 较为类似:通过向高风险地区增派警力等方式降低该地区的风险值,进而最终实现犯罪率的下降。事实证明,这一阻遏犯罪的努力是非常成功的。一项在波士顿开展的研究表明,根据预测警务 2.0 的预测结果进行警力介入之后,暴力犯罪下降了 17.3%,其中,抢劫犯罪下降了 19.2%,重伤害犯罪下降了 15.4%。^{②④}

比较分析大数据预测警务 1.0 和 2.0 不难发现,两者除了适用范围有所不同之外,其预测依据也存在差别。大数据预测警务 1.0 还无法摆脱对过往犯罪活动相关数据的依赖,甚至可以说主要是建立在过往犯罪活动的数据库上的。而大数据预测警务 2.0 则“使用一个地点的地理和环境属性来预测未来的犯罪,并不依赖于该地区过去犯罪活动的相关数据”,因而被认为是“真正的预测警务”。^{②⑤}但是,大数据预测警务 1.0 和 2.0 也存在一些本质上的相同之处:首先,二者遵循的基本逻辑是相同的,即某一地区存在的特定的环境漏洞可能诱发犯罪,进而导致该地区发生犯罪的风险升高;其次,无论是大数据预测警务 1.0 还是 2.0,都不会对某一具体犯罪作出预测,而只会对犯罪的风险值作出预测。

(三) 大数据预测警务 3.0 阶段

如果说大数据预测警务 1.0 和 2.0 只是实现了对犯罪地点的预测的话,那么预测警务 3.0 则实现了对可能实施犯罪的具体的人的预测。大数据预测警务 3.0 建立在如下认知上:负面的社交网络如同环境漏洞一样,有可能诱发犯罪行为。犯罪学理论发现,一小部分人群受其所处的社会关系网络的影响,成为犯罪的受害者或者实施者的风险比其他人高很多。基于这一认知,研究人员构建了人与人之间关系和连接的地图,并将地址、电话号码以及其他数据源连接起来,以此为正在进行的侦查活动提供见解和思路,同时识别新的犯罪类型。^{②⑥}

2013 年,研究人员在芝加哥进行了一项实证研究,研究的目的是对预测警务 3.0 进行实践和验证。研究人员通过分析住址、社交关系情况(过去与枪支暴力受害者和帮派发生关系的经历)以及年龄等指标来预测一个人成为枪支暴力受害者或者实施者的可能性。^{②⑦}在实验的过程中,每天都有数百名警官被派往芝加哥最危险的区域,这些警官收

^{②①} See Joel M. Caplan & Leslie W. Kennedy eds., *Risk Terrain Modeling Compendium*, Rutgers Center on Public Security, 2011, p.57.

^{②②} 研究表明,居住在公共住房设施(public housing facilities)的人比不居住在公共住房设施的人遭受枪击犯罪侵害的几率要高两倍。*Id.*, 69.

^{②③} *Id.*, 69.

^{②④} See Anthony A. Braga et al., *SMART Approaches to Reducing Gun Violence*, at <https://centerforimprovinginvestigations.org/wp-content/uploads/2018/08/SPI-Gun-Violence-Spotlight-FINAL-2014.pdf> (Last visited on June 25, 2024).

^{②⑤} See Kelly K. Koss, *Leveraging Predictive Policing Algorithms to Restore Fourth Amendment Protections in High-Crime Areas in a Post-Wardlaw World*, *Chicago-Kent Law Review*, Vol.90:301, p.310 (2015).

^{②⑥} See Andrew Guthrie Ferguson, *supra* note 5, 1138.

^{②⑦} See Megan Behrman, *When Gangs Go Viral: Using Social Media and Surveillance Cameras to Enhance Gang Databases*, *Harvard Journal of Law & Technology*, Vol.29:315, p.332-333 (2015).

集上述区域中上万名帮派成员、帮派成员的竞争对手以及帮派成员声称的领地等信息,然后运用大数据技术对上述信息进行综合分析。在此过程中,警方锁定了四百多人,这四百多人因为在社交关系网络中所处的位置,被认为最有可能成为谋杀犯罪的受害者或者实施者。^{②7}芝加哥警方将这个四百人名单称为“热点名单”(heat list)。一个人一旦被列入“热点名单”,就会成为警方实施干预行为的对象,更容易被警方采取拦截或逮捕等措施。有学者指出,虽然被列入“热点名单”这一事实并不足以构成“合理怀疑”或者“合理根据”,进而使警方对该人采取的拦截或者逮捕措施获得正当性,但是,在其他大数据监控措施的配合下,警方对该人采取措施的可能性会大幅增加。^{②8}

大数据预测警务 3.0 在美国的实验被证明是卓有成效的,一项在新奥尔良市进行的实验证明,通过界定那些可能卷入枪支暴力犯罪的高风险人群,并对这些人采取干预措施之后,新奥尔良市谋杀罪的犯罪率下降了 21.9%。^{②9}

三、大数据预测警务存在的风险

大数据预测警务给人们展示了一幅非常美好的图景:传统的建立在直觉、经验等个体判断基础上的主观预测被建立在海量数据和科学方法基础上的客观预测取代,预测的精准度大幅提高。根据这种预测来部署警力必然会极大地提升打击犯罪的效率,进而有效缓解刑事司法资源捉襟见肘的现状。不仅如此,大数据预测是计算机通过运行事先设计好的算法作出的,没有掺杂警员的个人喜好和私人恩怨,也不存在权钱交易和暗箱操作的空间,因而预测的结果更为客观公正,能够有效避免当事人在刑事诉讼活动中遭受偏见和歧视性待遇。这些优势很容易使人们迷信和盲目依赖大数据预测警务,对其隐含的问题视而不见。

然而,如果对大数据预测警务深入分析不难发现,预测结果的准确性和公正性不是必然的和无条件的;相反,其对数据和算法有着非常苛刻的要求,如果数据和算法不能满足这些严苛的要求,预测的结果不仅可能出现错误,甚至可能会对公民权利构成巨大威胁。在侦查实践中,存在大量影响数据质量的因素,也有很多算法难以克服的障碍,以致严重制约了预测警务的准确性和公正性。正因为如此,在域外国家,预测警务的应用一直存在激烈的争议,近年有些国家或地区甚至禁止警察在执法时使用预测警务。例如,美国加利福尼亚州圣克鲁斯市曾于 2011 年通过了一项“预测警务计划”,启动了预测警务的试点,从而成为美国最早试点预测警务的城市。但是,由于预测警务一直面临基础数据不准确、预测结果存在偏见等质疑,2020 年 6 月,圣克鲁斯市通过一项法

^{②7} See Monica Davey, *Chicago Tactics Put a Major Dent in Killing Trend*, N. Y. Times, June 11, 2013, at A1.

