

# 论大数据证据

刘品新

**内容提要:**大数据开启了证据法制建设的时代转型。将形形色色的大数据材料用作证据以证明案情,是当下司法实践无法回避的一道题目。该现象的背后存在着重要的价值论和方法论基础。概要地说,大数据证据以海量电子数据凝练的规律性认识发挥证明作用,其主要以分析结果或报告的形式呈现。对于这一证据如何定位,各国法律界存在着不同的理论争论和实务处理,主要包括鉴定意见说、专家辅助人意见说、证人证言说等。考虑到大数据证据具有专业性和科学性,中国现实的便宜选择是将其纳入鉴定意见的证据法定形式。对于这一证据如何审查判断,我国应当聚焦真实性与关联性规则进行创新。具体来说,针对海量数据本身的真实性问题,要建设以“大”真实性为主的真实性规则;针对大数据分析结果的真实性问题,要构建判断机器算法是否可信的真实性规则;针对大数据分析结果所揭示的关联性结论,要构建基于整体数据与具体数据分层的关联性规则,特别是探索超越人类经验判断的关联性规则。

**关键词:**大数据 大数据证据 鉴定意见 机器证言 证据规则

刘品新,中国人民大学法学院教授。

## 一 引言

“大数据开启了一次重大的时代转型。就像望远镜让我们能够感受宇宙,显微镜让我们能够观测微生物一样,大数据正在改变我们的生活以及理解世界的方式,成为新发明和新服务的源泉,而更多的改变正蓄势待发”。<sup>[1]</sup> 这是2013年英国学者维克托对大数据改变社会的预言。近年来,世界范围内司法领域确实发生了巨变,突出表现为基于机器算法、机器人和高级人工智能等新兴科技的各种大数据材料涌向法庭。

[1] [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1页。

美国实务中已出现了将汽车事件数据记录系统数据<sup>[2]</sup>、谷歌地球卫星图像和 GPS 坐标数据<sup>[3]</sup>、“查找我的 iPhone”功能数据<sup>[4]</sup>等用作证据的案例。这些材料的特点是数据量巨大,并可被编译为分析报告。如汽车事件数据记录系统是专门记录汽车行驶数据的系统,它通常可以在撞车发生前、中、后,按照秒级自动提取车辆运行、系统状态、司机指令、撞车特征等各种数据,以在需要时还原汽车事件真相。我国的司法舞台也上演着同样的故事。早期,海量的通信数据、资金流数据、账户信息、网页数据、网络舆情信息、GIS 地理信息及专业数据库的数据已被运用于侦查活动中。新近,人们又接续看到“舆情分析报告”<sup>[5]</sup>、“百度指数”<sup>[6]</sup>、“淘宝客反作弊系统数据”<sup>[7]</sup>、“轨迹大数据”<sup>[8]</sup>、“公安大数据平台数据”<sup>[9]</sup>、“证券交易监控系统数据”<sup>[10]</sup>、“大数据资金分析报告”<sup>[11]</sup>等用于证明的情景,同时遇到控辩双方产生争议及裁判者任意裁决的种种问题。这些都需要我们针对性地进行制度建设。本文尝试对大数据证据的基础问题作出回应,以期推动实务发展和法制变革。

## 二 大数据用作证据的理论基础

证据是证明的根据,但并不是一切用于证明的根据都属于法定的证据。在司法实践中,某些可用作证明根据的特定材料会被习惯性地排除出证据范围。一些新生事物常会遭遇这一规律。它们通常难以为人们快速地普遍接受,如视听资料、电子数据等也是经过很长一段时候后才入法正名。<sup>[12]</sup>现如今,形形色色、以海量信息特征呈现的大数据材料,便处于有待法律认可的境地。

大数据材料具有“容量大”、“种类多”、“价值密度低”等特点,尤以“容量大”为基础特征。笔者调研 e 租宝案件时发现,该案在侦查阶段收集的海量数据包括:(1)e 租宝及

[2] Commonwealth v. Salas, 95 A.3d 304, 308, 309 (Pa. Super. Ct. 2014).

[3] United States v. Lizarraga-Tirado, 789 F.3d 1107, 1109, 1110 (9th Cir. 2015).

[4] Pickett v. State, 112 A.3d 1078, 1090 (Md. Ct. Spec. App. 2015).

