

逮捕社会危险性量化评估研究*

——以自动化决策与算法规制为视角

高通

摘要:逮捕社会危险性量化评估主要是使用形式化、量化或精算方法来预测犯罪嫌疑人、被告人的社会危险性,其自 20 世纪 80 年代以来在世界范围内快速发展。逮捕社会危险性量化评估以法律现实主义以及循证实践为思想基础,以降低审前羁押率为目的,并以科学技术的快速发展为支撑。从实践来看,逮捕社会危险性量化评估模型的预测正确率只达到中等程度,且并未超越司法人员的预测正确率。逮捕社会危险性量化评估在带来便利的同时,也在科学性和公正性方面存在多重风险,如固化司法实践中既存的偏见歧视以及剥夺当事人辩护权等。为保障逮捕社会危险性量化评估的有效适用,一方面应建构科学且公正的评估模型,另一方面应确保逮捕社会危险性量化评估在建构和实施过程中受到正当程序原则的规制。

关键词:逮捕 社会危险性 自动化决策 算法规制

中图分类号:DF61 **文献标识码:**A **文章编号:**1673-8330(2021)06-0131-14

DOI:10.13893/j.cnki.bffx.2021.06.011

一、问题的提出

逮捕条件是逮捕制度改革的重要内容。我国逮捕条件的改革主要包括严格证据条件和细化社会危险性条件两方面内容,而近些年来细化逮捕社会危险性条件的改革获得立法者更多的青睐。^①学者们也开始反思逮捕社会危险性条件与证据和罪责条件的关系,逐渐认可逮捕社会危险性为逮捕条件的核心。^②但当逮捕制度改革聚焦于细化逮捕社会危险性条件时,逮捕社会危险性条件精确性不足、缺乏可操作性等问题也逐渐暴露出来。^③这些问题不仅降低了法律控制的密度,也减损了法律保留原

* 本文系南开大学亚洲中心资助项目“刑事诉讼中域外证据使用问题研究”(AS2101)的阶段性成果。

[作者简介]高通,南开大学法学院副教授,南开大学·中国社会科学院大学 21 世纪马克思主义研究院研究员,法学博士。

① 例如,1979 年《刑事诉讼法》对逮捕社会危险性条件仅规定为“有逮捕必要”的表述,1996 年《刑事诉讼法》则扩展为“不足以防止其发生社会危险性,而有逮捕必要”,2012 年《刑事诉讼法》删除“有逮捕必要”并将社会危险性条件细化为五种情形,2018 年《刑事诉讼法》进一步规定将“涉嫌犯罪的性质、情节,认罪认罚等情况”作为逮捕社会危险性审查的考量因素。

② 参见孙谦:《司法改革背景下逮捕的若干问题研究》,载《中国法学》2017 年第 3 期,第 30 页;杨依:《以社会危险性审查为核心的逮捕条件重构——基于经验事实的理论反思》,载《比较法研究》2018 年第 3 期,第 130 页。

③ 参见李训虎:《逮捕制度再改革的法释义学解读》,载《法学研究》2018 年第 3 期,第 156 页。

则的功能。^④ 所以,实现逮捕社会危险性条件的精确化日益成为我国以及其他国家或地区逮捕制度改革共识。逮捕社会危险性条件的精确化主要有两种方法,一种方法是通过立法、判例等进一步细化逮捕社会危险性条件以强化立法的密度,另一种方法则是在逮捕社会危险性审查中引入风险评估机制来强化审查的科学性与公正性。两种方法并行不悖。但随着逮捕功能向防范社会公共风险的转变以及大数据、人工智能等技术在司法领域的广泛应用,在逮捕审查中引入风险评估的做法得到越来越多的认可,从而使得逮捕社会危险性量化评估在世界范围内快速发展起来。美国、英国、加拿大等国家的刑事司法实践中逐步将危险性评估作为逮捕审查的重要内容,传统大陆法系国家如德国、法国等也开始试点使用危险性量化评估。^⑤

虽然我国对逮捕社会危险性条件的改革主要是沿着细化社会危险性条件的思路进行,但司法实践中也试图将量化评估方法引入逮捕社会危险性判断之中。早在 2005 年,上海市闵行区人民检察院就曾开发出未成年人非羁押措施可行性评估制度,之后北京、太原等地都曾开发出多种量化评估工具。^⑥ 2017 年 2 月,中央政法委部署由上海市高级人民法院研发“推进以审判为中心的诉讼制度改革软件”任务,该软件中就包括逮捕社会危险性的量化评估系统。但与实务界对量化评估的热衷不同,理论界对逮捕社会危险性量化评估的实践发展及其对刑事诉讼的影响等问题目前仍缺乏深入的探讨。随着科技对刑事司法介入程度的日益加深,运用定量研究方法来量化逮捕社会危险性在刑事司法实践中的应用也将越来越广泛。但量化评估逮捕社会危险性也是一把双刃剑,其在提升诉讼效率的同时,也会带来一定的风险。所以,如何利用量化评估方法来更好地评估逮捕社会危险性是逮捕制度改革的重要理论课题。基于上述考虑,本文以逮捕社会危险性量化评估为研究对象,通过研究试图回答如下几个基础性问题:第一,逮捕社会危险性量化评估的发展背景和基本原理分别是什么?第二,逮捕社会危险性量化评估可能存在哪些科学性和公正性风险?第三,如何更好地在判断逮捕社会危险性过程中使用量化评估方法?

二、逮捕社会危险性量化评估的发展背景与基本原理

逮捕社会危险性评估是犯罪预测的一种表现形式。犯罪预测在犯罪学中并非一个全新的课题,但通过风险评估工具来预测再犯风险则是最近一百年的事情。20 世纪 20 年代,芝加哥大学社会学家欧尼斯特·伯吉斯开发了世界上首个风险评估工具,利用罪犯的社会经济特征等因素来预测其再次犯罪的风险。^⑦ 20 世纪 80 年代之后,风险评估在世界范围内迅速发展起来。时至今日,风险评估已经遍布刑事司法程序的各个阶段,各种风险评估模型也层出不穷。据统计,当前世界范围内有超过 200 种风险评估工具被用于刑事司法和司法精神病领域。^⑧ 风险评估,包括逮捕社会危险性量化评估,在美国刑事司法实践中的应用最为广泛。其中,应用较普遍的逮捕社会危险性评估工具包括美国肯塔基州的公众安全风险评估系统(以下简称“PSA 系统”)、美国联邦审前风险评估工具(以下简称“PTRA 系统”)以及最新的以替代惩罚为目标的惩教管理分析系统(以下简称“COMPAS 系统”)。^⑨

^④ 参见林钰雄:《刑事诉讼法》(上册·总论编),中国人民大学出版社 2005 年版,第 232 页。

^⑤ 参见[德] Hans-Jürgen Kerner:《德国刑事追诉与制裁》,许泽天、薛智仁译,元照出版社 2008 年版,第 108—109 页。

^⑥ 参见王贞会:《审查逮捕社会危险性评估量化模型的原理与建构》,载《政法论坛》2016 年第 2 期,第 74 页。

^⑦ Gwen J. Eijk, *Inclusion and Exclusion through Risk-Based Justice: Analysing Combinations of Risk Assessment from Pretrial Detention to Release*, 60 *British Journal of Criminology* 1080, 1082 (2020).

^⑧ Douglas T., Pugh J., Singh I., Savulescu J., Fazel S., *Risk Assessment Tools in Criminal Justice and Forensic Psychiatry: The Need for Better Data*, 42 *Europe Psychiatry* 134, 134 (2017).