^{②8} See Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, University of Pennsylvania Law Review, Vol.163:327, p.386 (2015).

^{②9} See Andrew Guthrie Ferguson, *supra* note 5, 1142.

令,禁止本市警察机构继续使用预测警务。^{③0}

(一) 数据的风险: 准确性、完整性和新鲜性难以保证

从前文对大数据预测警务的介绍可以看出,大数据预测警务的运行逻辑是: 收集一种或者多种类型的数据; 运用算法对收集的数据进行分析,以预测出某一地域乃至某个人发生(实施)犯罪的可能性; 警方根据预测的结果实施警务行为。由此可见,如果将预测警务视作一幢大厦,数据就是这座大厦的基石,数据的质量直接决定着预测结果的准确性。可以想见,作为基石的数据哪怕出现细微的偏差,也会对预测结果造成严重影响。

1. 数据的准确性和完整性难以保证

司法实践中,保证数据的准确性和完整性绝非一件易事,在采集、录入、整合和共享数据的过程中,存在诸多可能导致数据错误或者不完整的因素。美国联邦司法统计局曾经就这一问题发出喟叹:“在绝大多数专家看来,犯罪记录准确性和完整性的不足是影响国家犯罪记录信息系统的最为严重的缺陷。”^{③1}

以收集数据为例,完整地收集各类犯罪的数据对于警方来说是一件非常困难的事情。虽然某些犯罪,如杀人罪或者涉及交通工具的犯罪,因为犯罪的性质、严重程度或者某种激励机制(如保险公司理赔程序的要求)的存在,被害人或者相关人员通常会向警方控告或者报案。但是,大量犯罪中的被害人或者相关人员可能不会向警方报案或者控告,这将严重影响警方掌握的数据的完整性。例如,性侵害类犯罪的被害人出于免遭二次伤害等原因,往往选择不向警方报案。又如,家庭暴力类犯罪的被害人出于保护隐私、维系家庭关系等考量,也经常选择隐忍。正因为如此,警方掌握的犯罪数据往往是残缺不全的。据美国联邦司法部测算,在2006年至2010年期间,只有不到一半的暴力犯罪被警方统计,被警方统计的家庭财产犯罪的比率甚至更低。^{③2}

再以数据的整合和共享为例,为强化数据资源的功能,执法机构往往会对不同数据库中的数据资源进行整合,在整合数据的过程中也容易出现各种错误。整合数据的方法各有不同,可能仅仅基于姓名和住址进行整合,也可能加上社会保险号、出生日期等附加信息,由于姓名和住址数据是以多种格式捕捉的,这就很可能导致出现错误。^{③3}

2. 数据的新鲜性难以保证

除准确性和完整性外,大数据预测警务的有效运作还要求数据具有新鲜性。社会生

^{③0} See Matthew Guariglia, *Technology Can't Predict Crime, It Can Only Weaponize Proximity to Policing*, Electronic Frontier Foundation (Sept.30, 2020), at <https://www.eff.org/zh-hans/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing> (Last visited on June 15, 2024).

^{③1} Alex R. Hess, *Herring v. United States: Are Errors in Government Databases Preventing Defendants from Receiving Fair Trials*, *Journal of High Technology Law*, Vol.11:129, p.147 (2010).

^{③2} See Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, *New York University Law Review Online*, Vol.94:15, p.24 (2019).

^{③3} See Anita Ramasastry, *Lost in Translation-Data Mining, National Security and the Adverse Inference Problem*, *Santa Clara Computer & High Technology Law Journal*, Vol.22:757, p.775 (2006).

活瞬息万变,警务预测系统必须及时跟进社会生活的急剧变化,否则将无法保证预测结果的准确性,而如何保障数据的新鲜性一直是困扰大数据预测警务的一个问题。

如前文所述,支持大数据预测警务的一项重要理论基础是犯罪学的研究成果:特定的环境漏洞是导致犯罪发生的重要原因。也就是说,某一地方之所以一而再、再而三地发生某种或某些犯罪,是因为存在某些环境漏洞,比如照明系统比较昏暗、缺乏必要的警力资源、有便捷的逃跑路线等。如果这些环境漏洞一直没有得到修复,预测警务系统会一直提示该地方为“犯罪高风险区”,警方基于该提示实施相应的警务行为并无疑问。但值得注意的是,上述环境漏洞是可能被修复的。比如,在某一地方被列为“犯罪高风险区”以后,警方采取了一系列补救措施,比如改善照明系统、增派警力、在可能的逃跑路线上增设卡点等,那么该地方的环境漏洞就可能因此得以修复,发生犯罪的风险也会随之降低。在此情况下,如果不及时更新相应的数据,那么预测警务系统还将维持该地方为“犯罪高风险区”的提示,此时警方如果盲目依赖预测系统的提示实施警务行为就会引发问题,轻则导致警力资源的浪费,重则可能对公民权利造成侵害。

然而,保持数据的新鲜性,如同保证数据的完整性和准确性一样绝非易事,需要投入大量的人财物力资源,需要建立相应的长效机制;如果投入的人财物力资源不足,或者缺乏相应的长效机制,数据库很容易出现老化问题,严重影响预测结果的准确性。在美国,虽然其预测警务系统比较发达,相关警察部门也非常重视预测警务数据的质量,但即便如此,美国刑事司法数据库的数据质量也一直广受诟病。如美国联邦司法统计局在2005年的一份报告中指出:“调查显示,刑事历史报告面临很多问题,包括大量积压、没有处置的旧记录以及缺乏保障记录准确性的必要审计。”^④金斯伯格大法官在“赫林诉美国案”(Herring v. United States)中发出这样的感叹:“这些数据库中的错误产生的风险并不小,赫林案给我们的警示是,执法部门的数据库没有受到充分的监督,并且经常是过时的。”^⑤

