理论前沿

网络刑事电子数据算法取证难题及其破解*

何邦武

内容提要:通过算法取证已经成为网络时代搜集网络电子数据的唯一选择。由于网络电子数据存储和呈现方式的特殊性,算法取证面临着关联性如何界定、可靠性如何保证及合法性如何守护的难题。与此同时,算法取证还遇到传统刑事证明理念及证明模式所形成逆向效应的阻遏。解决上述难题的可行性路径是,本诸法教义学的原理,推求证据关联性及可采性规则的应然法理,重新解读关联性规则中的经验法则,引入科技证据的可采性标准,参酌欧盟《一般数据保护条例》中数据权的规定,确立算法取证中合法性的基本原则和相应规则。算法取证难题的真正突破,还应摒弃印证的理念和惯习,回归现代自由心证,承认并践行案件事实调查的法律论证属性,重新审视法律论证中语言的角色和功能。

关键词:网络犯罪惩治　网络电子数据　算法取证　刑事证明模式

何邦武,南京审计大学法学院教授。

一　引　言

信息时代大数据的指数级增长所掀起的阵阵数据巨浪,正在“吞噬”着整个人类并改变其生活方式,促使其蜕变成一种“数字化”生存。这无疑是继工业文明之后人类文明的又一次大变局、大转型,以致由此产生人工智能将战胜并取代人脑的恐惧。〔1〕 与之相应,

* 本文的写作是与阿里集团安全部总监连斌先生多次晤谈的结果,有些观点即来自连斌先生,谨致谢忱。

〔1〕 关于人工智能到底能否战胜人脑问题的争论,已经持续了很长时间。自从 2016 年开始 AlphaGo 先后在几次围棋赛中战胜了人类之后,未来人工智能一定能战胜人脑的言论就甚嚣尘上,加之霍金关于人工智能战胜人脑预言的影响,人工智能将战胜人脑似已成为不易之论,甚至有人主张赋予机器人法律主体资格。笔者认为,上述问题的求解路径仍应回到人工智能工作的基本原理上来,即人工智能所使用的语言实际上是一种以 0 和 1 这两个数码为单位的二进制“算法”,与人脑在思维时融合了感觉、知觉、情感等理性和非理性因素而形成的包含多重复杂因素而形成的自然语言有着本质的不同,后者因其自身蕴含的丰富性无法被简单还原为二进制的计算机语言。因此,人工智能只能取代人类的某些认知活动,但无法像人类那样基于道德预设或目的预设作出判断并继之作出相应的选择,也不会出现行为的非理性(机器人行为的选择是计算的结果)。因此,担心人工智能将取代人脑实际上是多余的,也有学者从法律主体资格的概念内涵出发对之进行了否定性论证。参见吴习彧:《论人工智能的法律主体资格》,《浙江社会科学》2018 年第 6 期。

网络空间治理的现实需求，迫切需要法律理论与实务部门为其提供相应的制度供给，或者在规则罅隙或模糊之处提供可资争端解决的法理指引。其中，作为大数据载体之一的网络电子数据如何收集取证以取得证据资格，并随之得到法庭认证，是一切网络法律争端中最终无法回避的问题，成为网络治理各类法律问题实质上的核心和研究的热点。

在有关网络电子数据收集取证问题的研究方面，作为对此类证据迅捷发展现实的回应，既有研究成果在以下领域作了很多开掘性的研究：例如对网络电子数据的属性及其与传统意义上的电子数据的区别，网络电子数据收集取证，鉴定及保全的方式，电子数据真实性审查、区块链技术在证据保管中的运用等，其学术创新价值及实践指导意义值得肯定。〔2〕

然而，既有研究对如何从海量大数据中筛选相关的网络电子数据，即算法取证，以及其中所遵照的收集理念与传统的处于“信息孤岛”状态下的电子数据的收集理念又有哪些差异等问题，并无涉及。这不仅使网络电子数据算法取证在其研究的起点上陷入理论盲区，也使后续的关于网络电子数据鉴定、审查认证等的研究缺乏相应的基础，甚至蹈入某种纯粹理论演绎的虚空，与实际的网络电子数据收集、取证、认证所依凭的应然规则与法理凿枘不投。同时，出于对人工智能的非理性崇尚乃至神化，一些研究者抱持将网络电子数据“唯科学化”的观点，试图探索建构统一的电子数据采信标准。〔3〕此外，不同于应然的证据资格理论，我国以证据“三性”为特征的“本土化”言说所构建的证据资格理论，〔4〕导致了我国现有网络电子数据资格理论的变调，而与应然的证据资格法理不谐和（笔者将在下文中详述）。此外，印证证明模式和“客观真实”的证明标准所产生的倒逼效应，也对网络电子数据的收集取证形成掣肘。所有这些研究现状都说明，有关网络电子数据收集取证的研究，尚缺乏作为研究出发点的基础性共识，因此，分析网络电子数据取证中存在的诸多制度和观念性障碍，澄清有关网络电子数据收集取证的基本理念和应然的证据资格法理，重建网络电子数据取证的基础性共识，已形迫切且深有必要，唯此方能推进网络电子数据取证的研究。

需要说明的是，本文特别以“网络电子数据”作为专有名词和研究对象，是因为如前文所述，以大数据、物联网、人工智能、5G 为核心特征的大数据取证中的对象，存在着与传

〔2〕 涉及上述领域的电子数据论述参见刘品新：《论大数据证据》，《环球法律评论》2019 年第 1 期；刘品新：《电子证据的关联性》，《法商研究》2016 年第 6 期；刘品新、唐超琰：《互联网金融犯罪案件证据海量问题及应对》，《人民检察》2018 年第 20 期；龙宗智：《寻求有效取证与保证权利的平衡：评“两高一部”电子数据证据规定》，《法学》2016 年第 11 期；陈永生：《证据保管链制度研究》，《法学研究》2014 年第 5 期；刘译矾：《电子数据的双重鉴真》，《当代法学》2018 年第 3 期；胡铭：《电子数据在刑事证据体系中的定位与审查判断规则：基于网络假货犯罪案件裁判文书的分析》，《法学研究》2019 年第 2 期；谢登科：《电子数据的取证主体：合法性与合技术性之间》，《环球法律评论》2018 年第 1 期；张庆立：《区块链应用的不法风险与刑事法应对》，《东方法学》2019 年第 3 期。其他相关论述笔者将在下文中介绍。

〔3〕 参见刘品新、陈丽：《数据化的统一证据标准》，《国家检察官学院学报》2019 年第 2 期。刘品新：《印证与概率：电子证据的客观化采信》，《环球法律评论》2017 年第 4 期。笔者将在下文中对此作进一步的分析。

〔4〕 沈德咏先生在论及 1979 年以后我国证据法学的研究状况时，曾经指出，证据法学研究中的许多思想和观念还是在资料匮乏、视角单一的历史条件下“独创”出来的，而且，这些思想和概念还成为我们无法抛弃的认识前提和知识基础。参见沈德咏著：《刑事证据制度与理论·总序一》，法律出版社 2002 年版。

统的处于“信息孤岛”状态下的电子数据在表现形式、存储方式等方面的根本性区别，以致其取证理念、取证方式等亦有较大的差异。尽管后者如单位自身的数据库现在体量一般也很大，但仍然是独立的存在方式。因此之故，区分两类不同电子数据的研究势所必然，也理应成为研究者的理论自觉。