^⑨ 美国肯塔基州作为保释制度改革的试点州,其开发的 PSA 系统是目前在应用最广泛的审前风险评估模型。PTRA 系统是美国联邦拘留托管所办公室于 2009 年开发的审前风险评估模型,其在美国联邦系统内广泛适用。2017 年美国联邦地区法院中接近 90% 的被告人通过 PSA 系统评估审前风险。COMPAS 系统是美国最新一代的风险评估工具,其可以针对再犯、逃避审判风险等提供单独的评估。

在我国,上海市高级人民法院开发的刑事案件智能辅助办案系统(以下简称“上海206系统”)、贵州基于大数据开发的智能办案辅助系统等也引起社会的高度关注。

(一) 逮捕社会危险性量化评估的发展背景

逮捕社会危险性量化评估之所以在世界范围内快速发展,理论基础、现实动因及技术支撑等因素缺一不可。

第一,法律现实主义和循证实践是逮捕社会危险性量化评估产生与发展的思想基础。法律现实主义是西方社会20世纪初期发展起来的一个法学流派,主张要关注法律的实际运作,特别关注刑事司法中的司法行为问题。^⑩法律现实主义主张要进行实证性研究或经验性研究,并通过科学方法收集和分析实际的数据来检验假设是否成立。^⑪世界上首个风险评估模型是由社会学家通过统计分析建立的,之后的各种风险评估也是建立在实证性研究基础之上,并遵守社会科学研究的基本规范。相反,风险评估在传统法律形式主义下很难获得发展,数据量化甚至被视为一种危险。^⑫所以,风险评估的产生明显受到法律现实主义的影响。此外,循证实践(Evident - Based Practice)也影响到逮捕社会危险性量化评估的发展。循证实践最初产生于医学领域,强调要严格评估行为的实践效果,后该理念被引入刑事司法领域。^⑬时至今日,社会危险性量化评估已经成为循证刑事司法最前沿的领域。

第二,降低过高的审前羁押率是逮捕社会危险性量化评估快速发展的现实动因。传统上逮捕的主要目的在于预防犯罪嫌疑人、被告人逃避审判以确保程序的顺利进行。但20世纪中期以后因社会更加关注社会公共安全以及行政权力扩张等原因,预防性羁押在各国和地区刑事司法实践中大量出现。但由于预防性羁押的羁押标准并不清晰,^⑭司法实践中出现法院疏于审查犯罪社会危险性或者简单以支付保释金能力来判断其社会危险性等情形,由此造成审前羁押率的大幅上升。如美国联邦司法系统在1984年之前被告人被羁押至审判的比例是24%,但2010年的审前羁押率已达到64%,曾在刑事诉讼程序中某个时间段被羁押的比例则达到76%。而且金钱保释在保释中所占比例大幅增加,如美国联邦司法系统在1990年至1994年间无保释放和保释释放的比例分别为41%和24%,2002年至2004年间二者的比例则分别为23%和42%。^⑮事实上,我国逮捕实践中引入量化评估也是基于降低羁押率的考虑。^⑯所以,羁押率上升是逮捕社会危险性量化评估快速发展的现实原因。

第三,科学技术的发展为逮捕社会危险性量化评估提供了技术支撑。人身危险性对强制措施适用、量刑以及行刑等均有重要影响,刑法学中将这一理论称为人身危险性理论。但人身危险性理论又是一个非常有争议的理论,争议的核心便是人身危险性评估的科学性和可操作性。无论是人身危险性理论的支持者还是反对者,均承认人身危险性评估并非一件易事。^⑰如果能开发出科学的人身危险性评估方法,人身危险性理论及量化评估将在刑事司法领域中发挥更大的作用。随着科技的发展,制约人身危险性评估科学性和可操作性的障碍正在逐步被攻克。早期的风险评估主要依赖于观察、模拟、访谈等实验室研究方法,风险评估的原始数据以及自变量均不多,使得早期风险评估模型的预

^⑩ 胡铭:《法律现实主义与转型社会刑事司法》,载《法学研究》2011年第2期,第53页。

^⑪ 参见邓矜婷:《新法律现实主义的最新启示与发展》,载《法学家》2014年第4期,第11页。

^⑫ [荷]马丁·W.海塞林克:《新的欧洲法律文化》,魏磊杰、吴雅婷译,中国法制出版社2018年版,第152页。

^⑬ Megan Stevenson, *Assessing Risk Assessment in Action*, 103 *Minnesota Law Review* 303, 312—313 (2018).

^⑭ 美国联邦最高法院曾在合众国诉萨勒诺案中指出,如果预防性羁押是不合理的或过度的或其带有惩罚性目的,那么预防性羁押就违反正当程序条款。但究竟何种程度会构成过度或不合理,美国联邦最高法院并未有明确指示。See *United States v. Salerno*, 481 U.S. at 746—48.

^⑮ Sandra G. Mayson, *Dangerous Defendants*, 127 *The Yale Law Journal* 490, 506—508 (2018).

^⑯ 参见钟亚维、招阳:《审查逮捕如何做到更客观精准 广州南沙:探索建立社会危险性量化评估机制》,载《检察日报》2020年10月10日第2版。

^⑰ 参见陈伟:《人身危险性评估的理性反思》,载《中国刑事法杂志》2010年第6期,第9—10页。

测错误率甚至高达 54% 至 99%。¹⁸ 但随着计算机以及大数据、人工智能等技术的发展,社会危险性评估实现了对较大数量数据复杂的统计分析,预测的准确性也有大幅提升。如通过大数据对刑事司法预测的正确率要比临床法的预测正确率高 10% 左右。¹⁹ 人工智能技术的发展也促成机器学习式风险评估的产生,这使得社会危险性评估摆脱了人类对犯罪认识水平的限制,甚至可发现新的犯罪模式。²⁰ 可见,科技的发展是逮捕社会危险性量化评估快速发展的技术支撑。

(二) 逮捕社会危险性量化评估的基本原理

虽然不同国家和地区开发出多种逮捕社会危险性量化评估模型,但其基本原理大同小异,基本上都是通过形式化、精算或算法等预测未来犯罪或违法行为的发生概率。²¹ 具体来说,逮捕社会危险性量化评估的基本原理主要包括如下五方面内容。

第一,逮捕社会危险性量化评估的功能在于预测犯罪嫌疑人、被告人再犯新罪、逃避审判以及实施干扰刑事诉讼顺利进行行为的危险。与审判针对的是已经发生的犯罪事实不同,逮捕社会危险性评估面向的是尚未发生的危险。所以,逮捕社会危险性评估本质上是一种风险预测手段,通过犯罪嫌疑人、被告人的一些特殊表现或特征来预测其发生社会危险性的可能性。结合逮捕的具体目的来说,逮捕社会危险性量化评估主要是用来预测犯罪嫌疑人、被告人再犯新罪、逃避审判以及实施干扰刑事诉讼顺利进行手段的危险。当然,不同评估模型的预测重点并不完全相同,有的模型更侧重于对犯罪嫌疑人、被告人再犯新罪或再犯特定犯罪可能性的预测,有的模型更侧重于对逃避审判风险的预测,还有的模型会对再犯新罪以及逃避审判风险进行综合审查。

第二,逮捕社会危险性量化评估模型主要是利用犯罪学理论中的社会学习理论,通过分析影响人们社会学习进程的因素以实现对未来发生社会危险性的预测。早期的社会危险性评估只进行统计分析而不注重理论建构,后来的评估模型则借鉴犯罪学中的社会学习理论、犯罪人格、犯罪社会心理学、社会控制等理论,并通过理论基础来整理评估模型中的变量。总体来看,社会危险性评估的主要理论基础还是社会学习理论。²² 社会学习理论是当前西方犯罪学主流理论之一,源自法国犯罪学家塔尔的模仿理论,其理论核心为:当一个人接触已经或有意愿实施犯罪行为的人时,该人实施犯罪和越轨行为的概率会增加而且遵守法律的概率会降低。²³ 依据社会学习理论,社会危险性评估主要通过审查影响犯罪嫌疑人、被告人社会学习进程的因素,以实现对对其社会危险性的预测。