(二) 算法的风险: 错误、歧视难以避免和纠正

在充分收集与犯罪有关的数据之后,警务人员必须利用计算机系统,采用一定的算法,对所收集的数据进行分析,从而得出有关犯罪可能性的结论。由此可见,算法是连接犯罪数据与犯罪可能性的枢纽,是计算机对数据进行分析、整理、挖掘所依据的步骤和程式,算法设置是否合理直接决定着预测结果是否准确。域外预测警务的实践表明,预测算法作为评估未来发生犯罪可能性的模型,是极有可能出现错误的,甚至可能会对刑事诉讼的一些基本原则构成威胁。如果对这些问题视而不见,片面迷信预测算法,不仅可能无法实现预测警务提升诉讼效率的功能,反而会对司法公正的价值目标造成巨大冲击。

1. 算法可能出现错误

预测算法的工作机理是根据一系列与犯罪有关的因素判断某一地区或者某个人未来发生或实施犯罪的概率,而不是在有充分证据的基础上断定某个地区是否必然会发生

^④ Alex R. Hess, *supra* note 31, 147.

^⑤ Herring v. United States, 555 U.S. 135, 155 (2008).

犯罪或者某个人是否必然会实施犯罪。这一工作机理决定了预测的结果是有可能出现错误的。“预测技术是存在错误率的,错误涵盖假阴性(false negative)和假阳性(false positive)两种。”^{③⑥}“假阴性”是指算法预测某一地区不太可能发生犯罪或者某人不太可能实施犯罪,但是该地区事后发生了犯罪或者该人事后实施了犯罪。“假阳性”是指算法预测某一地区可能发生犯罪或者某人可能实施犯罪,但事后该地并未发生犯罪或者该人并未实施犯罪。无论出现了“假阴性”还是“假阳性”的预测结果,都会对刑事诉讼的价值目标造成冲击。“假阴性”的预测结果可能导致警方防范不足,被错误认定为“低犯罪风险”的地区或人员因为警力分配不足最终发生或实施了犯罪。“假阳性”的预测结果则会导致警方对被错误认定为“高犯罪风险”的地区或人员配置过多警力,甚至对无辜公民的权利进行不当干预和侵害。相比之下,“假阳性”对刑事诉讼价值目标的冲击更为严重。“假阴性”尽管可能导致一些犯罪未能得到有效控制,但警方通过刑事追诉程序通常可以使大多数犯罪造成的损害后果得到修复;而“假阳性”则可能导致警方对无辜公民的人身权进行限制甚至剥夺,而对公民人身权的限制或剥夺在发现错误时很难弥补。

不仅如此,相较于预测警务 1.0 和 2.0 而言,“假阳性”对预测警务 3.0 造成的冲击更大。原因在于,预测警务 1.0 和 2.0 只是预测某个地区是否可能发生犯罪,而预测警务 3.0 直接预测某个特定的人是否存在较高的犯罪风险,一旦出现“假阳性”的错误结果,将直接导致国家公权力对公民个人权利进行不当干预和侵犯。“就预测警务 3.0 而言,当警方使用大数据技术在大型数据库中开启匹配嫌疑模式时,错误导致的风险上升了。假阳性的预测结果会导致无辜的公民被单独挑出来,并就非犯罪行为接受调查。这些调查可能包括警察的肢体接触、威胁甚至暴力行为。”^{③⑦}

2. 算法可能存在歧视

由于预测算法建立在客观数据和科学分析的基础上,而非基于办案人员的主观直觉和过往经验,所以很容易使人们误以为大数据预测警务得出的结论是客观公正,不存在偏见的。然而,算法在消费、求职等其他社会领域的应用早已表明,算法绝非如其外观呈现出来的那样客观公正,而是存在严重的歧视问题。“算法在诸多领域的运用中产生了歧视性后果,常见的表现形式有价格歧视、就业歧视、信用歧视。”^{③⑧}之所以算法也会出现歧视,一个重要原因在于预测算法并非脱离社会生活而存在,它服务于社会生活,因而必须接受源自社会生活的大量数据的训练,如果社会上存在对某一群体的歧视和偏见,那么用来训练算法的数据也会存在歧视和偏见,最终导致算法被打上歧视和偏见的烙印。预测警务的算法也无法例外,“如果作为基础的数据是歧视性的,那么建立在该数据基础上的算法也会是歧视性的,那些所谓的无歧视算法(color-blind algorithms),所做的不

^{③⑥} Andrew Guthrie Ferguson, *supra* note 5, 1158.

^{③⑦} Andrew Guthrie Ferguson, *supra* note 5, 1158.

^{③⑧} 洪丹娜:《算法歧视的宪法价值调适:基于人的尊严》,载《政治与法律》2020年第8期,第28页。

过是强化刑事司法体系中业已存在的偏见而已”^{③⑨}。

以美国为例,美国很多司法辖区都以拦截、逮捕或者监禁等数据训练预测算法,表面看来,这些数据的确与一个地区的犯罪风险程度存在相关性,并且这些数据并非族裔、收入等敏感数据,貌似与歧视和偏见无关。然而,由于美国社会存在严重的种族歧视和贫富差距,对少数族裔、穷人适用拦截、逮捕和监禁的比率比对白人、富人高得多,因而其拦截、逮捕和监禁等数据本身就存在严重的歧视、偏见问题,用这些数据训练出来的算法也很难摆脱歧视和偏见的色彩。以拦截的适用为例,“尽管 14 至 24 岁之间的黑人男子和拉丁裔男子只占纽约市人口的 4.7%,但是对他们的拦截却占到了 2011 年纽约市拦截总量的 41.6%”,不仅如此,“对年轻黑人拦截的次数已经超过了整个城市黑人的数量”,与此形成鲜明对比的是,上述被拦截的人当中只有很少的人最终被逮捕。^{④⑩} 黑人和白人群体拦截数据的巨大差异可能不是因为黑人群体更可能实施犯罪,而是因为“刑事司法体系存在内在的种族偏见,例如警察更倾向于对少数族裔而非白人采取措施,又如少数族裔社区往往被布置了更多的警力”^{④⑪}。