二　关联性确立：网络电子数据算法取证的核心难题

利用网络犯罪的手段多种多样，和传统的物理空间犯罪一样，采用各种隐蔽手法逃避监管是犯罪嫌疑人常有的思维。笔者在调研过程中，从某网络平台了解到的信息即具有代表性。该网络平台安全部负责人举出了以下虚拟的案件：某犯罪嫌疑人涉嫌非法制造枪支的犯罪。其犯罪手法是利用网络购买各类枪支配件，然后进行组装。为了逃避侦查，该嫌疑人同时在网上注册了多个账号用于交易，并将收货地点写成不同的地方。但是，网络平台通过对海量数据进行清洗、检索和逐层筛选，终于将购买主体，收货主体等锁定为同一嫌疑人。网络平台将犯罪线索提交给公安部门后，后者在嫌疑人住所找到了组装的枪支，其中的各类配件与网络上交易的配件一一对应。从正常的侦查逻辑判断，这样的案件从线上的交易记录到线下的物证、快递公司交货的证人证言等，已经构成完整的证据链，可以证明嫌疑人的犯罪行为。但由于数据筛选过程的复杂，涉及多种数据模型和算法，警方对此感到疑虑：网络平台的数据筛选过程中算法的可信性程度如何，凭什么就认定多种交易主体即为犯罪嫌疑人一人？这样的模型算法得出的关联性可靠吗？加上犯罪嫌疑人的否认性供述以及其枪支是从外面捡来的辩解，最终的结果可能是警方担心证据不足不予立案。

另外，据该负责人介绍，他在叙说这样的案件时，已经作了很多简化，而实际过程由于数据繁杂，数据检索缺乏明确的目的性，其筛选过程事实上非常复杂。由于数据建模和算法等涉及平台自身很多的技术信息，同时也是为了防止犯罪嫌疑人在了解了这些模型后衍生出更加复杂的反侦查技术手段，因此，数据清洗和筛选过程又不能公开。更为棘手的是，这样的数据建模及算法会随着网络大数据的发展而不断翻新，其纯技术性的脸孔更加难以为计算机专业技术人员以外的人员所知晓。〔5〕 因为网络发展所带来的知识增量与科技证据一样，已经逾越常人的认知能力。这一结果客观上使犯罪侦查陷入僵局，也是所有网络电子数据取证面临的真正难题，即，数据建模与算法本身的可信性从何而来，能否如科技证据一样，通过确立一个算法标准，而使检索数据在合法性的基础上，因具有可信性而可以采纳为证据？

上述案例争议的核心就是网络电子数据取证的关联性和数据筛选中算法的可信性问

〔5〕 该负责人讲到的这种情形，可比拟的是李昌钰先生物证鉴识科学中的化学制剂显性方法。在李氏侦破的美国康涅迪克州“锯木机杀妻案”中，罪犯即被害人丈夫理查德·克拉夫兹购买的一台被丢弃在水中的油锯被打捞上来后，李昌钰先生就是用显性法将油锯上被罪犯磨去的号码显示出来（存留时间非常短），以此固定了该油锯为罪犯购买的事实。这里的疑问同样是，这种化学制剂显性法可信性的依据何在？

题。在上述二者之间,又存在着互相依附的关系:关联性是数据建模和算法的出发点,而数据建模和算法的可信性又是关联性得以保障的基础。这里首先分析关联性问题,因算法的可信性涉及证据能力问题,拟于文章下一部分分析。

在以 5V 为特征的大数据时代,如何从海量电子数据中,根据相关性原理筛选用以证明某项事实的数据,是网络电子数据搜集的首要问题。如果是主动型侦查,这一问题将更加突出。[6] 虽然与传统的物理空间中犯罪证据的收集以关联性为搜集取证的逻辑起点一致,但置身网络大数据的语境中,这里的关联性具有不同的内涵与外延。在传统的证据法律理论中,关联性是指有助于证明有关假设的属性,这种假设一旦成立,将从逻辑上影响争议事项。关联性描述了向法庭提交的证据与某一案件中的关键命题或可证明的命题之间的逻辑关系,是证据与待证事实之间的逻辑关系。[7] 美国《联邦证据规则》第 401 条关于关联性的定义为,某证据"具有使任何对于决定诉讼结果的事实的存在比没有该证据时更有可能或更无可能的趋势"。尽管这样的描述性界定(不是属加种差的所谓定义概念的本质主义方法)存在很多模糊之处,但以下的意思是清楚的:关联性是一种基于证据收集、采用主体的经验性判断,是对命题事实之间真值关系的判断,其逻辑基础是一种因果关系。其中,作为"因"的命题事实与"果"的命题事实范围是明确的、特定的。如在一起盗窃案件中,被盗窃的财物、盗窃现场嫌疑人留下的痕迹等是侦查人员进一步搜集其他证据(盗窃物的去向、犯罪嫌疑人身份等)的"因",以此展开的证据搜集都是以关联性为基础,搜集的目的是明确的、对象是可知的,取证中倚赖的是有经验的侦查人员。

然而,网络电子数据的搜集,其对象关联性的确定是以算法模型为基础的自动捕捉,或者是人工智能的自动识别,而非自然人的经验感知,所依据的原理是数据算法中的模糊理论而非因果关系理论,遵循的是一种数理逻辑。在数据爬取阶段,需要通过 ETL 工具进行处理,其过程包括从数据来源端经过抽取(extract)、清洗(cleaning)、转换(transform)、加载(load)到目的端 4 个环节。具体来说,这一过程首先是从数据源抽取出所需的数据,通过网络爬虫和一些网站平台提供的公共应用程序接口(Application Programming Interface,API)等方式从网站上获取非结构化或半结构化的网页数据,然后将其提取、清洗、转换成结构化的数据,统一存储在数据库中。[8] 其后还要通过数据建模,依照算法对数据仓中的数据进行分析和处理,找出关联性电子数据。这一过程可图解如图 1。

---

〔6〕 大数据的 5V 特点系 IBM 公司提出的,即 Volume(大量)、Velocity(高速)、Variety(多样)、Value(低价值密度)、Veracity(真实性)。其核心特点是因摩尔定律所致的数据存储总量的急遽扩张。摩尔定律是 1965 年英特尔的创始人之一戈登·摩尔在考察了计算机硬件的发展规律之后提出的。该定律认为,同一面积芯片上可容纳的晶体管数量,一到两年将增加一倍。摩尔定律为大数据时代的到来铺平了道路。有关摩尔定律的详细知识参见涂子沛著:《数据之巅:大数据革命,历史、现实与未来》,中信出版社 2014 年版,第 259 页以下。

〔7〕 Black's Law Dictionary (5th ed), p. 1160.

〔8〕 大数据采集系统包括系统日志采集、浏览器数据采集和数据库采集 3 种,网络电子数据采集是其中的一种。三种系统数据采集所倚赖的算法模型均不相同,详细论述参见《大数据时代必须要了解的数据系统知识》,http://www.sohu.com/a/258229051_100123000,最近访问时间[2019-04-18]。

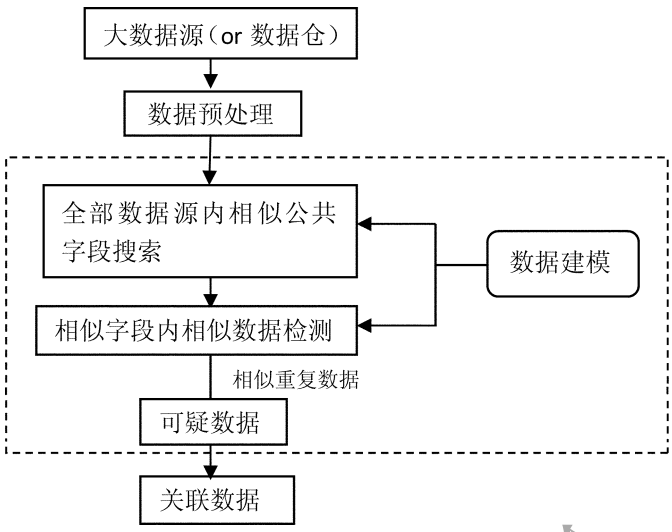

图 1　数据处理过程

由此可以看出，无论是前期的数据爬取，[9] 还是后期对数据仓中的数据通过建构模型的分析和处理，关联性电子数据的获得，都是计算机系统通过一定的数据模型自动识别的结果，是科学性而非经验性结论。易言之，网络电子数据算法取证的规则与法理已经逾越了传统的基于物理空间证据收集的关联性规则与理论。这是网络电子数据取证实践对传统证据法理与规则的第一重挑战。