第三,建构逮捕社会危险性量化评估模型时从单纯使用定性或定量研究方法到结合使用定性与定量研究方法。传统上对逮捕社会危险性的预测主要是由司法官员依据经验法则,结合法律规定、个案实际情况如已知的不利证据分量、人格及其私人关系等情形来综合决定。²⁴ 这种评估方法虽然有司法官员的职业素养以及职业经验等保障其准确性,但这种方法也因为主观性过强、容易产生偏见等原因而遭致批判。²⁵ 所以,20 世纪 90 年代之后的逮捕社会危险性风险评估更注重评估的科学性和专业性,评估通常通过相关专家对逮捕社会危险性进行专业评估或利用统计学方法建构逮捕社会危险

¹⁸ Danielle Kehl, Priscilla Guo & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, <https://cyber.harvard.edu/>, 2020-9-1.

¹⁹ Ric Simmons, *Qualifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, 2016 Michigan State Law Review 947, 953 (2016).

²⁰ Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 University of Pennsylvania Law Review 327, 369 (2015).

²¹ 前引¹⁸, 第 314 页。

²² Tim Brennan, William Dieterich, Beate Ehret, *Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System*, 36 Criminal Justice and Behavior 21, 23 (2009).

²³ Ronald L. Akers, Wesley G. Jennings, *The Social Learning Theory of Crime and Deviance*, in Marvin D. Krohn, Alan J. Lizotte, Gina Penly Hall eds., *Handbook on Crime and Deviance*, Springer, New York, 2009.

²⁴ 参见[德] Claus Roxin:《德国刑事诉讼法》,吴丽琪译,三民书局股份有限公司 1998 年版,第 324 页。

²⁵ 参见卢建军:《人身危险性评估的基本方法》,载《人民检察》2011 年第 14 期,第 77 页。

性评估模型来实现。定量研究方法强化了模型的客观性,但实践中的模型却陷入完全定量研究的“陷阱”中,无法摒除先前判决书中偏见、歧视等因素的影响。为克服定量研究存在的问题,当前的逮捕社会危险性量化评估模型大都使用定性定量相结合的方法,先运用定量研究方法选择变量和形成算法,再运用定性研究摒除有损公正实现的部分自变量。^{②6}在具体定量研究方法的选择上,早期的社会危险性评估模型主要使用线性回归、逻辑回归等分析方法,新近的评估模型已使用功能更为强大的多元统计分析、模型平均、混合模式集成以及人工智能深度学习等方法。^{②7}但从我国司法实践来看,逮捕社会危险性量化评估正处于从定性评估向定量评估转型的过程之中,建构量化评估模型时主要使用访谈、调研等定性研究方法,统计分析等定量研究方法发挥的作用较为有限。^{②8}

第四,虽然不同逮捕社会危险性量化评估模型的指标体系存在一定差异,但总体上围绕“核心八项指标”展开。由于各评估模型所依赖的原始数据以及算法不尽相同,各量化评估模型的指标体系也存在不同。如美国 PSA、PTRA 和 COMPAS 系统中分别有 9 个、13 个和 15 个指标,^{②9}我国的上海 206 系统中确定了 32 个逮捕社会危险性评价指标,广州市南沙区的逮捕社会危险性量化评估模型围绕人身危险性、社会危害性、诉讼可控性等形成了 43 个指标。但总体来看,经过犯罪学理论和实践的多年发展,主流的逮捕社会危险性量化评估模型中的评估因素已大致相同,形成所谓的社会危险性评估“核心八项指标”。核心八项指标包括犯罪历史或逃避诉讼历史、反社会人格、犯罪态度、犯罪联系、教育背景和工作情况、家庭和服役情况、药物滥用以及娱乐和休息情况,这是逮捕社会危险性评估中的核心指标体系。^{③0}当然,为了预测方便,逮捕社会危险性评估中的评价指标数量不会太多,而且在指标选择上越来越侧重于无需访谈或问卷调查就可直接提取到的客观信息。

第五,逮捕社会危险性量化评估结果表现为风险分数或风险等级,该结果通常仅作为司法官员裁判的参考。首先,社会危险性量化评估的结果是风险分数或风险等级。如美国 PSA 系统中依据风险分数从低到高,将审前风险等级划分为五级。^{③1}当然,不同模型给出风险分数的方式并不完全相同。如美国 PSA 系统给出的是再次被逮捕、不出庭和因暴力犯罪被捕风险的综合分数,而 COMPAS 系统则会针对不同风险分别给出单个的风险分数。其次,逮捕社会危险性评估结果通常仅作为司法官员决定逮捕与否的参考,司法官员在决定逮捕与否时拥有决定权。如在美国威斯康星州诉卢米斯案(以下简称“卢米斯案”)中,威斯康星州最高法院认为,法官在使用此类风险评估时必须谨慎行事,而且使用风险评估的法官必须说明评估以外的其他支持判刑的因素。^{③2}但这一情况似乎有所改变,逮捕社会危险性评估结果对司法官员的影响作用正在强化。美国肯塔基州 2011 年公共安全与犯罪者责任法案规定,法官必须对所有被羁押候审的犯罪嫌疑人、被告人进行审前风险评估,并将所有低

②6 前引②,第 21—22 页。

②7 前引②,第 21、23—24 页。我国的上海 206 系统也已经使用到人工智能的深度学习算法来建构模型。参见崔亚东:《人工智能与司法现代化》,上海人民出版社 2019 年版,第 168 页。

②8 笔者调研广州市某区人民检察院开发的量化评估模型后发现,其在设计社会危险性的影响因素时使用定量研究与定性研究相结合的方法。除法律规定的 27 种情形外,又从三千余个案件中提炼出 16 种自变量。但在自变量权重涉及上则更依赖于定性研究,首先通过专家给分的方式确定不同自变量的权重值,之后再通过案例来检验模型正确与否。

②9 前引⑤,第 568 页。

③0 笔者以广州市某区人民检察院制作的 43 项逮捕社会危险性量化评估指标为例分析后发现,除“重要证件尚未完全固定”“可能判处三年以上徒刑”无法归入外,其余指标均可纳入核心八项指标中。具体来说,犯罪历史或逃避诉讼历史:有自杀自残经历、经传唤不到案或逃跑、接触相关诉讼人、威胁恐吓被害人或证人或举报人、毁灭物证;反社会人格:多次作案、连续作案、流窜作案、看守所表现恶劣、密谋其他犯罪、暴力抗拒抓捕;犯罪态度:自首、立功、坦白、认罪认罚、积极赔偿退赃、被害人谅解;犯罪联系:他人过错、防卫或避险过当、从犯、预备、中止、未遂、教唆犯、预谋犯罪、手段残忍、故意犯罪、数罪并罚、重点打击犯罪类型、犯罪期间、犯罪对象;教育背景和工作情况:年龄、教育程度、职业情况、是否存在身体残疾或精神障碍;家庭和服役情况:固定住所;药物滥用:是否有吸毒或赌博恶习等个人自然情况。

③1 参见张吉喜、梁小华:《美国司法部审前风险评估模型及其对我国的启示》,载《中国刑事法杂志》2010 年第 7 期,第 111 页。

③2 参见江溯:《自动化决策、刑事司法与算法规制——由卢米斯案引发的思考》,载《东方法学》2020 年第 3 期,第 78 页。

风险和中风险的被告人通过非金钱保释快速释放。^③

三、逮捕社会危险性量化评估的科学性与公正性风险

逮捕社会危险性量化评估的有效开展依赖于量化评估模型的科学性与公正性,虽然当前量化评估模型已经发展多代,甚至将大数据、人工智能等最新科技应用于建立评估模型过程中,但量化评估模型的科学性与公正性仍然是制约量化评估有效开展的主要因素。