很明显,如果执法机关用过往适用拦截、逮捕或者监禁等措施的数据训练算法,必然导致在其他变量相同的情况下,相对于白人社区,黑人或者拉丁裔等少数族裔聚居的社区更容易被算法界定为“高犯罪风险区”。一旦被界定为“高犯罪风险区”,这些地方就会受到过度监控,居住在这些地区的居民的权利,包括人身权就更容易受到公权力的侵害。

3. 算法出现问题很难被发现和纠正

无论是建立在直觉、经验基础上的“临床预测”,还是建立在大数据、算法基础上的“精准预测”,都可能出现错误和歧视等问题。然而,与警员个体的“临床预测”不同的是,预测警务这种“精准预测”由于有客观数据和科学方法的加持,因而更容易呈现出一种客观而准确的外观,让人们很容易忽视其本身可能存在的各种问题。预测警务出现问题很难被发现和纠正,还与算法不透明息息相关。许多算法开发公司都寻求知识产权保护,拒绝公开算法。以美国为例,刑事司法系统中运用最广泛的两大算法 (COMPAS 和 PredPol) 的开发商均主张其软件的内在运作机制属于商业秘密,还经常要求警方签署保密协议,反对公开记录。^{④⑫} 其实,即使开发算法的公司不寻求知识产权保护,预测算法存在的问题也很难被发现和纠正,这与算法本身难以理解不无关系。算法在社会生活各个领域的运用都存在“黑箱效应”。“大数据分析使用的人工智能在处理数据以及根据数据作出决定时使用复杂的数学算法。这些算法通常是不透明的,产生了所谓的‘黑箱’

^{③⑨} Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, Michigan State Law Review, Vol.2016:947, p.980 (2016).

^{④⑩} See Kelly K. Koss, *supra* note 24, 316.

^{④⑪} Ric Simmons, *supra* note 39, 980.

^{④⑫} See Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, U.C. Davis Law Review, Vol.52:1067, p.1087 (2018).

效应(‘black box’ effect)。”^{④③}对于属于计算机外行的人士来说,理解算法并发现算法的错误是一件非常困难的事情。

四、大数据预测警务的法律规制

近年来,我国公安机关越来越重视预测警务的探索和运用,并取得了一定的积极成效。例如,苏州市高新区公安分局运用“蜂巢计划”预测警务系统对每日警情、发案数据进行自动分析,预测重点发案点位,指导巡防工作,极大地节约了警力资源,“2018年1月至11月,辖区违法犯罪警情同比下降16.5%,其中刑事案件发案数同比下降31.3%,而巡防警力未增加一人”^{④④}。又如,2021年,合肥市公安局研发了“守望者”综合反诈平台,该平台“能够依据结构化数据的关联性,自动建立数据连接,以预警数据中的人员信息为依据,挖掘拓展关联信息,输出预警线索,提高预警率”^{④⑤}。

然而总体上看,我国预测警务的探索尚处于起步阶段,目前尚未出台全国层面统一的制度规范,既没有规范性文件对处理数据的各个环节进行规制,也没有相关文件对预测警务算法开发和应用的条件作出明确规定,更没有相关文件对预测警务运行过程中警察权的约束以及公民权利的保护作出规定。在这种背景下,各地对预测警务的探索往往各自为政,缺乏统一规制。这种状况一方面可能导致数据和算法的质量无法得到保障,影响预测的准确性;另一方面,由于警察权在运作过程中缺乏约束,可能对公民权利造成侵害。

前文的分析表明,预测警务的数据和算法都可能出现错误,并且这些错误往往具有隐蔽性,很难被觉察和纠正。因而,我国在未来建构大数据预测警务制度时应当建立严密的约束机制,有效确保数据和算法的准确性,并在数据和算法出现错误时及时纠正;同时有效规制警察权的行使,确保公民权利免受不当干预。

(一) 规范数据采集和处理程序,确保数据质量

近年来,随着大数据警务活动的开展,许多公安机关已经收集并储存了海量数据,这为预测警务活动的开展奠定了坚实的数据基础。以辽宁省为例,截至2021年,该省“已经汇聚、整合内外部数据资源一百余类、二百余亿条数据……省公安厅已建各类信息化系统、平台共二百余个”^{④⑥}。然而遗憾的是,针对警务数据质量和管理的立法尚处于空白状态。虽然一些非官方机构很早就开始探索警务数据的行业标准,如2011年中国计算机软件与技术服务股份有限公司、中国人民公安大学公安管理学院等单位联合起草了《公安信息化数据质量管理规范》,对数据的采集、审核、校验、标记、更正等提供了行

^{④③} Moira Paterson & Maeve McDonagh, *Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data*, Monash University Law Review, Vol.44:1, p.3 (2018).

^{④④} 李海明、李路:《智慧警务“苏州方案”示范全省领先全国》,载《江苏法制报》2018年12月17日,第8版。

^{④⑤} 海世梅、乔磊:《合肥:“守望者”预警+“机器人”联手防电诈》,载《人民公安报》2021年10月11日,第5版。

^{④⑥} 佟冶、颜坤星:《公安数据质量监测的实践研究》,载《辽宁警察学院学报》2021年第3期,第15页。

业指引,但迄今为止,公安部尚未出台全国统一适用的规制警务数据处理的规范性文件。实践中,各地公安机关采集和处理数据时缺乏有效的规范和指引,导致处理数据的很多环节都存在问题。以数据采集为例,数据“虚采、漏采和信息不完整、不规范”仍然比较严重。^{④7}再以数据录入为例,有些地方的公安机关“不执行统一的业务规范与数据标准,或是使用的标准版本滞后,导致数据格式不一致”^{④8},给数据的融合、使用造成困难。又以数据的储存和维护为例,数据库信息泄露风险不容忽视,甚至存在公安人员泄露公民个人信息的情况。在一些出售、非法提供公民个人信息犯罪的案件中,泄露公民个人信息的源头就是公安机关工作人员。^{④9}为确保服务于预测警务的数据的质量,建议采取以下措施:

第一,由公安部出台统一的警务数据操作管理规范,对采集和处理警务数据各个环节的要求和标准作出严格规定。具体而言,操作管理规范应当根据不同类型执法数据的特点,对数据的收集范围、收集程序和收集方法分别作出规定,并对容易被遗漏的数据类型作出特别提示,从而确保收集数据的准确性和完整性。应当对录入数据的格式、创建条目的程序和要求等作出统一规定,从而确保录入数据的规范性和统一性。应当根据不同类型数据的特点,对更新数据的期限作出规定,并设置相应的程序提醒和督促工作人员及时更新数据,从而确保数据的新鲜性。应当明确数据的复核和校准机制,在一名工作人员完成数据的录入、更新之后,由另一名工作人员进行复核和校准,从而将错误率降到最低。“制定强制性的统一标准,指导如何收集数据并将其输入预测警务项目……将显著降低与人类收集大数据的易错性相关的潜在风险。”^{⑤0}

第二,公安机关设置专职的数据操作和管理人员,专门负责数据的采集、录入、传输、存储等工作。数字时代的到来意味着警务活动的样态发生了巨大变化,与数据相关的警务活动占据了警务活动的相当部分,这在客观上要求将更多的警力资源分配到处理数据的工作中。以数据的收集为例,虽然目前已经出现了一些自动化的警务数据收集工具,但大量的警务数据还是依赖警务人员的手动输入,例如警情、案件、入监所人员、涉案人员等触发型公安核心业务数据主要依靠警务人员手动输入;又如常住人口、暂住人口、出租房屋等常态型公安警种业务数据,有相当一部分也依靠警务人员手动输入。^{⑤1}为保证这些数据能够被完整收集、准确录入,以及在发生变化后及时更新,必须投入大量的专职警力资源。

(二) 建立算法审核机制,对算法的准确性和风险进行审查评估

目前,警务预测算法的设计和运用基本上是一种“警方+技术提供方”的二元封闭

^{④7} 参见前注^{④6},佟冶、颜坤星文,第17页。

^{④8} 朱晓瑜、郑挺:《公安大数据场景下的数据治理研究》,载《公安学刊》2020年第3期,第114页。

^{④9} 参见山东省青岛市中级人民法院(2016)鲁02刑终397号刑事判决书;山西省吕梁市离石区人民法院(2019)晋1102刑初137号刑事判决书;广东省广州市越秀区人民法院(2015)穗越发刑初1601号刑事判决书。

^{⑤0} Kelly K. Koss, *supra* note 24, 329.

^{⑤1} 参见前注^{④6},佟冶、颜坤星文,第16页。

模式,也就是由警方负责提出建设要求和相关数据,由技术提供方负责设计算法。^{⑤②}在这种模式下,算法进入刑事司法程序缺乏来自第三方的审查和评估,这对确保算法的准确性与合理性是明显不利的。

第一,有些算法所依据的理论基础可能是不可靠的,缺乏第三方审查可能导致不合理的算法进入刑事司法程序,冲击刑事诉讼的价值目标。近年来,随着大数据技术的迅猛发展,算法开发的步伐不断加快,一些公司开发出了功能强大、对执法机构极具吸引力的算法。然而,有些算法所依据的理论基础并不可靠,甚至可能存在严重问题。例如,美国联邦政府很早就开始测试一种算法,该算法能够根据一个人的生理数据预测该人是否意图实施恐怖行为。^{⑤③}挪威的研究走得更远,一个研究项目尝试将某些儿童预先归类为“可能的罪犯”。挪威政府收集了大量公民的数据,并将这些数据与单一的身份文件关联起来,研究人员希望通过挖掘儿童及其父母的身份文件,通过儿童出生时的环境来预测该儿童是否会在18岁生日之前实施犯罪。^{⑤④}这种算法对执法机构具有吸引力,因为该算法声称,它们能够将可能危害社会的危险分子精准地筛查出来。然而,只要稍加分析就不难发现,这种算法其实就是建立在天生犯罪人理论的基础之上,而天生犯罪人理论的缺陷毋庸多言,建立在该理论基础上的算法的准确性是难以得到保障的。退一步而言,即便算法的准确性能够得到保障,这些算法的内在逻辑——牺牲某些特定群体的利益以实现社会安全——也是令人不安的。

第二,构建算法时需要对一些重要的法律价值进行权衡和取舍,缺乏第三方审查可能导致算法对法律价值的选择背离法律的原则与精神。预测警务算法的构建包含许多复杂的步骤,其中很多步骤都涉及对不同法律价值的判断和取舍,因而构建预测警务系统时需要由执法机构之外的第三方进行审查和评估。例如,“在构建算法系统时,必须选择如何权衡不同的错误:假阳性和假阴性。这意味着在开发预测警务系统时必须判断哪一种情况更糟糕:将无辜者视为潜在的罪犯(假阳性)或者让算法忽视潜在的犯罪(假阴性)”;“编程人员必须对分析方法进行‘校准’,使之偏向于一种错误而不是另一种,或者,如果可能的话,必须尝试平等地对每一种算法错误进行权衡。每一种‘校准’不仅是一种技术决策,也是一种政治决策”。^{⑤⑤}很显然,在对上述问题进行权衡和选择的时候,执法机构基于打击犯罪的立场,可能会作出与法律专家截然不同的选择;负责开发算法的技术专家也有可能作出与法律专家完全不同的选择,因为科学技术具有追求效率的内生动力,而公正具有难以量化的特点,当公正与效率发生冲突时,技术专家往往会倾向于选择效率。例如,在假阳性错误(冤枉无辜)和假阴性错误(放纵犯罪)之间,执

⑤② 参见李金锋、陈如超:《大数据警务风险及其规制》,载《中国人民公安大学学报(社会科学版)》2023年第5期,第135页。

⑤③ See Jackson Polansky & Henry F. Fradella, *Does Precrime Mesh with the Ideals of U.S. Justice: Implications for the Future of Predictive Policing*, *Cardozo Public Law, Policy and Ethics Journal*, Vol.15:253, p.268 (2017).

⑤④ *Id.*, 258-259.

⑤⑤ See Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, *Fordham Urban Law Journal*, Vol.46:364, p.381-382 (2019).