三　证据能力确认：网络电子数据算法取证的又一难题

证据能力即证据具备可采性的资格，包括可信性（credibility）与合法性（legitimacy）两方面内容。[10] 网络电子数据的证据能力（可采性）同样受制于既有的可信性与合法性规则。但由于网络电子数据由算法收集形成，其形成过程同样对证据的可信性与合法性构成新的挑战。

需要申明的是，规范意义上的证据资格或证据能力（competency of evidence）理论，又叫证据的可采性（admisibility of evidence），是诉讼当事人或其他有关人员提交的证据符合法律规定的采用或采纳标准。但是在中国证据法学语境中，对证据资格的判断直接表述为符合证据客观性（真实性）、合法性和关联性等三性即证据内在属性的要求。这实际上是一种关于证据概念属性在外延上的错置，即将“什么是证据”（证据概念）和“什么证据

〔9〕 从技术上看，目前常用的网页爬虫系统有 Apache Nutch、Crawler4j、Scrapy 等程序系统。其中，Apache Nutch 是一个高度可扩展、可伸缩性的分布式爬虫框架。Apache 则通过分布式抓取网页数据，由 Hadoop 支持，通过提交 MapReduce 任务来抓取网页数据，还可以将网页数据存储在 HDFS 分布式文件系统中。Nutch 可以进行分布式多任务进行爬取数据、存储和索引。Crawler4j、Scrapy 框架则大大降低了开发人员开发速率，开发人员可以很快地完成一个爬虫系统的开发。参见 http://www.sohu.com/a/258229051_100123000，最近访问时间[2019－04－18]。

〔10〕 证据资格（可采性）是对具备关联性的证据进行的合法性、可信性审查，即关联性是证据资格的前置属性，属于经验与逻辑问题，而可采性是法律问题，是一个反面、消极、纯粹的法律性概念，其适用目的旨在排除有关联性的证据。参见郭志媛著：《刑事证据可采性研究》，中国人民公安大学出版社 2004 年版，第 20 页以下。

可以被采用”(证据资格)混同。[11] 并且,证据三性的界定中还存在一种泛哲学化的思维,因为,即使是伪证,换一种视角也具有真实性(客观性),从而使证据的真实性(客观性)成为一种诡辩式断言和文字游戏。

质言之,我国目前主流证据法理论关于证据真实性或客观性的概念,是哲学理论中物质的客观性原理在证据属性中的运用,是一种特定时期泛哲学化思维的产物,不仅意义难以界定(如伪证也具有客观性、真实性)且无实践价值,甚至使人在实践中产生究竟什么是客观性的疑思。然而,如沈德咏先生所言,1979 年以后中国大陆证据法学研究中的许多思想和观念实际上是在资料匮乏、视角单一的历史条件下“独创”出来的,[12] 阅读者可以想象并还原大陆主流证据知识理论最初形成时的场域:受居支配地位的泛哲学化思维的影响,在没有其他证据理论知识可资借鉴的情境下,构建现有的主流证据知识必然依靠当时君临一切的哲学本体论思维,对证据概念的界定及其属性依照哲学理论中关于物质本质的论述进行想象化的构建。尽管这一理论构建不乏创新和匠心,但由此也使大陆主流证据知识体系遁入按照哲学话语体系的逻辑进行构建的误区。而且,这些思想和概念由于先入为主的效应还成为当前证据理论无法抛弃的认识前提和知识基础。理论上的集体无意识,与实践中形成的以证据三性为证据资格审查标准的通行且无法变更的话语现实,就是明证。正因为这种“路径依赖”,大陆主流证据理论在关于网络电子数据证据资格问题的论述时,实际上包括了本文在先讨论的关联性在内。

此外,在将真实性视为电子数据证据资格的元素时,会使得真实性这一概念本身在所指上出现混乱,既包括电子数据的可靠性,也包括电子数据内容在证明力上是可以采信的,以致相关的理论分析越来越复杂,越来越使人费解。[13] 本文遵从严格的证据资格理论,以合法性和可信性作为证据资格的要素,这一研究进路也可视为一种对应然证据资格理论的回归。[14]

(一)网络电子数据的可信性

这包括基础数据本身即数据库及算法模型分析的结果两方面的可信性问题。笔者认为,传统的规制电子数据可信性的证据规则即传闻证据规则及最佳证据规则仍然是规制网络电子数据的有效规则。其中,传闻证据规则作为英美法系最具特色的证据规则,因其具有发现案件真实、保障程序公正以及完善证据制度等诸多价值,符合我国正在推进的以审判为中心的当事人主义庭审制度改革所需要的配套证据制度规则的需

〔11〕 详细论述参见何家弘著:《新编证据法学》,法律出版社 2000 年版,第 104 页。

〔12〕 参见沈德咏著:《刑事证据制度与理论·总序一》,法律出版社 2002 年版。有关 1979 年以后中国大陆证据法学的研究还参见何邦武:《近代证据法学知识系谱研究:意旨、方法与进路》,《求索》2015 年第 2 期。

〔13〕 此类文章理论的基础都在以真实性替代可采性分析,以致难以排除可采性与可信性的混乱。参见褚福民:《电子证据真实性的三个层面:以刑事诉讼为例的分析》,《法学研究》2018 年第 4 期;刘品新:《论电子证据的理性真实观》,《法商研究》2018 年第 4 期;刘译矾:《论电子数据的双重鉴真》,《当代法学》2018 年第 3 期。

〔14〕 这种研究路向的理据还在于,即使在大陆法系,也存在与英美法系可采性规则相近的理论与制度,因此,可作为证据资格的普适性法理,成为一种理想类型(马克斯·韦伯)。参见郭志媛著:《刑事证据可采性研究》,中国人民公安大学出版社 2004 年版,第 64 页以下。

求,具有引入该制度的必要性与可行性。[15] 可借鉴域外一些国家的相应规则,作为正常业务活动记录的例外予以采纳。[16] 而且,我国现行的《中华人民共和国合同法》(第 11 条)及《中华人民共和国电子签名法》(第 2、4 条)等也已经以功能等同的方式确立了电子数据具有书面形式的证据资格,为其作为传闻规则的例外奠定了制度基础。关于最佳证据规则,其证据保管链制度在应对网络电子数据时尚无新的挑战。但必须注意的是,此类规则尚不足以有效规制网络电子数据。因为,不同于单纯的"电子设备存储记录与衍生记录",因其由计算机系统自动生成,可作为传闻证据例外而被采用。[17]

在网络电子数据的取证和认证中,因算法而搜集的数据具有了人为的因素(数据建模)。由此必然使侦查人员和司法人员产生困惑:这样取得的证据所倚赖的算法可信与否?这里的疑思涉及三个方面问题:一是用作算法对象选取的相对合理性。大数据的流动性使得任何时间段截取的数据都具有抽样性,不是也无法得到反映全面情况的数据。由此使得数据的收集和以此为基础的算法结论只具有相对性。二是算法过程本身受多重因素影响。算法可以被理解为"计算的方法和技巧",是在高级语言发展了很多年之后,更多地被封装成了独立的函数或者独立的类,开放接口供人调用的程序性设计。计算机中的算法大多数指的就是一段或者几段程序,告诉计算机用什么样的逻辑和步骤来处理数据和计算,然后得到处理的结果。因此,算法的运用是一个辩证的过程,不仅在于不同算法间的比较和搭配使用有着辩证关系。即使在同一个算法中,不同的参数和阈值设置同样会带来大相径庭的结果,甚至影响数据解读的科学性。[18] 三是与前述算法的自身属性相关,大数据算法结论自始即为概率性存在而非百分之百的确定。有学者因此指出,"'大数据'通常用概率说话,而不是板着'确凿无疑'的面孔……当我们试图扩大规模的时候,要学会拥抱混乱",因此,"除了纠结于数据的准确性、正确性、纯洁度和严格度之外,我们也应该容许一些不精确的存在"。[19]