(一) 逮捕社会危险性量化评估的科学性风险

量化评估模型的科学性风险既可能来自于人类对犯罪认识水平的限制,也可能来自于评估指标的不科学、原始数据质量的不高以及缺乏对评估结果正确性的科学评估等。

第一,逮捕社会危险性量化评估的准确性受人们对犯罪认识水平的限制,且量化研究方法在还原司法人员裁决心理时存在一定限制。社会危险性量化评估以法律现实主义为理论基础,将裁决视为理性的司法人员在掌握全部信息后作出的理性决策,并试图将其内心过程以形式化的方式呈现出来。^④ 由此推之,只要能够发现影响司法官员裁决羁押与否的所有因素,就可以建构量化评估模型以还原司法官员裁决时的心理。但这在司法实践中几乎不可能完全实现。因为无论量化评估模型是通过定量方法分析案例建立,还是依据规范性文件中的规定来建构,亦或兼采定量和定性方法来形成,社会危险性评估本身并不能超越人类对犯罪的认识水平。虽然大数据以及人工智能等技术可尽可能还原司法实践中判断逮捕社会危险性的指标和方法,并使得量化评估模型的预测正确率有了较大幅度的提升。但只要人类无法完全认识犯罪的形成机制,量化评估模型的预测准确率就不可能达到100%的程度。所以,逮捕社会危险性量化评估模型的预测准确率存在天然缺陷,受人类对犯罪认识水平的制约。^⑤ 此外,虽然社会危险性量化评估旨在发现司法官员裁决社会危险性时的心理,但其只是通过概念化、形式化的方式来描述司法人员的裁决过程以及影响因素,而非司法人员真正的心理过程。而且,描述过程中可能会出现失真、数据丢失等多种情形,无法完全体现司法人员的心理过程。

第二,逮捕社会危险性量化评估的指标体系不尽科学。量化评估指标是逮捕社会危险性量化评估的关键,但即便除去人类对犯罪认识水平以及量化研究方法的限制外,当前的量化评估指标在科学性上还有较大提升的空间。其一,逮捕社会危险性量化评估模型中的指标不够精细化。基于我国是成文法国家以及我国法律中对逮捕社会危险性已有较详细的规定,我国的逮捕社会危险性量化评估模型主要是依据法律规定来建构,评估模型中的绝大多数指标只是对法律规定的重复和细化。但这种细化却仍然较为粗疏,甚至影响到评估模型的准确性。如有无违法犯罪史是评估社会危险性的重要内容,我国立法和司法实践中也将有无累犯、前科、曾被行政处罚等情形作为评估犯罪嫌疑人、被告人社会危险性的重要因素。但实际上,违法犯罪史对社会危险性的预测是可进一步细化的,如美国几款社会危险性评估模型中将过往违法犯罪的发生时间、违法犯罪的种类等均纳入到评估中。其中原理已被实证研究所证明,虽然从犯罪学角度来说违法行为与再次犯罪风险间存在正相关,但早期违法行为模式越强其重新犯罪风险越高。^⑥ 所以,我国逮捕社会危险性量化评估模型中的指标体系仍然较为粗疏。其二,逮捕社会危险性量化评估中大量使用需要司法人员认定的量刑情节作为评价指标,有损量化评估的客观性。量化评估模型的客观性需要以指标的客观性为支撑,但我国社会危险性量化评估指标体系的客观性是有所欠缺的。如我国的社会危险性量化评估中直接使用部分量刑情节作

^③ Public Safety and Offender Accountability Act, H. B. 463, 2011.

^④ [美]里德·黑斯蒂:《陪审员的内心世界——陪审员裁决过程的心理分析》,刘威、李恒译,北京大学出版社2006年版,第5页。

^⑤ John Monahan, Jennifer L. Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*, 26 Federal Sentencing Reporter 158, 162 (2014).

^⑥ 参见曾贇:《服刑人员刑满释放前重新犯罪风险预测研究》,载《法学评论》2011年第6期,第136页。

为指标,如自首、立功、坦白、犯罪中止、正当防卫等。直接使用量刑情节作为评估指标虽然方便且易于操作,但也可能使得量化评估中带有更多的主观性判断。因为量刑情节并非仅仅是一种客观行为,还涉及到司法人员的主观认定问题,不可避免地会带有司法人员的主观判断。所以,世界上主流的社会危险性量化评估模型更倾向于使用客观性较强的指标,特别是无需访谈或问卷调查就可直接提取到的客观信息。其三,模型建构过程中受到较多推断、猜测等主观性因素的影响。理想的逮捕社会危险性量化评估模型应当是主要建立在定量分析基础之上的,通过定量分析来发现自变量、自变量的权重及之间的相关关系,以尽可能减少猜测、主观推断等因素的影响。但我国的社会危险性量化评估模型则更多使用定性为主、定量为辅的方法,特别是对于自变量的权重以及自变量之间关系我国更青睐于通过定性研究方法来确定,这也使得主观推断、猜测等对我国量化评估模型的影响更大。

第三,逮捕社会危险性量化评估所依赖的原始数据质量不高。逮捕社会危险性量化评估模型主要是运用统计学知识从原始数据中提取变量并建立数据模型,所以,原始数据的准确性对于模型的科学性有着至关重要的基础作用。但从当前原始数据获取途径以及更新情况来看,量化评估模型所依赖的原始数据在准确性、及时性方面都存在一些问题。其一,法律领域中数据的“表象性”特征制约着原始数据的客观性。虽然当前的逮捕社会危险性量化评估模型主要使用定量与定性分析相结合方法建立,但最基础的原始数据仍然是刑事案卷材料,如社会危险性判断的“核心八项指标”大都是从刑事案卷材料中能直接提取出来的信息。但由于我国法律领域的信息存在突出的表象性特征,使得从刑事案卷材料中提取出来的信息可能无法充分、真实地反映司法人员决策时真正采用的“实质信息”。^⑤如司法人员的性格特征、逮捕的社会效果以及领导的关照等,但这些内容在原始数据中很难体现出来。其二,当前的数据提取技术无法确保相关数据的准确提取。统计分析中的一个基础步骤就是将裁判文书中的相关信息提取出来并将其转化为可供计算机识别的语言,特别是随着大数据技术在评估预测中的应用,提取数据信息的准确性问题更加值得关注。因为要想在系统中录入百万份级别的裁判文书仅依靠人工是不现实的,必须要借助特殊的技术才能实现。虽然人工录入也会因工作人员知识结构限制而影响其准确性,但当前能达到人工录入准确性的技术还非常少见。如普通的“爬虫”技术虽然可以大规模检索词汇但无法识别字符的准确含义,即便是上海 206 系统依赖的光学字符识别系统也存在识别率瓶颈、干扰因素过多等问题。^⑥所以,如何确保数据信息提取的准确性是亟需解决的技术问题。其三,相关评估模型的数据信息缺乏及时更新。随着社会的发展和我们对社会危险性认识水平的提升,影响社会危险性的因素也应是不断变化的。这些变化也应当及时反映在社会危险性评估模型中,否则会严重影响到评估模型的准确性。但从相关评估模型的原始数据来看,这些模型很少会及时更新原始数据。如美国 COMPAS 系统依赖的是 2004 年 1 月至 2005 年 11 月间的裁判文书,PTRA 系统依赖的是 2001 年 10 月至 2007 年 9 月的裁判文书。我国的逮捕社会危险性评估模型尚处于初创阶段,所依赖的原始数据大都比较新,如上海 206 系统是以 2012 年至 2016 年的数据为支撑建构起来的。^⑦但 2016 年之后我国的刑事司法实践中出现较大变化,如果这些变化未能体现到模型中,那么模型的数据更新也是不及时的。