法机构和技术专家可能认为后者更加难以容忍。又如,在保护公民个体权利与实现社会安全方面,执法机构和技术专家可能认为后者更值得保障。为避免算法中法律价值选择的失衡,算法的设计过程不仅需要技术专家的参与,也需要法律专家的参与。即使因为客观原因,在设计算法时未能邀请法律专家参与,也必须保证在执法机关将其应用于刑事司法系统之前由法律专家对算法进行审核,确保算法对法律价值的选择不悖离法律的原则和精神。

笔者认为,我国在未来建构预测警务制度时,也应当建立严格的算法审核机制。具体而言,可以在省级以上公安机关设立专门的算法审查委员会,市、县级公安机关开发或者引入预测警务系统时,必须申请省级以上公安机关算法审查委员会进行审查。算法审查委员会由法律专家和技术专家组成,同时可以吸收社会公众参与。委员会应当审查算法所依据的理论基础是否可靠,是否会造成对某些群体的歧视,是否与法律的基本原则、规则存在冲突等。只有经审查,算法在以上方面都没有问题,执法机关才能予以引进。事实上,这也是开发与应用人工智能的许多国家、地区的共同要求。如欧盟基本权利机构在2022年12月发布的《算法中的偏见:人工智能与歧视》中就呼吁,人工智能算法的开发与应用者应当就算法对公民及其基本权利可能造成的影响进行审查,“实现技术创新与基本权利的平衡是欧盟规制人工智能及相关技术的政策努力的核心”^{⑤6}。

(三) 规制预测警务系统的设置与使用,确保对公民权利的保障

在预测警务模式下,受益于大数据带来的技术红利,国家管理社会和预防犯罪的能力获得大幅提升,但与此同时,国家公权力与公民个人权利的冲突样态也发生了变化。在传统警务模式下,犯罪事实发生以后警方才会采取执法措施;而在预测警务模式下,犯罪尚未发生时警方就会积极主动预测,并采取相应的执法措施。在传统警务模式下,警方执法一般针对特定嫌疑人;而在预测警务模式下,警方执法经常针对不特定人、不特定群体。这两大特征导致在预测警务模式下,警察权力更容易被滥用,公民权利更容易被侵犯,因而必须对警察权的行使进行严格规制,防止公民权利受到侵犯。

1. 建立预测警务数据库的严格管理制度,确保公民隐私和个人信息的安全

长期以来,我国刑事诉讼法对公民隐私权、个人信息保护都明显不足,立法中有关公民隐私权与个人信息保护的规定寥寥无几。不仅如此,有限的与公民隐私权和个人信息保护相关的条款也主要适用于传统的对已发生犯罪的侦查,并不适用于预测警务活动。可以说,我国预测警务活动中公民权利的保护尚处于立法空白状态。这种立法现状导致我国预测警务数据库的建立缺乏规制,大量适用于人员预测的数据库(类似于预测警务3.0的数据库)被建立起来,数据采集也存在无序扩张的风险。以某地建立的禁毒数据库为例,该数据库目前采集的主要是涉毒人员的数据,然而,有学者因为担心“信息采集不足,无法做到完整而准确的毒情态势感知”,主张“数据库系统不仅要关注已经初

^{⑤6} European Union Agency for Fundamental Rights, *Bias in Algorithms-Artificial Intelligence and Discrimination*, Publications Office of the European Union, 2022, p.17, at https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf (Last visited on June 15, 2024).

步标定的涉毒目标数据,还要广泛采集和存储大量看似并未涉毒的目标数据,即全维度地采集社会综合数据信息”。^{⑤7}扩大数据库的规模虽然对于预测和控制犯罪具有意义,但如果数据的采集不受任何限制,可能会对公民权利构成威胁。因此,我国未来立法应当建立严格的约束机制,以实现预测犯罪与保障公民基本权利的平衡。

(1) 对人员数据库的入库信息进行分类管理

如前文所述,预测警务分为 1.0、2.0 和 3.0 阶段,与之相应,预测警务数据库可以分为两类:一类是服务于犯罪地点预测的数据库,主要是预测警务 1.0 和 2.0;另一类是服务于相关人员预测的数据库(简称人员数据库),主要是预测警务 3.0。不同于前者,后者(人员数据库)通常需要储存大量的公民个人信息,如姓名、住所、职业、身份证号码等信息,有些数据库可能还需要储存公民的生物特征、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等敏感个人信息。这类数据库的建立和使用会对公民的个人信息权和隐私权构成威胁,因此有必要对人员数据库中的数据进行分类管理。犯罪人或者被追诉人由于被认定为实施或者涉嫌犯罪,因而对执法机构收集和储存其个人信息负有一定的忍受义务,执法机构不仅可以收集和储存其非敏感个人信息,也可以收集和储存其敏感个人信息。但被害人和证人等并没有被认定为实施或者涉嫌犯罪,收集和储存他们的敏感个人信息可能导致刑事执法活动价值取向的失衡。因而未来立法时有必要明确规定,除非征得被害人、证人的同意,执法机构不得收集和储存他们的敏感个人信息。事实上,有些国家和地区已经注意到了这一问题,例如,《欧洲议会和理事会指令 2016/680》第 31 条明确规定,“应当尽可能明确区分不同类别数据主体的个人数据,例如:犯罪嫌疑人、被定罪的人、被害人以及其他诉讼参与人,如证人、掌握相关线索和联系的人、嫌疑人或者罪犯的同伙”^{⑤8}。

(2) 对人员数据库的使用进行严格规制

人员数据库中的信息与公民身份是一一匹配的,如果不对其适用进行严格规制,容易导致侵犯公民隐私权和个人信息权;同时,人工智能算法能够对海量数据进行挖掘、整理和分析,如果不对人员数据库的使用进行必要限制,很容易导致警方运用人工智能算法,通过信息碰撞实现对公民各个方面信息的碎片化整合,实现对公民生活、工作、社交等各个方面的“全景监控”,侵犯公民个人隐私。正因为如此,域外有些国家限制甚至禁止广泛收集公民个人信息进行犯罪预测。例如,荷兰曾开发了一款用于预防和打击社会保障领域欺诈行为的预测警务系统——“系统风险提示”(System Risk Indication)预警系统。为了保证预测效果,该系统需要收集大量公民个人信息,如就业数据、公民一体化数据(civic integration data)、债务数据、健康保险数据、个人基本资料(姓名、住址、出生日期等)。由于该数据库可能对公民隐私构成巨大威胁,2020 年荷兰海牙地区法院裁定,因为该预测警务系统侵犯了《欧洲人权公约》第 8 条所规定的隐私权,因而不得使

^{⑤7} 参见梁险峰:《侦控一体化导向下的禁毒数据库建设》,载《广西警察学院学报》2023 年第 3 期,第 122 页。

^{⑤8} Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016, Official Journal of the European Union, April 5, 2016, p.93.