因算法而改变数据原值的案例不难在实际生活中发现,以下的事例就曾被热议且难以否认存在这种可能。2019 年 4 月 13 日,爱否科技的前高级主笔王跃琨在微博中,针对华为 P30Pro 手机拍摄月亮的图片所发表的评论,大意是华为 P30Pro 手机拍摄的月亮照片存在 PS 的情况(后改称存在 AI 计算的情况),即拍摄的照片不排除有算法的

〔15〕 有关传闻证据制度引入我国的必要性与可行性论述,参见何邦武著:《刑事传闻规则研究》,法律出版社 2009 年版,第 251 页以下。

〔16〕 例如,根据美国《联邦证据规则》803 条第 6 款的规定,由知情人以任何形式制作的关于行为、事件、情况、意见或诊断的备忘录、报告、记录或者数据汇编(包括电子计算机存储器)。如果是在当时或其后不久制作的,或者是根据传来信息制作的,并为正常业务活动所保存,而且,该类制作是此业务活动的正常做法,并能由保管人或其他适格证人作证来证实,则可予以采纳。英国《1984 年警察与刑事证据法》第 69 条则规定了源自计算机记录的证据可采性问题。且其《2003 年刑事审判法》第 129 条关于"人以外的描述"也是有关计算机类证据的规定。

〔17〕 有关电子数据可信性的规则、理论及我国现行法律中的相关规定的梳理,参见何邦武:《论电子数据取证程序的法律规制》,《云南大学学报(法学版)》2012 年第 3 期。

〔18〕 参见高扬、卫峥、尹会生:《终于有人把数据、信息、算法、统计、概率和数据挖掘都讲明白了!》,数据来源于微信大数据公众号,https://mp.weixin.qq.com/s/mV6NLOl_uZK3-9HcdSOlAg,最近访问时间[2019-04-19]。

〔19〕 [英]维克托·迈尔舍·恩伯格著:《大数据时代:生活、工作与思维的大变革》,周涛译,浙江人民出版社 2012 年版,第 46、240 页。

介入。[20] 显然,因算法搜集的证据已超出传闻证据规则对传统电子数据规制的范围。而且,就网络电子数据产生的过程本身判断,这样的数据天然地存在着受质疑的风险,从而极大地动摇了其可信性。

(二)网络电子数据对证据合法性的冲击

表面看来,由于搜集对象的非特定性以及数据获取的随机性,网络电子数据取证具有"海选"的特点,类似于传统物理空间犯罪侦查中的"初查",而无需受刑事强制侦查的令状审查原则、比例原则、保密原则等的限制。然而,在大数据与物联网时代,传统法律制度及理念中有关政府、企业、个人等之间的权力与权利关系格局正在改变。大数据分析中,数据的最终使用情况,已经远远超出个人的意图层面,甚至超出个人的认知范围,个人数据的可能用途实际上已难以为个人所预测。因为,主要来自政府部门和网络平台企业的程序员正在编写着各种程序,分析和处理海量数据,以致使数据时代成为一个"算法"统治的世界,且复杂的算法正在使当下的信息时代成为"黑箱社会",出现算法歧视或算法对人的支配现象,置身其中的个人越来越透明,而来自政府和网络平台企业的力量对个人的控制变得越来越晦暗。[21] 大数据算法的这种非人格化、弥漫性和以数理定律的方式形成的滥权极易使普通公民的权利遭受侵害。[22] 传统的强制侦查与任意侦查行为之间的界限在大数据侦查中已经变得越来越模糊。

网络经济与社会交往的发展使网络安全问题越发凸显,个人数据信息保护的需求将更加迫切,个人数据权利作为个人基本权利必将成为不争的事实。因为基本权利是"论证社会制度正当性的最终依据",具有制度设定原理和基础上的先在性,而且,"一旦把个人权利作为正当性的最终根据,正当的社会组织再也不是高于个人的有机体,而是为个人服务的大机器,甚至家庭和国家亦变成了一个契约共同体"。[23] 有鉴于此,欧盟于 2018 年 5 月 25 日正式实施的《一般数据保护条例》(*General Data Protection Regulation*,GDPR),以基本权利的方式,对个人数据保护设定了范围最为广泛、权利类型最为多样的保护条款,成为当前各法域中对个人数据保护最为严格缜密的法律。[24]

〔20〕 该微博引起了网络上的诸多讨论,有网友通过反复实验,证实王跃琨的说法在一定意义上具有一定合理性,即华为 P30Pro 手机在进行拍摄时,如果被摄物被识别为是月亮,则会触发相应的算法,该算法会在最终成像进行处理,包括但不限于减少或是增加个别内容,而这里的减少或者增加,有观点认为可能是参照某张特定的月亮图片进行的。参见《华为 P30 Pro 拍月事件为何能引发持久争议?》,https://www.xianjichina.com/news/details_113911.html,最近访问时间[2019-05-04]。

〔21〕 参见郑戈:《在鼓励创新与保护人权之间:法律如何回应大数据技术革新的挑战》,《探索与争鸣》2016 年第 7 期,第 82 页。

〔22〕 参见郑戈:《在鼓励创新与保护人权之间:法律如何回应大数据技术革新的挑战》,《探索与争鸣》2016 年第 7 期,第 82 页。

〔23〕 参见金观涛著:《探索现代社会的起源》,社会科学文献出版社 2010 年版,第 12 页以下。有关个人数据权利保护方法的论述参见丁晓东:《个人信息私法保护的困境与出路》,《法学研究》2018 年第 6 期,以及李伟民:《"个人信息权"性质之辨与立法模式研究:以互联网新型权利为视角》,《上海师范大学学报(哲学社会科学版)》2018 年第 3 期。

〔24〕 根据该《一般数据保护条例》第三章(第 12 至 22 条)的规定,个人享有的数据权利十分广泛,包括数据主体有权获取控制者的身份及详细联系方式、处理目的,以及对个人数据的访问权、更正权、擦除权(被遗忘权)、限制处理权、数据携带权、一般反对权、反对自动化处理权等。在美国,加州消费者隐私保护法(CCPA)在个人数据保护中亦具有代表性,相关研究参见孙海鸣:《GDPR VS 加州隐私法:欧美这两部个人数据保护法规有什么差异?》,http://www.sohu.com/a/316058079_786964,最近访问时间[2019-05-31]。

2019 年 5 月 14 日，作为美国高科技革命阵地，谷歌、脸书等高科技公司根据地的旧金山通过了一项法案，率先对人脸识别技术采取了“反对潜在滥用”的立场，禁止执法部门和其它机构使用人脸识别技术。这一现象也说明了当前人工智能技术与公民个人权利保护之间存在着较大的冲突和紧张关系。因此，在涉及个人数据权利关系问题的处理上，国家机关、任何社会组织都应该信守谦抑的立场，设定并严守自己应有的边界，在合理使用个人数据和保护个人数据权之间维持适当的平衡，这是任何在信息时代建设网络法治国家都应遵守的法则。

根据上述个人数据权保护与合理使用的应然法理，在我国网络犯罪侦查的算法取证中，以下行为对相对人数据权利或利益的侵害，极有可能发生：由于算法取证实际上是一种人工智能对数据的自动化处理，因而，这一处理过程本身不仅构成了对与数据有关的相对人权利的侵害，而且，还可能对相对人的隐私权、取证时的参与权、取证时的知情权等造成侵害。此外，如果在取证中使用了本该销毁的数据，或者过度处理了相对人的数据，同样构成对相应数据权利的侵害。2015 年发生在北京的任甲玉案[25]，说明被遗忘的数据权利已经与中国的法律和社会正式碰撞。有理由相信数据权利将成为中国网民维权的重要手段。这是证据合法性规则在算法取证中面临的又一重困境。

四、传统证明模式的影响：网络电子数据算法取证的外部挑战

我国庭审中证据调查和证据认证制度及与之相应的理论产生的“倒逼”效应，对审前取证制度和实践形成的制约，在网络电子数据取证中同样存在。其中，客观真理观理念下的客观真实的证明标准及印证证明模式，是影响网络电子数据取证的重要外部因素。笔者曾就有关网络犯罪中因刑事证明中的印证证明模式及客观真实证明标准的影响所致证据搜集困难的问题进行过分析，认为既有的印证模式及其所固守的证明标准观念产生的逆向效应，固化乃至强化了侦查取证中的证供结合取证模式及办案思维，漠视甚至忽视经由间接证据形成的证据链证明案件事实的收集间接证据的取证模式。[26]