第四,对逮捕社会危险性量化评估结果的正确性缺乏科学评估。实施逮捕社会危险性量化评估的主要目的是降低羁押率,但通过量化评估来降低羁押率又不能影响到犯罪控制目标的实现,即量化评估不能为了降低羁押率而降低羁押率。所以,依据循证实践的要求,需要对逮捕社会危险性量化评

^⑤ 参见左卫民:《关于法律人工智能在中国运用前景的若干思考》,载《清华法学》2018年第2期,第115页。

^⑥ 参见前引②崔亚东书,第146页。

^⑦ 参见严剑漪:《揭秘“206”:法院未来的人工智能图景——上海刑事案件智能辅助办案系统164天研发实录》,载《人民法治》2018年第2期,第40页。

估结果的正确性进行科学评估。完整的逮捕社会危险性量化结果的评估应当包括两方面内容,一是评估量化评估对羁押率的影响,二是评估模型的预测正确率。其中,评估社会危险性量化评估对羁押率的影响较为简单,主要是通过分析采用社会危险性量化评估后羁押率的变化来发现其影响;评估模型的预测正确率,则是通过被预测为具有高度社会危险性的人最终实施危害社会行为的比例来实现。我国司法实践中通常只关注逮捕社会危险性量化评估对羁押率的影响,对于模型预测正确率的评估则极少涉及。较之于对羁押率变化的分析,模型预测正确率的评估更为重要、也更为客观。因为虽然从理论上来说,在控制其他变量而仅改变社会危险性评估方法的情况下,羁押率变化等指标可表明量化评估的预测正确率较高,但实践中很难完全控制住其他自变量对羁押率变化的影响,如逮捕理念、逮捕程序、考核指标等因素很难完全保持不变。所以,美国、英国等在评价量化评估结果科学性时依据的是模型的预测正确率。如研究发现,美国 PSA、COMPAS 等系统的预测正确率差别不大,仅达到中等程度的预测正确率,如预测为高风险的被告人其再次犯罪的比例为 40% 左右;^④牛津大学新开发的再犯风险评估模型中,认定为高风险的犯罪嫌疑人、被告人在两年内再次实施暴力犯罪的比例达到 60%。^⑤ 故而,我国司法实践中对逮捕社会危险性量化评估结果的科学性评估是不完整的。

(二) 逮捕社会危险性量化评估的公正性风险

逮捕社会危险性量化评估不仅存在科学性风险,还存在着公正性风险。因为逮捕社会危险性量化评估不仅可能会强化办案人员对特定人员的刻板印象从而有违个别判断的要求,还可能会固化司法实践中既存的偏见歧视等因素,甚至会产生剥夺当事人辩护权的风险。

第一,逮捕社会危险性量化评估可能强化司法人员对特定群体的刻板印象,不利于社会危险性个别化判断的实现。逮捕社会危险性量化评估旨在限制司法人员的自由裁量权,而且作为人工智能的重要应用领域,逮捕社会危险性评估的终极目标则是完全取代司法官员的裁决,用计算机来模拟和取代人类的思考过程。^⑥ 但司法人员的自由裁量权并非一无是处,其对于确保社会危险性判断的个别化有着重要的意义。因为逮捕社会危险性评估是利用群组特征来审查特定个人的社会危险性,无法完全反映特定个人的社会危险性。如通过统计分析会发现,低学历、农村户口、无工作等与再次犯罪之间存在显著影响,^⑦提示此类人员具有更高的社会危险性。如果司法人员不再进行个别化审查而是直接依据此结果作出羁押决定,会持续强化司法人员对此类人员的高社会风险刻板印象,最终会形成司法人员对特定群体的偏见、歧视。实际上高风险只是代表一种可能性,并不意味着凡是符合低学历、农村户口或无工作情形的犯罪嫌疑人、被告人就会实施具有社会危险性的行为。而且,犯罪是个人生物因素、环境因素和心理因素等相互作用的结果,任何人都不是天生的罪犯。^⑧ 所以,运用逮捕社会危险性量化评估结果时,要防范司法人员对特定群体形成刻板印象的风险。

第二,逮捕社会危险性量化评估可能固化既有的偏见、歧视,有损逮捕社会危险性评估的公正性。偏见、歧视固定化是逮捕社会危险性评估面临的主要质疑之一,且随着人工智能自主学习能力的加强,这一问题会日益加剧。^⑨ 因为逮捕社会危险性评估主要是通过生效裁判文书中提炼变量并建构模型的,而司法官员在现实裁决中会受到偏见、歧视、意识形态以及性格特征等主观因素的影响,^⑩这

^④ Min Yang, Stephen C. P. Wong, Jeremy Coid, *The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools*, 136 *Psychological Bulletin* 740, 757 (2010).

^⑤ Christopher Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, 15 *Ohio State Journal of Criminal Law* 583, 584 (2018).

^⑥ 参见郑戈:《算法的法律与法律的算法》,载《中国法律评论》2018年第2期,第76页。

^⑦ 参见汪晓翔、刘仁文:《不同再犯类型的差异化风险因素研究》,载《中国人民公安大学学报(社会科学版)》2020年第5期,第8页。

^⑧ 参见狄小华:《智能化再犯风险评估司法应用的法律风险及其防范》,载《学术界》2020年第5期,第76页。

^⑨ 参见王燃:《大数据时代侦查模式的变革及其法律问题研究》,载《法制与社会发展》2018年第5期,第124页。

^⑩ 参见[美]理查德·波斯纳:《法官如何思考》,苏力译,北京大学出版社2009年版,第43页。

些主观因素也会反映到裁判结果中。虽然统计分析可以去除很多个性化的因素从而实现司法官员智慧的聚合效应,^{④7}但当这些主观性因素普遍存在时,统计分析则无法将这些主观性因素从算法中去除。而且,即便是直接将这些主观性因素从最终模型中剔除,也无法完全剔除主观性因素对模型算法的潜在影响。因为犯罪历史、曾被监禁情况、教育以及就业情况等,本身就可能体现了这种偏见或歧视。如果量化评估模型中将这些偏见、歧视固化下来,将会对特定群体造成更为不利的影响。如美国新闻调查性网站 ProPublica 在分析一种主流风险评估工具后认为,该评估工具将黑人被告人错误划分为“高风险”的概率要比白人被告人更高,将白人被告人错误划分为“低风险”的概率要比黑人被告人更高。^{④8}所以,逮捕社会危险性量化评估会将司法官员的主观因素固化,反而不利于社会危险性的公正判断。此外,偏见、歧视等主观因素也可能来自于开发量化评估工具的企业或个人。因为当前的逮捕社会危险性量化评估工具大都是由企业单独或企业与司法机关联合开发的,企业也可能会将特定价值、偏见等糅合到评估模型中以谋取私益。^{④9}