用。^{⑤⑨} 借鉴域外经验,建议我国未来对人员数据库的使用设置严格的控制机制。首先,人员数据库只能适用于比较严重的犯罪,如可能判处3年以上有期徒刑以上刑罚的犯罪。其次,人员数据库的使用必须经过严格的批准程序,如必须经过县级以上公安机关负责人批准。最后,建立数据的定期删除制度,如对于超过一定期限的刑事案件,必须删除数据库中相关人员的数据。

2. 除非有其他证据证明相对人实施了涉嫌犯罪的行为,不能仅凭大数据预测结论对其权利进行限制或者剥夺

在传统警务模式下,警察不可能仅仅根据一个人的身份特征对其权利进行限制或者剥夺。很难想象,警察仅仅因为一个人处于失业状态、居住在贫民区或者属于某个特定族群就对其采取拦截或者盘查措施。在传统执法模式下,警方必须掌握了一定的证据证明相对人可能实施了违法犯罪行为才能对其权利进行限制或剥夺,如执法人员自己目击相对人实施了违法犯罪行为,或者有证人指认相对人实施了违法犯罪行为,或者相对人身上、住所遗留有犯罪证据等。但在预测警务模式下,这一切很可能发生变化。如前文所述,大数据预测往往是根据周围环境、人员构成、收入状态、年龄、性别等群体性特征来判断某个地区是否是犯罪“高风险区”,或者某个人是否应当被列入犯罪“热点名单”。由于算法具有“黑箱效应”,同时执法人员往往存在“自动化偏见”(automation bias)^{⑥⑩}和“确认偏见”(confirmation bias)^{⑥⑪},因而执法人员很可能对算法存在不切实际的依赖,在算法背后的运作机理尚未得到充分证明的情况下就将其奉若圭臬,完全依赖算法的结论实施警务行为。这就很有可能导致仅凭一个人的身份特征就对其权利进行限制或者剥夺。“很难想象现实社会中会出现这样的场景:群体特征本身就达到了合理怀疑的标准,但是在机器预测时代,理论上是有可能达到这一结果的。”^{⑥⑫}

仅仅因为一个人的身份特征而非具体行为就限制或者剥夺其权利不仅是错误的,而且是非常危险的。首先,这打破了国家公权力与公民个人权利之间的平衡关系。根据法治的一般原则,国家公权力与公民个人权利之间应当维持一种合理的比例和平衡关系,只有当公民个人存在危害社会的现实风险时,国家公权力才能对公民个人权利进行限制或者剥夺。而只有当公民可能实施了危害社会的违法犯罪行为时,才能认为其存在危害社会的现实风险。如果没有任何证据证明公民可能实施了危害社会的行为,不能仅仅因为其具有某种群体性特征,如属于失业群体或者特定族群,就认定其存在危害社会的现实风险,进而采取限制或者剥夺其权利的措施。否则,将打破国家公权力与公民个人权利之间的平衡关系,突破国家行使公权力的正当性底线。其次,仅凭身份特征就限制或

^{⑤⑨} See Litska Strikwerda, *supra* note 2, p.423-424.

^{⑥⑩} 自动化偏见是这样一种心理倾向,即相信计算机生成的结论要优于人工生成的结论。See Molly Griffard, *A Bias-Free Predictive Policing Tool: An Evaluation of the NYPD's Patternizr*, *Fordham Urban Law Journal*, Vol.47:43, p.54 (2019).

^{⑥⑪} 确认偏见,是指人们一旦接受了某种假定,认为其是正确的,那么此后其处理信息的模式将会发生改变,会搜索和记忆那些能够证明假定成立的信息,并以能够证明假定成立的方式解释信息。换言之,人们在接受假定之后心理上就会努力证明假定而不是推翻假定。*Id.*, 55.

^{⑥⑫} Ric Simmons, *supra* note 39, 986.

者剥夺公民权利会导致严重的歧视和社会不公。如果允许仅凭一个人的身份特征对其权利进行限制或者剥夺,那么,从理论上说,具备该身份特征的每一个社会成员,无论如何谨小慎微地约束自己的行为,只要其没有摆脱自己的身份特征,就可能随时面临被限制或者剥夺权利的风险。“仅仅因为一个人属于某个特殊的群体(a specific group)就对他实施拦截、搜查或者逮捕是不合适的,这样做等于是说一个人的群体特征是如此可疑,以至于任何特定的时刻都有理由相信他有可能从事犯罪活动。”^③这意味着某些群体成员仅仅因为身份烙印就被迫承受来自公权力的更多干预甚至侵犯,而无法与一般社会成员一样享受到平等的权利保障,这将动摇公正、平等、正义等现代法治的基石。

基于以上原因,笔者认为,除非有其他证据证明相对人实施了违法犯罪行为,否则不能仅仅依据大数据预测结论对其权利进行限制或者剥夺。“仅仅因为一个人属于某一特定的阶层或者处于某个群体就对其采取措施,这种做法无论在道德上还是法律上都是错误的。这就是为什么预测算法也需要将被告人的特定行为作为考虑因素的至关重要的原因,预测算法必须这样设计——不将预测完全建立在身份和地位的基础上。换句话说,预测不能建立在一个人是谁的基础之上,必须建立在一个人做了什么的基础之上。”^④一些域外国家或地区已经注意到这一问题,并在立法上作出明确规定。例如,欧盟2024年颁布的《人工智能法案》(Artificial Intelligence Act)第42条明确规定:“根据无罪推定原则,永远只能依据自然人的实际行为对其作出判定。绝不能在缺乏客观可核实的事实基础证明一个人卷入刑事犯罪,且缺乏人工评估的情况下,由AI仅仅基于一个人的概貌、个人特征或者性格,如国籍、出生地、住所、子女数量、债务水平或者车型对一个人的行为作出预测”^⑤。

随着我国预测警务实践的快速发展,一些地方已经运用预测算法打击犯罪。以江苏省海安市为例,该市公安系统搭建了“流窜盗窃犯罪预警模型”,“通过数据赋能,加强对比碰撞,实现‘露头就打’。2023年12月以来,海安公安通过巡防预警抓获盗窃嫌疑人48名”。^⑥在预测警务法律规制不完善的情况下,未来有可能出现主要依靠、甚至完全依靠算法预测结论限制或者剥夺公民基本权利的现象。为防范这种风险,建议我国在未来立法时明确规定,除非有其他证据证明相对人实施了特定的违法犯罪行为,否则不得仅仅依据大数据预测结论对相对人的权利进行限制或者剥夺。值得注意的是,我国《个人信息保护法》第24条第3款明确规定:“通过自动化决策方式作出对个人权益有重大影响的决定,个人有权要求个人信息处理者予以说明,并有权拒绝个人信息处理者仅通

^③ Ric Simmons, *supra* note 39, 985.