印证模式及客观真实的证明理念产生的负效应在算法取证中同样存在，以前文所举的通过网络购买配件组装枪支为例，公安检察部门之所以不予立案或不起诉，就是出于对

〔25〕 参见(2015)一中民终字第 09558 号，(2015)海民初字第 17417 号。该案的基本情况是，原告任甲玉系管理学领域的从业人员，其于 2014 年 7 月 1 日起在无锡陶氏生物科技有限公司从事相关的教育工作，至 2014 年 11 月 26 日解除劳动关系。但从 2015 年 2 月初开始，原告陆续在百度网站上发现“陶氏教育任甲玉”“无锡陶氏教育任甲玉”等字样的内容及链接。原告认为，由于陶氏教育在外界颇受争议，“陶氏教育任甲玉”“无锡陶氏教育任甲玉”等侵权信息给原告名誉造成极大侵害，且原告曾多次发邮件给被告要求删除相关内容，但是被告没有删除或采取任何停止侵权的措施，讼争由此引起。有关该案的详细内容及讨论参见杨立新、杜泽夏：《中国“被遗忘权”第一案任甲玉诉百度公司名誉权纠纷案裁判理由评述》，《法律适用·司法案例》2017 年第 16 期。

〔26〕 参见何邦武：《小额多笔网络电信售假和诈骗犯罪取证问题研究》，《政治与法律》2016 年第 8 期；何邦武：《近代证据法学知识系谱研究：意旨、方法与进路》，《求索》2015 年第 2 期。有关审判程序对审前程序的影响参见龙宗智著：《刑事庭审制度研究》，中国政法大学出版社 2001 年版，第 232 页以下。

此类经算法取得的间接证据证明链所形成的证明力以及算法本身的质疑,因为没有可以相互印证的相关证据尤其是口供的支撑性印证,甚至只有口供对算法取得的证据证明链的否认性证明(如以枪是自己捡来的理由抗辩)。即使援引最高人民法院、最高人民检察院、公安部 2016 年颁行的《关于办理电信网络诈骗等刑事案件适用法律若干问题的意见》,“综合认定”此类算法所取得的证据,也难以解决实践中的难题。因为,“综合认定”虽然有部分放弃印证证明模式的意向,然而在以印证证明模式和客观真实证明标准为核心的制度系统中,这样的解释性规则显然无法撼动刑事诉讼基本法的制度和理念。该“意见”实施以来的实践也证明,由于规则缺乏操作性,理论及实践中难以就“综合认定”形成共识,以致实践中对如何适用“综合认定”规则仍然缺乏可操作性,导致“规则虚置”,已经成为不争的事实。〔27〕

正是由于受客观真实证明标准的影响,在电子数据的认证上,有学者试图借助概率理论,通过概率的乘积规则设计电子证据印证公式,以此建构统一的认证标准,使电子数据证明力数字化。该学者坚信“电子证据采信应当走向客观化”,认为“在证据采信标准的客观化转型过程中,数学是基础,概率论是钥匙,概率化则是标志”。而要解决电子证据的采信问题,必须在设置可操作性标尺方面取得突破。该学者试图通过计算概率,使“事实清楚,证据确实、充分”的证明标准实现一定程度的具体化,以此打通一定区域内公、检、法办案流程,防止公、检、法人员在电子数据证明力认证中的迷茫,或者防止其认证过程中的恣意,并举出司法实践中的几则电子数据适用的案例补强其观点。〔28〕

笔者认为,概率论无非是归纳逻辑在现代社会的一种呈现方式,是“用概率论的定量分析和公理化、形式化的手段探索有限的经验事实对一定范围内的普遍原理的归纳支持和确证程度”。〔29〕而且,在人文社会科学领域,这种“经由统计和概率所辨认出的一系列社会规律,与自然规律有着实质上的不同。虽然同样被称为规律,但人们新发现的这类规律脱胎于偶然性,而非必然性。”〔30〕因此,没有必要对概率持一种神化的态度,以为一旦通过计算就能实现确信无疑。这实际是一种流于表面上的科学精神和理性而实质上的反科学精神和非理性,其实质无异于中世纪的法定证据制度在现代科技文明光影下的复活。

〔27〕 该《意见》第二部分(四)规定:“因犯罪嫌疑人、被告人故意隐匿、毁灭证据等原因,致拨打电话次数、发送信息条数的证据难以收集的,可以根据经查证属实的日拨打人次数、日发送信息条数,结合犯罪嫌疑人、被告人实施犯罪的时间,犯罪嫌疑人、被告人的供述等相关证据,综合予以认定。”第六部分(一)还规定:“办理电信网络诈骗案件,确因被害人人数众多等客观条件的限制,无法逐一收集的,可以结合已收集的被害人陈述,以及经查证属实的银行账户交易记录……等证据,综合认定被害人人数及诈骗资金数额等犯罪事实。”根据上述规定,可以看出,该《意见》出台的意图,在于突破网络新型犯罪中电子数据收集、提取中的困难,以及被害人人数众多、分布广泛且无法确定、搜集等的难题,使被害人陈述不再作为认定案件事实的关键证据。有关网络电信犯罪中“综合认定”的法理及实际运用分析,容笔者另文论述。事实上,所谓综合认定,从来就是法庭证据调查后对证据证明力进行综合性衡量和评价的方式,在其他非电子数据的法庭调查中,早就存在。作为一种心证的方式,中外理论一直孜孜以求的是,以何种方式使其可视化(如威格摩尔的心证图谱)或者可量化(如贝叶斯定理的运用等)而外显于人,但心证过程的复杂性使一切的解释都沦入一说即俗或者永远说不清的境地。就此而言,该规定本身属于无制度增量的解释性规则,其实践中的无果实属必然。有关网络犯罪中综合认证问题的详尽分析参见何邦武:《“综合认定”的应然解读与实践进路》,《河北法学》2019 年第 8 期。

〔28〕 本段引用有关观点参见刘品新:《印证与概率:电子证据的客观化采信》,《环球法律评论》2017 年第 4 期。

〔29〕 崔清田著:《现代逻辑科学》,天津教育出版社 1990 年版,第 223 页。

〔30〕 张保生著:《证据科学论纲》,经济科学出版社 2019 年版,第 114 页。

此为其一。其二,认为用概率这一数字化手段可以决疑的观点实际是对概率方式的误读。因为,与根据纯粹的数学知识进行计算得出某一结论一样,根据算法模型所得出的某一现象的概率,都是一种数字化的证据,如何根据某一数据作出判断,仍然需要法官(人脑)结合其他证据综合衡量,而不能简单地将其理解为由数据得出结论。以女王诉莎莉·克拉克案(Regina v. Sally Clark)为例,这是一个经典的运用内曼—皮尔逊准则的案例,其前后相左的判决可以说明所谓数据作为证明标准的非至上性,同样的情形还发生在美国人民诉柯林斯案中。[31] 从该文举出的几则适用概率裁决的网络案例来看,其判决的运思过程中,都是将概率作为裁决的一种依据。而在这一对数据概率的解读过程中,经历了从纯粹的数据(人工语言)到案件事实判断(自然语言)的跃升,正是在后一环节,融入了人类对数据所欲证明事实的价值判断,是纯粹的数据无法完成的。总之,试图通过数据、概率实现裁判标准的客观化仍然是一种不切合实际的幻想,反映了作者对问题思考和解决方式的非理性。