第三,逮捕社会危险性量化评估可能会产生剥夺当事人辩护权的风险。由于逮捕社会危险性量化评估模型及算法的非公开性,可能使得律师进行强制措施辩护变得更加困难,从而实质剥夺了当事人获得辩护的权利。其一,当前的量化评估模型更多地体现了公权力机关的意志,缺乏听取辩方意见的渠道。虽然当事人及其辩护人可在量化评估过程中提出具体意见,但由于逮捕社会危险性量化评估依据的是已经定型的评估模型,当事人及其辩护人提出意见更好的阶段是在模型建构过程中。当事人及其辩护人通过参与量化评估模型建构,从源头上防止公权力机关将量化评估模型变为其恣意的工具。但从当前司法实践来看,非但当事人及其辩护律师极少能参与到这一过程中来,其意见也很难反映到量化评估模型中去。如上海市高级人民法院研发 206 系统时,是从上海法院、检察院和公安系统中抽调 64 名业务骨干组建的研发法律团队。^{⑤0}其二,逮捕社会危险性量化评估原理的不公开,实质上剥夺了当事人的辩护权。基于保守商业秘密以及量化评估模型建构的复杂性等原因,辩护律师以及法官、检察官等具体办案人员通常并不清楚逮捕社会危险性量化评估模型的原理,只是在被动使用量化评估工具。如在卢米斯案中,威斯康星州最高法院大法官在协同意见书中写道,“本案中法院对 COMPAS 系统缺乏了解是一个关键问题,虽然在口头辩论中法院不断询问州和被告人的律师以了解 COMPAS 系统如何运行的,但极少有问题被恰当回答。”^{⑤1}辩护的前提是了解,如果律师对某一问题都不了解当然也就谈不上辩护。在当事人及其辩护律师不了解逮捕社会危险性量化评估原理的背景下,即便是量化评估结果存在问题,当事人及其辩护律师也难以提出有效的质疑。故而有学者认为当前量化评估这种不予公开、不接受质询、不提供解释、不进行救济的做法有演化为“算法霸权”的风险。^{⑤2}所以,逮捕社会危险性量化评估模型原理的不公开,可能会产生剥夺当事人辩护权的风险。

四、逮捕社会危险性量化评估的优化路径

虽然逮捕社会危险性量化评估存在一定科学性和公正性风险,但从司法实践来看,试点逮捕社会危险性量化评估的地方均出现羁押率下降的现象。如美国肯塔基州 2011 年立法要求法官在判断被告人社会风险时必须使用风险评估工具后,低风险和中风险被告人被无条件释放的比例分别较之前增加了 22% 和 16%。^{⑤3}我国司法实践中也有数据表明逮捕社会危险性量化评估对降低羁押率有影

^{④7} 参见马长山:《司法人工智能的重塑效应及其限度》,载《法学研究》2020年第4期,第29页。

^{④8} Julia Angwin et al., *Machine Bias*, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, 2020-9-4.

^{④9} 参见郑曦:《人工智能技术在司法裁判中的运用及规制》,载《中外法学》2020年第3期,第694页。

^{⑤0} 参见前引^{④9},第39页。

^{⑤1} 881 N. W. 2d 774 (Wis. 2016).

^{⑤2} 参见前引^{④7},第34页。

^{⑤3} 前引^{④3},第354页。

响,如广州市南沙区检察院适用社会危险性量化评估系统后,因无社会危险性不批准逮捕案件同比上升 17.24%。^④ 上述数据表明,逮捕社会危险性量化评估与采用非羁押性强制措施之间可能存在一定的因果关系。在我国羁押率居高不下的背景下,我们应扬长避短,规制逮捕社会危险性量化评估的风险并实现其优化适用。

(一) 建构科学公正的逮捕社会危险性量化评估模型

逮捕社会危险性量化评估主要通过评估模型来实现,故而建构科学公正的量化评估模型是优化量化评估的关键所在。如前所述,既然评估模型的准确率与人类对犯罪认识水平、评估模型指标体系的科学化、原始数据质量以及评估结果的科学性评估等因素有关,那么优化量化评估模型亦可从如下几方面展开。

第一,通过社会危险性理论与实践经验分析以及两者间的检验与校正,不断优化逮捕社会危险性量化评估模型。评估某人的社会危险性并非一件容易的事情,虽然当前社会风险评估模型的预测正确率已经有了很大提升,但距离社会期望的预测正确率仍存在较大差距,这需要进一步完善模型建构方法。如前所述,单纯使用定性或定量研究方法来建构社会危险性模型都可能面临一定的困难,所以当前主流的社会危险性量化评估模型均采用定性与定量相结合的方法来建构。其一,充分运用社会学习理论发掘新的可能影响社会危险性的因素。当前逮捕社会危险性量化评估的主要理论基础是社会学习理论,但受原始数据来源不足以及数据提取技术不完善等现实因素的影响,当前的量化评估模型中并未将所有可能影响社会学习进程的因素都包含进去。但理论研究可不去考虑现实因素制约的问题,通过对社会学习理论的研究与分析,我们可进一步挖掘可能影响发生社会危险性的新的要素,从而在学理上实现量化评估模型的优化。其二,充分提炼逮捕社会危险性评估的实践经验。与美国将社会危险性条件的认定完全交由司法人员自由裁量不同,我国立法上对逮捕社会危险性有着较为细致的规定,但这并不意味着司法实践中就没有自由裁量的空间。司法人员在认定社会危险性时也会用到大量的经验法则以及利益考量,如立法规定的社会危险性如何进一步细化、各情形分别代表何种程度的社会危险性、出现不同情形时如何综合认定社会危险性以及司法人员的个人利益等,这些经验和利益考量恰恰是建构社会危险性量化评估模型所必需的。所以,应充分提炼逮捕社会危险性评估中的实践经验。其三,强化逮捕社会危险性量化评估理论模型与实践经验的相互验证,从而实现量化评估模型的不断优化。理论分析与实践经验总结并非非此即彼的关系,二者完全可以实现相互验证与相互优化。通过相互验证,一方面可修正理论分析中的漏洞,另一反面也可检验实践经验是否具备学理上的正当性基础。^⑤

第二,优化逮捕社会危险性量化评估模型的指标体系。优化逮捕社会危险性量化评估模型指标体系,可从强化指标的客观性、重视指标的地区差异性以及去除不正当指标三方面来实现。其一,强化指标的客观性与精细化。如前所述,我国的逮捕社会危险性量化评估模型中将不少量刑情节直接作为具体指标,使得指标体系的客观性不足。对此可对相关指标做技术化处理,将其还原为客观行为的指标。如将自首拆解为主动投案和如实供述自己所犯罪行两个指标,将坦白替换为如实供述自己所犯罪行,将正当防卫拆解为正当防卫的具体构成要件等。再如,对于累犯、前科以及曾被行政处罚等指标可限定发生的具体时间,如限定为两年之内曾发生特定类型犯罪、一年内曾因特定行为而被行政处罚等。其二,关注指标的地区差异性。由于数据来源的地域差异、文化差异、司法人员使用评估工具的熟练程度等原因,逮捕社会危险性量化评估的指标也存在一定的地区差异。如美国有实证研

^④ 参见前引⑩。

^⑤ 参见孙道萃:《人工智能辅助精准预测量刑的中国境遇——以认罪认罚案件为适用场域》,载《暨南学报(哲学社会科学版)》2020年第12期,第76页。

究发现,美国许多常见的暴力犯罪再犯风险评估模型对欧美白人比少数族裔有更高的适用性;^⑤我国亦有实证研究表明,不同地区在受贿罪量刑结果上存在差异,且部分地区之间的量刑差异非常显著。^⑥考虑到我国刑事司法实践以及地区差异的现实情况,可以考虑以省级行政区划为单位分别开发相应的社会危险性评估模型,从而实现关注地区差异与确保法律统一适用的有效平衡。其三,限制或去除模型中的不正当指标。当前建立逮捕社会危险性量化评估模型时采用定量与定性研究相结合的方法,其中很重要一个原因就是要通过定性分析来去除模型中可能包含的不正当指标,如偏见、歧视等。所以,在建立逮捕社会危险性量化评估模型之后应当对模型进行正当性审查,仔细甄别各指标中可能包含的司法人员主观因素。对于那些为法律所禁止的主观因素则必须在评估模型中予以去除,对于法律尚未禁止但有可能造成对特定人群歧视的主观因素也应当适当限制其在评估模型中的作用。此外,还可通过将人权保障、正当程序等理念加入评估模型,来稀释偏见、歧视对模型的影响。