^④ Ric Simmons, *supra* note 42, 1076.

^⑤ European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM(2021)0206- C9-0146/2021 -2021/0106(COD), at https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (Last visited on June 15, 2024).

^⑥ 参见杨青青、曹钰华、邓茹萍:《海安:“大数据+”巡防模式提升街面警务战斗力》,载《人民公安报》2024年3月25日,第3版。

过自动化决策的方式作出决定。”这一规定值得我国在将来建构预测警务的相关法律制度时予以借鉴。

3. 保障被追诉方有权知悉警方进行大数据预测所依据的算法,并有权对算法的科学性、合理性提出质疑

(1) 保障被追诉方对算法的知情权是程序正义的基本要求

程序正义有两项重要要求:保障个体的尊严和确保个体对程序的充分参与。前者要求当事人在刑事诉讼中受到人道和有尊严的对待,如果被提起刑事指控,有权知悉指控的罪名及指控所依据的证据、事实和法律。后者要求当事人有充分的机会参与诉讼,全面收集和提出对本方有利的证据,有效反驳对自己不利的证据。

在一般人的印象中,算法是一套机械化的运算程序,会抹杀公民作为人的个体化特征,甚至将公民物化为一个个抽象的数据和概率值;在算法的世界里,公民个人更像是流水作业生产线上的产品而不是有血有肉的人。在这种认知背景下,如果相对人仅仅知道其被限制或者剥夺权利是算法预测的结果,而既不被告知算法预测所依据的因素及各因素的权重,也不被告知算法运作的机理,那么他就很容易产生不被尊重、没有尊严的感觉,也丝毫体会不到自己作为程序参与者的主体地位,因而,相对人很容易认为程序是不公正的,甚至会对程序产生敌意。为了解决这一问题,立法必须明确规定执法人员有向被追诉人解释算法的义务,确保被追诉人对算法的知情权。如果被追诉人了解算法的运作机理,知悉支撑算法的理论基础,特别是知晓算法只是协助执法人员作出判断的依据之一,自己被限制或者剥夺权利的主要原因在于自己实施了涉嫌违法犯罪的行为,这样其认为自己被客体化、丧失尊严的认知将大为弱化,甚至完全消失。

(2) 保障被追诉方对算法的知情权有利于强化对算法的监督

如前文所述,算法是有可能出现错误的,因而必须强化对算法的监督。对算法的监督可能来自公权力机构,如在引入算法之前由相关机构对其进行审查和评估。然而,仅仅依靠公权力机构对算法进行监督显然是不够的。公权力机构行政化的运作方式及其对算法错误在感知上的疏离使其对算法错误的监督往往较为迟滞和被动。要解决这一问题,最有效的办法就是引入私权利主体,特别是被追诉方的监督。被追诉方是算法判断结果的直接承受者,一旦算法出现错误,其很有可能被错误地限制甚至剥夺权利,因而被追诉方有强大的动力去探寻和质疑算法可能存在的错误。

要实现被追诉方对算法的监督,就必须保障被追诉方有知悉算法的权利。而且,这种知悉权应当是全面的:不仅应当有权知悉算法的结论,而且应当有权知悉警方是依据何种算法得出判断结论的,还应当有权知悉算法的运作机理以及背后的理论依据。美国有学者明确主张,刑事辩护律师应当致力于推动检察官披露在逮捕和起诉被告时使用的预测性警务工具(predictive policing tools)和决策支持模型(decision support models)。美国纽约州2020年1月1日生效的一项法律要求“检察官在被告被传讯后的15个工作日内与被告共享其发现的材料”。根据该法律,检察官需要与被告共享的材料范围非常广泛,警方运用算法得出判断的过程“应当被解释为法律执行机构的报告(law

enforcement report)”,属于应当向被告披露的内容。^{⑥7}

为了充分保障被追诉人的合法权益,同时强化对算法的监督,建议我国在未来立法时充分保障被追诉方对算法的知情权,规定警方在依据算法对被追诉人的权利进行限制或者剥夺时应当向被追诉人告知警方使用的算法,当被追诉人及其辩护人要求知悉算法的运作机理和理论基础时,执法机关应当告知并进行详细解释。值得注意的是,我国《个人信息保护法》第 24 条第 3 款明确规定:“通过自动化决策方式作出对个人权益有重大影响的决定,个人有权要求个人信息处理者予以说明”;第 48 条进一步规定:“个人有权要求个人信息处理者对其个人信息处理规则进行解释说明。”这两条规定值得我们在将来建构预测警务的相关法律制度时予以借鉴。

Abstract: The emergence of big data predictive policing has greatly improved the capacity of police to solve cases, prevent and crack down on crimes, but it also brings about certain risks that need to be regulated. Extraterritorial predictive policing has gone through three stages, namely 1.0, 2.0 and 3.0. The prediction capacity has been continuously improved, and the operation mechanism has been gradually optimized. The development of big data predictive policing faces two risks: first, it is hard to ensure the accuracy, completeness, and timeliness of data; second, it is difficult to avoid or correct errors and discrimination in the algorithm. Big data predictive policing shall be regulated in the following three ways: first, the procedures for data collection and processing should be standardized to ensure the quality of data; second, the algorithm review mechanism should be established to supervise and evaluate the accuracy and risks of the algorithm; third, the establishment and use of predictive policing systems should be strictly regulated to ensure the protection of civil rights.

(责任编辑:王 楷)

^{⑥7} See Molly Griffard, *supra* note 60, 75, 78.