五 完善网络电子数据算法取证的适恰之路

网络电子数据算法取证过程中产生的诸多困境,再次验证了信息时代人工智能对传统法律制度及其理念的冲击效应,[32]说明了法律应及时跟上人工智能发展步伐,对人工智能进行规制的必要性、迫切性。基于法的安定性考量,结合当下人工智能的发展水平及其应用状况,运用法教义学的方法,以"传统法律修正模式"对人工智能进行规制应是一种可行的进路。[33] 申言之,针对网络刑事犯罪惩治中的算法取证,可以在既有的刑事证据规则体系内,依照法教义学的原理,通过对既有证据制度及网络电子数据规则条文"语义空缺"或者"规范漏洞"的合目的性弥补,使算法取得的电子数据具有可采性,回应前文关于网络电子数据算法取证所面临的挑战,实现对网络电子数据算法取证的有效规制。

(一)算法取证关联性的合理解释

第一,虽然算法取证依照的是数理逻辑中的模糊理论,其数据抓取过程自身是人工智能的自动识别的结果,与传统证据法律中基于经验而作出关联性判断进行收集证据有别。然而,不应忽视的是,算法所依赖的模型是由人脑建构的,其设计、目的、成功标准、数据使

〔31〕 See[2000]EWCA Crim 54. 该案的基本情况是,1996 年 11 月,莎莉的孩子仅 11 周的克里斯托弗(Christopher)在她面前离世。死亡原因被归结为婴儿猝死综合症(sudden infant death syndrome,SIDS)。大约一年后,其另一个孩子仅 8 周的 Harry 离世。两个孩子死亡时,仅莎莉一人在场。莎莉被控谋杀自己的两个孩子,并于 1999 年被判终身监禁。莎莉于 2000 年第一次上诉,后败诉。2003 年,莎莉再次上诉,被判无罪。关于 People v. Collins 案的评述,参见易延友:《通过计算实现正义》,《数学文化》2013 年第 5 期。

〔32〕 有研究者断言,人工智能技术的发展将改变人们对法律的认知,重塑法律的规则形态,乃至法律的价值导向,参见 Drew Simshaw, NicolasTerry, KrisHauser and M. L. Cummings, Regulating Healtheare Robots: Maximiizing Opportunities While Minimizing Risks, *Richmond Journal of Law & Technology*, 2016, 22 (3): 1 - 38。

〔33〕 参见汪庆华:《人工智能的法律规制路径:一个框架性讨论》,《现代法学》2019 年第 2 期。但笔者认为,法的安定性考量不是唯一和排他的价值目标,当制定法明显不法,超出人们可以忍耐的限度,失去其之所以为法的"法性"时,就可以否认既有立法,唯此方可谓符合拉德布鲁赫公式的要求。

用等都是设计者、开发者的主观选择。在这一过程中,设计者和开发者可能将自己所怀有的偏见嵌入算法系统,以致算法在本质上是“以数学方式或者计算机代码表达的意见”,[34] 离不开人的主观因素,这自然无法排除设计者自身所带有的基于经验的“前见”。因而,在算法建模过程中,自然无法排斥关联性的考量,即如何找到其想要的数据,而这个所谓“想要”,就是一种自然人基于某种现实理由且带有经验性的判断。该算法建模中所依存的思维方式由此将与证据收集中的关联性规则暗合。并且,在提供给算法模型识别的数据中(如相应的数据仓的建立),仍然需要建立在初始的关联性基础上。

第二,被抓取的数据最终能否被用于某一案件,最终离不开人脑的判断,而这一判断的基础首先就是关联性考量。因此,立足算法取证全过程的纵向视角,关联性仍然是其中必备的因素,不能仅仅因为算法环节的纯数理形式而否认其中基于经验的关联性因素的存在。归纳言之,在算法取证的深层逻辑中,仍然有人类经验性思维的基础。或者说,证据收集中的关联性规则在算法取证中只不过是换了一种表达方式。

(二)确立和完善算法取证可采性的相关制度

1. 保证算法取证的可靠性

鉴于算法取证中网络科技和算法建模运用的特点,笔者认为,算法取证所得的证据宜视为科技证据的一种,应当借鉴科技证据可采性的规则与理论,在对算法取证进行可靠性全方位思考的基础上,确立相应的可靠性规则。目前而言,可以借鉴的科技证据可靠性规则与理论,当为美国《联邦证据规则》702 条以及联邦最高法院判决的 Daubert 案和作为此案延伸的“锦湖轮胎案”(将可靠性标准延伸到技术领域)中所确立的规则。这一规则主要包括以下内容:该科学原理的可证伪性;该原理已知的错误率;该理论已经同行评议或者公开发表的程度;该原理或技术在相关科学团体中达到“普遍接受”的程度。[35] 概括起来,该规则包括某一原理自身的“科学”程度(可证伪性和社会接受度)及依据该原理得出结论的实践状况(错误率)表现原理与实践两个方面。以之衡量作为一种技术的算法取证(类似于锦湖轮胎案)算法取证规则的具体化应当是:最终能使算法模型及其原理具有可解释性,[36] 且能够得到相关领域专家的认同,同时,应当公开根据某一算法的错误率,以增强算法的社会接受度。另外,强化算法运行前数据供给的客观、全面、有效,即通过数据采集的充分性保证算法基础的可靠性。而数据供给的有效、全面,唯有验诸数据源的收集程序,通过算法模型运行前数据来源的公开、合法、全面来保证。

---

〔34〕 腾讯研究院:《算法决策兴起:人工智能时代的若干伦理问题及策略》,http://www.sohu.com/a/143165651_455313,最近访问时间[2019-06-06]。

〔35〕 Daubert v. Merrell Dow Pharmaceuticals, Inc., Supreme Court of the United States, 1993, 509 U.S. 578; Kumho Tire Company, Ltd. v. Carmichael, Supreme Court of the United States, 1999, 526 U.S. 137, 119 S. Ct. 1167, 143 L. E. 2d 238. 在 Daubert 案之前的 Frye 案中,美国联邦最高法院还确立了关于科技证据的普遍接受标准,但 Daubert 案则在关于科技证据的采纳中,援引联邦证据规则 702 条,增加了法官作为此类证据采纳的守门人义务,赋予法官对证据采纳一定的自由裁量权限,强化了法官在此类证据采用中的主体性地位。笔者认为这是有关科技证据可采性标准进步的表征,详尽论述容笔者另文撰述。

〔36〕 算法的解释包括内部解释和外部解释,算法的内部解释针对的是技术人员,其标准是可诊断性,目的是为了解决技术问题。内部解释涉及算法源代码和商业秘密,不宜公开。这里指涉的是外部解释,目的是为了实现算法公开、公平,最终使那些与算法有关的人员在其数据权利受到侵害时有救济的机会。

2. 建立和完善算法取证合法性的制度

这方面可资借鉴的是欧盟《一般数据保护条例》(以下简称《条例》),该《条例》第2章专门为数据处理设定了6项原则,对于确立算法取证的合法性规则具有借鉴意义。这些原则包括:合法性、合理性和透明性原则;目的限制原则;数据最小化原则;准确性原则;限期储存原则以及诚实与保密原则。该《条例》还规定,数据控制者有责任对上述原则的遵守与否提供证明。[37] 结合个人数据权利面对其取证行为过程中强制侦查与任意侦查之间日渐模糊的界限的现状,为使个人数据信息免于被自动化决策,在进行算法取证时,应当参照刑事侦查取证尤其是技术侦查的基本原则和理念,[38] 设定适当的程序性原则并据此确定相应的取证规则,以保证算法取证的合法性。