第三,确保量化评估原始数据的质量。确保原始数据的质量,可从统计分析方法的理性选择、原始样本的代表性以及信息抽取技术的完善等方面实现。其一,理性选择统计分析方法。先前的逮捕社会危险性量化评估模型所依据的样本主要是通过抽样来获取的,^⑦近年来司法实践中特别重视通过大数据技术进行全样本分析来建构量化评估模型。但从统计学角度来说,全样本分析并非必然优于抽样分析,二者各有利弊。如抽样分析中结果的全面性不足,而全样本分析中结果的精确性则有所欠缺。^⑧而且虽然当前很多逮捕社会危险性量化评估模型均宣称使用大数据,但其所依赖的样本量却并不大,与传统的抽样分析样本容量差别不大。^⑨所以,在建立量化评估模型时可根据情况选择抽样分析或是全样本分析。其二,确保原始样本的代表性。数据信息的代表性即建构评估模型所依赖的样本要具有代表性,能反映司法裁判的一般情况。通过抽样选取原始样本时,要确保抽样的随机性、代表性,否则会影响到结论的准确性。全样本分析虽然可以不用考虑样本的随机性,但也需要考虑样本的代表性问题。如对于某些在原始样本中极少存在但对社会危险性影响较大的指标,为确保模型的准确性,可通过增加特定样本以将此类指标纳入到模型中来。其三,完善信息抽取技术,提升原始数据的准确性。一个完整的信息(知识)抽取过程包括法律数据获取、法律数据标注、法律数据训练和法律模型生成四个环节。^⑩无论是大数据分析还是小样本的抽样分析,都涉及到从原始样本中抽取数据信息的问题。但客观来说,当前的信息抽取技术还难以满足量化评估模型对抽取信息准确性和全面性的要求。所以,为确保抽取信息抽取的准确性与全面性,一方面应继续完善信息提取技术,尽可能提升人工语言与机器语言转化的正确率;另一方面也应坚持机器抽取与人工核验相结合的方式,由人工查验机器抽取信息的准确性与全面性。此外,逮捕社会危险性量化评估模型建立之后并非一劳永逸,之后也应当根据司法实践的需求不断更新原始数据从而实现评估模型的更新换代。

(二) 逮捕社会危险性量化评估中的正当程序保障

逮捕社会危险性量化评估作为刑事诉讼程序的一部分,也应受到正当程序原则的规制。

第一,适度公开逮捕社会危险性量化评估的算法,建立对量化评估模型科学性与公正性的司法审

^⑤ Stefanie Schmid, Roxanne Heffernan, Tony Ward, *Why We Cannot Explain Cross - Culture Differences in Risk Assessment*, 50 *Aggression and Violent Behavior* 1, 1 (2020).

^⑥ 王剑波:《我国受贿罪量刑地区差异问题实证研究》,载《中国法学》2016年第4期,第252页。

^⑦ 如美国佛罗里达州审前风险评估模型的原始数据来源于1757个案例,而俄亥俄州审前风险评估工具依赖的是1800多个案例,弗吉尼亚州风险评估模型的原始数据是1971个案例。John Logan Koepke, David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 *Washington Law Review* 1725, 1758 (2018).

^⑧ 参见陈刚:《解释与规制:程序法定主义下的大数据侦查》,载《法学杂志》2020年第12期,第8页。

^⑨ 如上海206系统虽然宣称依据的是30万起刑事案件,但具体到特定子系统样本量则少得多,如命案模型中依赖的是591起刑事案件,盗窃罪模型中依赖的是36779件盗窃案件。参见前引^⑦。

^⑩ 参见王禄生:《论法律大数据“领域理论”的构建》,载《中国法学》2020年第2期,第270页。

查机制。逮捕社会危险性量化评估的核心是算法,为确保量化评估的正当性,量化评估的算法应当接受司法的公正性审查。其一,建立对逮捕社会危险性量化评估科学性与公正性的司法审查机制。虽然逮捕社会危险性量化评估模型的开发大都由司法机关主导或参与,但量化评估模型并不因司法机关的参与而自动具有法律上的正当性。逮捕社会危险性评估关涉犯罪嫌疑人、被告人人身权利的限制与剥夺,故而依据司法审查原则的要求,必须经过司法机关依法定程序审查之后才可在刑事诉讼中使用。司法系统对逮捕社会危险性量化评估模型的审查应包括科学性与公正性两方面的内容,科学性审查主要是确保量化评估的预测正确率,而公正性审查则是尽力防范量化评估受到偏见、歧视的不当影响。其二,适度公开逮捕社会危险性量化评估的核心算法。审查的前提是公开,公开也是消除“算法黑箱”“算法歧视”以及“算法霸权”的关键所在。当然,算法公开并非公开所有内容。通常来说,当算法满足如下三个要求时就算有效公开,即政府就算法程序以及后续的实施和检验生成适当的记录,承包商向公共机构就如何开发算法披露足够的信息,以及公关机构和法院将商业秘密视为公开披露的有限例外。^② 依此规则,逮捕社会危险性量化评估算法的公开,至少应当包括公开原始数据的来源、自变量的选取规则与权重设计、建构模型的方法、评估工具的预测正确率以及评估工具的限制等内容。而且,考虑到不同量化评估工具使用的概念术语不完全相同,公开量化评估算法还应当公开概念、术语的具体内涵与外延。

第二,逮捕社会危险性审查量化评估不能完全替代司法人员的判断,量化评估结果是司法官员适用非羁押性强制措施的重要依据。其一,逮捕社会危险性量化评估不能完全替代司法官员的判断,司法官员享有最终的决定权。虽然逮捕社会危险性量化评估旨在限制司法官员裁决羁押时的自由裁量权,但却不能完全剥夺司法官员的自由裁量权。这一方面是因为量化评估无法穷尽所有影响社会危险性的因素,可能会忽视特定犯罪嫌疑人、被告人的具体情况;另一方面也因为量化评估本身只是一种概率,只表明犯罪嫌疑人、被告人实施社会危险性行为的可能性而已。所以,逮捕社会危险性量化评估结果不能完全代替司法官员的判决,司法官员对羁押与否享有最终的决定权,否则容易陷入“机械司法”中。如在卢米斯案中,威斯康星州最高法院在判决中强调了个性化量刑的重要性,并认为由于法院未将 COMPAS 系统的评估报告作为判决的唯一依据,所以 COMPAS 系统评估满足量刑个性化的要求。^③ 其二,逮捕社会危险性量化评估结果是司法人员适用非羁押性强制措施的重要依据。虽然逮捕社会危险性量化评估不能完全取代司法人员的判断,但其仍应对司法人员自由裁量权的行使发挥一定的限制作用。由于逮捕社会危险性量化评估结果表现为高、中、低三种风险等级,故对量化评估结果的使用可区分不同风险等级而有所区分。首先,当逮捕社会危险性量化评估结果为低风险或中风险时,司法官员原则上应采取非羁押性强制措施。此种设计与逮捕社会危险性量化评估制度的目的有关。对逮捕社会危险性进行量化评估的原因是要降低审前羁押率,故而其主要目的是筛选出具有较低社会危险性的犯罪嫌疑人、被告人并对其适用非羁押性强制措施。所以,原则上应当要求司法官员对评估为中风险或低风险的犯罪嫌疑人、被告人采用非羁押性强制措施。其次,当量化评估结果为高风险时,司法人员不得仅凭此结果对犯罪嫌疑人、被告人羁押而仍应综合考虑后裁决是否羁押。量化评估本质上是一种概率,高风险并不意味着其未来必然会实施具有社会危险性的行为。^④

^② Robert Brauneis, Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 *Yale Journal of Law & Technology* 103, 104 (2018).