从更广泛的视角来看,数据采集同样应遵守侦查程序的一般原理,受制于该一般原理下的侦查程序的基本原则。[39] 考虑到话语使用的可通约性和交流的便捷性,算法取证的合法性原则可化约为通行的关于侦查的基本原则,即程序法定原则、比例原则、保密原则和限制使用原则等。首先,程序法定原则旨在强化算法取证中的技术理性以实现刑事侦查法治的价值理性,通过设定算法取证主体的权力资格、技术资质,明确算法取证的程序等,防止政府及大数据平台的算法专制,避免算法黑箱,侵害相对人的数据权利等。其次,比例原则在其本原意义上包括合目的性原则、必要性原则和相称性原则等3项子原则,目的是通过对侦查手段在使用目的、使用限度上的限制,保障相对人的权利。对照欧盟《条例》,比例原则的内涵与其中的目的限制与数据最小化原则暗合。在具体适用该原则时,可以借鉴美国在数据权利保护中的场景原则,在个人数据权利与公共安全之间的利益衡量中,将比例原则具体化。[40] 再次,传统意义上的侦查保密原则,也被称为"侦查密行原则"或"侦查不公开原则",主要是指侦查活动中所涉及到的侦查内容对当事人保密和对社会成员保密。限制使用原则,主要指侦查资料的使用对象、使用范围、使用时间、保存手段及时间,以及无关材料的及时销毁等均应遵守严格的法律强制性规定。最后,保密原则和限制使用原则可对应于上述原则中的目的限制、数据最小化、限期存储及诚实与保密原则。在将保密原则和限制使用原则落实到算法取证时,可以将其内容参照《条例》具体化、特定化。还应说明的是,上述原则的确立,还具有宣示性价值,即明确告知公权力机关在进行算法取证时,必须也不能自外于刑事侦查的基本原则。

此外,根据以权利制衡权力的原理,通过明确赋予数据主体的数据权利,保障数据权利主体数据使用的知情权,以及被侵权后的救济权,强化数据平台(公私性质均应如此)数据搜集使用中对相对人的告知义务等,以此形成一种逆向监督效应,促使算法合

〔37〕 各原则的详细内容,参见欧盟《一般数据保护条例》第2章中的有关规定。

〔38〕 有关技术侦查的基本原则的分析,参见何邦武等:《现行职务犯罪技术侦查的基本原则评议》,《净月学刊》2016年第4期。

〔39〕 侦查程序与人权保障是刑事侦查和证据收集中的永恒话题,研究者代有其人,也常有新的研究成果问世,但笔者仍愿推荐孙长永教授的《侦查程序与人权:比较法考察》(中国方正出版社2000年版)一书。以资料的梳理和论证的深度而言,此书迄今仍为该领域的扛鼎之作,启迪甚多,值得细读。

〔40〕 有关美国数据保护场景化理解的阐述,参见丁晓东:《个人信息私法保护的困境与出路》,《法学研究》2018年第6期。

规,使公权力维系其谦抑的品性,在客观上也可以实现规范权力,保障取证合法的目的。数据权利作为一种基本人权,在我国还面临着完成法理上如何证成,以及如何走向符合国内网络经济发展与人权保障有效均衡的实际,完成顶层设计,走向立法的问题,涉及宪法和其他法律的法理与制度。这是一项法学研究者及立法者必须面对的现实,本文恕不一一展开。[41] 面对各类新型数据权利,尽管我国目前尚无个人数据保护的专门立法,但对个人数据权利进行保护将是我们信息经济健康发展的必由之路,同时,在我国既有的法典如《电子商务法》《数据安全管理办法(征求意见稿)》国家互联网信息办公室关于《儿童个人信息网络保护规定(征求意见稿)》等分散立法中,已经有一定的数据权利保护制度,具备了对数据权利进行保护的基础。立足我国国情,数据主体究竟应当赋予哪些权利,值得进一步探究,但立法确立和保护数据权利势在必行。在这方面,可资借鉴的立法有欧盟《条例》及美国加州消费者隐私法案(*California Consumer Privacy Act*,CCPA),除了前文提到的数据权利外,两部法典列出的数据主体反对数据处理的权利、选择退出权等,均对我国将来的立法有借鉴意义。[42]

(三)刑事证明理念与方式的转变

完善数据取证制度,需要我们转变刑事证明的理念,实现刑事证明方式向现代自由心证的回归,逆向助推算法取证制度的完善。如前所述,受传统认识论的影响,在我国刑事证明语境中,客观真实的刑事证明标准已经成为我国刑事证明的执念。由追求客观真实而衍生的印证证明模式,被一部分学者自负地视为高于自由心证且能保证发现案件真实的特色模式。然而,理论的自负无法消除实践中因追求客观真实或法律真实所致的弊端,修改刑事证明理念及其指导下的证明模式,并向对话和论证证明模式转变,势所必然。必须意识到,在刑事诉讼证明中,事实认定的对话和论证属性,是"从演绎证明到对话证明,从'封闭'到比较'开放'的推理形式,从不容质疑的权威到在不同解决方案之间辩证选择,已成为一种趋势,尽管这一趋势是在各种法律传统或法律制度内部发生的,是渐进的而非突发的"。[43]

不仅如此,与认知过程中人的主体性地位日渐凸显相一致,作为自然人的法官在证据评价中正日益发挥着更加积极的作用。借助伽达默尔现代诠释学和哈贝马斯的法律论证理论,可以断言,自由心证作为对证据和事实的"理解"已经具有"此在"的本体论意义,借助"理解"这一中介,理解的主体和理解的对象合二为一。同时,由于知识只具有相对确定性,是由参与对话的各理性主体经由对话协商的模式,在确保所有对话的参加者有平等的机会、能自由言谈、任何一方没有特权、对话在诚实且免于被强制的条件下,通过理想的

〔41〕 我国理论界对数据权利保护研究刚刚起步,数据权利概念本身亦处于如何界定、如何达成共识的初始阶段。详尽论述参见丁晓东:《什么是数据权利?——从欧洲〈一般数据保护条例〉看数据隐私的保护》,《华东政法大学学报》2018 年第 4 期;许娟:《中国个人信息保护的权利构造》,《上海大学学报(社会科学版)》2019 年第 3 期。

〔42〕 二者在保护数据权利上存在较大区别,从价值目标上看,GDPR 偏重对个人数据的保护,而 CCPA 偏重促进数据流通,详细权利内容参见孙海鸣:《GDPR VS 加州隐私法:欧美这两部个人数据保护法规有什么差异?》,http://www.sohu.com/a/316058079_786964,最近访问时间[2019-06-12]。

〔43〕 张志铭著:《法律解释操作分析》,中国政法大学出版社 1999 年版,第 208 页。

程序设制和严格的理由论证而获得的。其中,法官在使法律具体化于每一个特殊情况时,不乏创造性补充行为,发挥着积极的建构作用,[44] 因此,参与法庭调查的各方主体,通过法律论证,在对话、辩论和质证中,达成对案件真相的可接受性共识,是现代自由心证理念下,法庭调查和由此形成案件真相认知的应然进路。

申言之,在法律论证过程中,案件真相的形成具有建构的属性,是参与法庭调查的各方主体在不断地对话中逐渐建构的关于某一案件的"事相",而不应被视为一种简单的从前提到结论的三段论逻辑,即"前提—结论"(案件事实前提、法律规则前提和法律结论)的封闭模式。相反,其认定的前提是开放式的,承载的命题是断言式的,而其过程是诉讼各方参与情势下的辩驳、对话、交涉和说服,从而,最终认定的结果应当具有合理的可接受性。[45] 易言之,这种以断言性语句表现出来的命题,其内容的客观性(真值)不是取决于它与外在的"经验的"某种因果关系,而是取决于它是否可以在理想的认知条件下"兑现"。由此,原先与客观的外部事物的性质联系在一起的"真理"的概念(还原主义本体论),已转化为"理想的证明方式"或"理想的可接受的正当性"。[46]

回归现代自由心证,承认并践行案件事实调查的法律论证属性,还应摒弃以第一代认知科学为基础的工具主义语言观念,引入以第二代认知科学为基础的语用主义语言学理论,重新审视法律论证中语言的角色和功能。起源于20世纪50年代的第一代认知科学,以客观主义理论为基础,并从分析哲学中吸取了符号运算理论,对推理采取了形式分析的方法。其中,语言被视为世界的镜像,具有精确的表征功能和再现功能,是忠实反映世界的工具。认为只要方法得当,人们就能透过语言精确地了解世界,理解他人及其作品。显然,这一语言观"丢弃了人在认识范畴、形成概念、进行推理、建构语义系统中的主观能动性因素,忽视了人的身体经验、生理构造、认知方式、丰富想象力等所起到的作用"。[47] 以体验哲学为基础的第二代认知科学则坚持身体与心智不可分的观点,认为"概念和意义是一种基于身体经验的心理现象,是人类通过自己的身体和大脑与客观世界互动的结果",从而,意义是主客体之间互动的结果,而不是通过符号与世界之间客观的、直接的连接而产生的,这就颠覆了第一代认知科学基于客观主义的真值对应论、真值条件论。[48] 以上述语言学理论衡量我国当前刑事证明中的客观真实或者法律真实观,可以发现,我国当前有关案件事实的理论主张正是以第一代认知科学为基础的,必须予以更新,唯此方能

〔44〕 参见[德]阿图尔·考夫曼、温弗里德·哈斯默尔主编:《当代法哲学和法律理论导论》,郑永流译,法律出版社2002年版,第376页以下;郑永流:《出释入造——法律诠释学及其与法律解释学的关系》,《法学研究》2002年第3期。有关伽达默尔诠释学原理的分析及其与现代自由心证证明方式的契合,参见何邦武、马作武:《现代诠释学视野下的自由心证》,《山东社会科学》2005年第5期。

〔45〕 这里还要澄清的是,以合理的可接受性标准替代基于本质主义的客观真实标准,并不意味着案件事实认知上的相对主义,以及由此导致的证明标准的丧失。可接受性标准的客观性证成源于语言的先在性社会现实。详细论证参见何邦武:《"综合认定"的应然解读与实践进路》,《河北法学》2019年第8期。这里还涉及上述两个观点不同的基础性共识问题,笔者下文关于语言哲学的论述也可为疏证。

〔46〕 刘钢著:《真理的话语理论基础:从达米特、布兰顿到哈贝马斯》,人民出版社2015年版,第397页以下。

〔47〕 王寅著:《认知语言学》,上海外语教育出版社2007年版,第55页。

〔48〕 参见王寅著:《认知语言学》,上海外语教育出版社2007年版,第55、58页。有关第一代认知科学与第二代认知科学的区别以及后者在法律论证中的该当性,容笔者另文详为分析。

实现刑事证明理念向现代自由心证的真正回归。[49]

为使法律论证理论从制度上得以落实,还应消除刑事庭审制度的痼疾,通过落实证人出庭等制度,以保障庭审的论辩性、对话性,使摒弃印证证明模式后的心证模式有可以信赖的基础。[50] 与此同时,真正理解并落实《刑事诉讼法》第 57 条“没有被告人供述,证据确实、充分的,可以认定被告人有罪和处以刑罚”,即以间接证据、证据链证明案件真相的规则,解构印证模式所赖以存续的基础。[51] 此外,针对大数据算法取证所得证据而进行的法庭调查的特点,还可根据《刑事诉讼法》第 197 条“公诉人、当事人和辩护人、诉讼代理人可以申请法庭通知有专门知识的人出庭,就鉴定人作出的鉴定意见提出意见”的规定,有针对性地设立专家辅助人制度,让建构算法取证模型的人员出庭对有关算法的问题作出不涉及商业秘密的解释,接受法庭的质询,在有利于调查案件事实的同时,也可以实现看得见的正义。

六 结 语

信息时代,大数据对传统法律观念及其规则所提出的全方位挑战,以朱苏力教授刻画的法律人对待科技、信息试图采取视而不见的“鸵鸟政策”应对已绝无可能。[52] 正视信息科技对法律的深层次冲击,更新既有的法律观念及规则,或者“阐旧邦以辅新命”,以法教义学的方法,通过对既有规则的合法理性、合目的性弥补,为网络空间的治理提供合法有效的规则,将是植基于工业文明的既有法律制度及其法理实现内在超越的必由之路。

本文关于网络犯罪惩治中算法取证的论述即遵循上述理路,即以既有的证据法理为基础,在既有的证据规则范围内,涵摄算法取证这一新生现象。循此进路,算法的关联性可以在放宽的视野中得到新的合理解释,算法的可靠性可以在传统的科技证据法理中找到支撑,如此等等。其中,对既有证据规则及法理构成较为严峻挑战的当为算法取证的合法性问题,由于新生的数据权利的多样性、易受侵害性,对此类权利的保护从

〔49〕 关于法律真实与客观真实的分析著述较多,且相互还存在争论,但其实二者的理论基础是一致的,由此决定了二者都“不能为证明标准、证据规则提供正当性”。参见王敏远:《一个谬误、两句废话、三种学说:对案件事实及证据的哲学、历史学分析》,载王敏远《公法》(第 4 卷),法律出版社 2003 年版,第 172 页。对客观真实进行比较精致论证的当数张继成教授,该氏以早期分析实证主义的“真值语义学”为底色,以命题与事实的符合为基础,折中真理符合论、实效论、融贯论、语义论等真理学说,试图构建一个完整的真的保证性、真的核证性标准、真的有效性标准以及真的合理可接受标准的证明标准体系。但该论述自始即带有强烈的本质主义还原论色彩和客观主义的幻觉,不仅无法消弭各种真理观的相互对立之处,也无法消除其主张的内部矛盾性。参见张继成:《诉讼证明标准的科学重构》,《中国社会科学》2005 年第 5 期。

〔50〕 现行刑事庭审制度虽然设定了强制证人出庭作证制度,但既有的规定仍存在较大随意性,以致实际上很难保证证人能出庭接受对质询问,仍存在通过适当的法理阐释进一步完善的问题。有关分析参见何邦武:《证人出庭作证例外的裁量性标准问题探析:基于修订后〈刑事诉讼法〉第 187 条、第 188 条的分析》,《政治与法律》2013 年第 5 期。

〔51〕 经由间接证据证明案件真相的制度规则与法律实践之间相背离的现象,根源在印证证明模式及制约该模式的客观真实证明标准。详尽分析及解决之路参见何邦武:《小额多笔网络电信售假和诈骗犯罪取证问题研究》,《政治与法律》2016 年第 8 期。

〔52〕 参见苏力:《法律与科技问题的法理学重构》,《中国社会科学》1999 年第 5 期。

而对相关数据平台、公权力机构算法取证行为的限缩就显得尤为迫切。然而,发展数字经济,促进数据流通,又需要此类约束必须适度,如何维系保护数据权利与限制算法取证行为二者的平衡,将随着数据存储情势、取证样式的不断翻新而构成对算法取证行为长期的挑战。

尤应申明的是,面对网络算法取证这一新生事物,本文主张通过对既有规则的扩张性解释,将其纳入既有证据可采性(证据资格)规则内,基于既有规则的法理,对其进行规范。与这一思维进路相一致,本文同时认为所谓为网络电子数据新立统一量化的采信标准主张具有伪科学性,否认该主张的可取性,具体原因已在文中进行了阐述。总之,立足证据关联性及可采性规则的应然法理,坚守现代自由心证的刑事证明理论,从印证证明理念及其模式中真正解缚,重新审视语言的角色与功能,仍然是解决网络电子数据算法取证难题的不二法门。

[本文为作者主持的2018年度浙江大学互联网法律研究中心课题“网络犯罪惩治中的电子数据收集取证问题研究”(2018B02)的研究成果。]

[**Abstract**] Algorithmic forensics has become the only choice to collect network electronic data in the network era. Because of the particularity of network electronic data storage and presentation, algorithmic forensics faces the problems such as how to define relevance, how to ensure reliability and how to protect legitimacy. At the same time, algorithmic forensics also encounters the restraint of the adverse effect caused by traditional criminal proof concept and mode. A feasible way to solve the above problems is to trace the principles of Legal Dogmatics, inference the oughtness jurisprudence of the relevance and admissibility of evidence, re-interpret the empirical rules in association rules, introduce the admissibility criteria of scientific and technological evidence, and establish the basic principles and corresponding rules of legality in algorithmic forensics by referring to the provisions of data rights in the EU General Data Protection Regulations. abandoning the idea and practice of corroboration, returning to modern doctrine of discretional evaluation of evidence, and reviewing the role and function of language in legal argumentation.

(责任编辑:贾　元)