^③ 参见前引^②。

^④ 可能会有观点对此提出质疑,认为高风险的犯罪嫌疑人、被告人未来并不必然会实施具有社会危险性的行为,那么低风险犯罪嫌疑人、被告人未来也并不必然不会实施具有社会危险性的行为,为何此处只限制高风险评估结果的使用而不限制低风险评估结果的使用。这里还需考虑的是羁押例外原则。依据羁押例外原则,羁押候审本身就应是一种例外。强化低风险评估结果的使用以及限制高风险评估结果的使用均有利于羁押例外原则的实现,故而对量化评估结果的使用作出如上设计。

而且,高风险内部也可划分为不同的风险等级,将所有高风险犯罪嫌疑人、被告人全部羁押也违背比例原则的要求。所以,美国肯塔基州立法只要求将所有中低风险的犯罪嫌疑人、被告人予以快速释放,但并未对高风险的犯罪嫌疑人、被告人作出统一规定,而是由司法官员根据案件情况综合判断。德国司法实践在判断“逃亡危险”时,明确要求不能将犯罪学上的经验用于对被指控人作不利的评价,这种评价必须是来自于个案的具体事实,而且要将预示逃亡情况与阻碍逃亡情况相权衡。^⑤故而,不能仅凭逮捕社会危险性量化评估结果为高风险就将犯罪嫌疑人、被告人羁押。

第三,充分保障犯罪嫌疑人、被告人在量化评估过程中参与权与辩护权的实现。其一,辩方有权参与到社会危险性量化评估模型建立过程中并提出意见。基于辩护权的绝对性与广泛性,当运用逮捕社会危险性量化评估结果作羁押决定时,辩方的辩护既可针对羁押决定也可针对量化评估模型本身。但逮捕社会危险性量化评估模型作为一种模型化的方法一旦建立并适用后,辩方就模型本身提出的意见是很难获得认同的,这无疑限制了辩护权的实现。所以,考虑到量化评估的现实情况,应当允许辩方介入到量化评估模型建立过程中,并就模型本身的科学性与公正性提出意见。其二,强化逮捕社会危险性量化评估结果的说理机制。逮捕社会危险性量化评估将社会危险性细化为一个个具体指标,司法人员收集这些指标的具体数值并将其输入量化评估模型从而得出评估结果。而这一过程通常是以数学等格式化方式进行的,这极大增加了犯罪嫌疑人、被告人对量化评估过程和结果的理解难度,阻碍着犯罪嫌疑人、被告人辩护权的实现。所以,当司法人员运用逮捕社会危险性量化评估结果时,特别是对量化评估结果为高风险的犯罪嫌疑人、被告人,应当强化司法人员的说理责任。司法人员不仅应向其说明量化评估原理性的内容,还应当具体说明量化评估在本案中的具体适用,如相关指标的具体赋值、赋值时的考虑因素以及评估结果的可能适用范围等。只有允许犯罪嫌疑人、被告人及其辩护律师充分了解量化评估在本案中的具体适用,才能使其有针对性地进行辩护,避免辩护流于形式。

第四,强化对逮捕社会危险性量化评估模型中涉及个人信息的保护。逮捕社会危险性量化评估大量使用犯罪嫌疑人、被告人的身份、家庭、职业等个人信息,对这些信息的搜集与使用就涉及到个人信息保护的问题。虽然当前社会危险性评估模型主要使用的是生效裁判文书或逮捕决定书等,社会危险性评估收集的犯罪嫌疑人、被告人信息并未超越法律许可的范围。但如果想进一步提升评估的预测准确率,就需要掌握犯罪嫌疑人、被告人更多的个人以及社会环境等信息。如依据《未成年人刑事检察工作指引(试行)》第36条规定,未成年人案件社会调查报告中不仅包含与犯罪有关的情况,还包括个人兴趣爱好、生活习惯、家庭经济状况、家庭疾病史以及交往对象等属于个人隐私的信息。而且,随着大数据以及日常监控技术的大量使用,犯罪嫌疑人、被告人的日常生活、工作以及社交活动中的诸多信息也可能被用来评估其社会危险性。所以,未来的逮捕社会危险性评估可能会使用更多以及更加隐私的犯罪嫌疑人、被告人个人信息,这可能会侵犯到犯罪嫌疑人、被告人的隐私权领域。对此,我们有必要重申比例原则在逮捕社会危险性评估中对个人信息权保护的必要性。只有那些为了进行逮捕社会危险性评估所必须的个人信息,才能为公安司法机关所收集并作为评估的依据;且收集到的个人信息只能用作对犯罪的侦查、起诉和审判而不得用作其他目的,如用作个人目的、政治目的、违背法律精神的社会管控等。^⑥此外,当个人信息使用完毕后,犯罪嫌疑人、被告人有要求公安司法机关将相关信息删除或不再使用相关信息的权利。^⑦

^⑤ 参见《德国刑事诉讼法典》,宗玉琨译注,知识产权出版社2013年版,第99页。

^⑥ 参见程雷:《大数据的法律侦查》,载《中国社会科学》2018年第11期,第173页。

^⑦ 参见郑曦:《“被遗忘”的权利:刑事司法视野下被遗忘权的适用》,载《学习与探索》2016年第4期,第64—65页。

结 论

随着我国逮捕制度的进一步改革,如何进一步降低羁押率也为社会各界广泛关注。在这一背景之下,结合当下人工智能在司法领域中的应用,逮捕社会危险性量化评估被引入我国刑事司法实践中。但总体来看,当前司法实践中对逮捕社会危险性量化评估模型的应用仍处于起步阶段,量化评估模型中的指标体系不尽完善,司法人员对使用量化评估模型缺乏足够的信任,人工智能技术的不当应用则会进一步放大上述风险。所以,实现逮捕社会危险性量化评估的科学化与正当化是未来推广适用逮捕社会危险性量化评估的关键所在。为实现逮捕社会危险性量化评估的科学化,我们应在不断提升对犯罪认识水平的基础上完善模型建构,确保评估模型所依赖原始数据的准确性,不断完善评估模型的指标体系。而逮捕社会危险性量化评估模型的正当化则要求其要受到正当程序的约束,建立并完善量化评估算法和评估结果的司法审查机制,充分保障辩护方在量化评估过程中的参与权,同时实现对个人信息权利的保障。

A Quantitative Assessment on the Social Dangerousness in Arrest —From the Perspective of Automated Decision – making and Regulatory System

GAO Tong

Abstract: Quantitative assessment on the social dangerousness of arrest is a method to predict the social dangerousness of criminal suspects and defendants by formal, quantitative and actuarial assessment, which has developed rapidly in the world since the 1980s. The quantitative assessment of social dangerousness of arrest is based on legal realism and evidence – based practice, with the aim of reducing the pretrial detention rate and is supported by the rapid development of science and technology. In practice, the assessment has a moderate predictive accuracy which is not more accurate than the judges' prediction. Although the assessment helps the judicial system in bring convenience, there are still some scientific and judicial risks. For example, the assessment will increase the risk of prejudice or discrimination, and the risk of depriving the defendant of the right to defense. In order to make full use of the assessment, we should build more scientific and fair assessment models, and make sure that the principle of due process should be applied in the construction and implementation of the quantitative assessment on the social dangerousness of arrest.

Key words: arrest social dangerousness automated decisions regulatory system