[5] 参见《山西省晋城市中级人民法院刑事判决书》(2017)晋 05 刑终 271 号。在该案中,控方举证了公安局制作的关于刑满释放人员程幼泽出狱网上舆情相关数据统计及舆情专报。

[6] 参见《广东省高级人民法院民事判决书》(2013)粤高法民三终字第 565 号。在该案中,原告提交了公证人员百度指数查询 qunar 和 quna 两个关键词的指数对比情况。

[7] 参见《杭州市余杭区人民法院民事判决书》(2014)杭余民初字第 3 号。在该案中,被告(淘宝)提供了抓取淘宝客异常推广行为的反作弊系统的数据。

[8] 参见《江苏省高级人民法院刑事裁定书》(2017)苏刑终 135 号。在该案中,公安机关依法调取的警务大数据平台之车辆大数据(截图)。

[9] 参见《四川省成都高新技术产业开发区人民法院刑事判决书》(2018)川 0191 刑初 392 号。在该案中,控方提交了公安大数据平台及情报系统查询资料。

[10] 参见《北京市第一中级人民法院行政判决书》(2013)一中行初字第 1171 号。在该案中,中国证监会提交了证券交易监控系统记录的数据作为证据,证明这些证券账户网上委托交易所用的 IP 地址、Mac 地址高度重合。

[11] 参见《湖南省长沙市中级人民法院刑事判决书》(2018)湘 01 刑终 18 号。在该案中,控方提交了利安达鉴审财报字[2016]第 1209 号鉴定意见书,以证明非法集资的金额和投资人数。

[12] 在我国,视听资料为 1996 年《刑事诉讼法》修法时入法,电子数据为 2012 年《刑事诉讼法》修法时入法。此前,关于视听资料、电子数据的证据地位一直存有争议。

芝麻金融数据；(2)集团 OA 系统中关于会议、财务、合同的数据；(3)关于公司及其产品介绍的电子数据；(4)涉案人员的手机数据等。它们的数据体量大到难以想象的程度。其中，“e 租宝及芝麻金融数据”至少包括从 4000 多家银行、247 家第三方支付平台、164 家保险公司、114 家券商汇总的 1 万多个账户的几十亿条资金交易流水信息，“集团 OA 系统中关于会议、财务、合同的数据”主要是部署在阿里云平台的、涉及 200 多台服务器的数据。全案数据的总量达到 30TB 左右。这些海量数据交由司法人员一条一条地筛选使用根本不现实，但可以借助数据统计、数据碰撞<sup>[13]</sup>、数据挖掘<sup>[14]</sup>与机器学习<sup>[15]</sup>等方法形成分析报告，用作办案支撑。这一现象不是特例，随着司法实践的发展，各种类似的报告还会花样翻新。那么，面对大数据材料，我们是抱陈守旧地将之局限于作为侦查线索，还是将其延展一步用作审判环节的证据？这就需要从基础理论层面进行考量。

### (一) 大数据用作证据的价值论

当下，依靠大数据证明案件事实已经成为一种客观需要。这首先是因为案件专业化、巨型化演变态势所决定的。在越来越专业、复杂的案件中，用于证明案件事实的证据也必然趋于专业化和复杂化。当今涉众型犯罪就呈现出涉案人数不特定、证据巨量的特点。据报道，我国金融领域涉众型犯罪案件的涉案金额已超过百亿元、千亿元，涉案人员达到十万人、百万人，数额之大、人数之众令人瞠目。笔者调研发现，在一线司法机关办理的涉众型犯罪案件中证据材料就达到数千卷，背后是司法人员庞大的人力与时间付出。此种情形下，仅仅凭借传统证据来证明案件事实面临着难以承受之重。与之相反，通过大数据证明案件事实就明显高效得多，尤其是基于海量数据的分析报告通常浓缩为具体的一份或多份文书证据，使用起来更具可操作性。

其次，这符合降低司法证明难度的导向。证据是一种稀缺的资源，证据短缺是人类司法的永恒规律。千百年来人类一直在寻找有效化解证明难题的方法，其中包括对新型证明方法的探索与推广。当前我国正面临着惩治以网络犯罪为代表的新型犯罪的艰巨任务，不同层面的法律规范一直千方百计地试图降低证明难度，包括更多引入“综合认定”<sup>[16]</sup>、“综合评估”<sup>[17]</sup>的规则，甚至尝试推出含有“证明责任移转”<sup>[18]</sup>意味的条款。学理

[13] 数据碰撞意指对两个以上数据库资源进行智能查询比对。

[14] 数据挖掘意指从大数据中通过算法搜索隐藏于其中的信息。

[15] 机器学习意指让机器从数据中学习，得到一个或多个更符合现实规律的模型，通过使用模型使得机器表现得更好。

[16] 最高人民法院、最高人民检察院、公安部《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》中规定：“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，无法逐一收集被害人陈述的，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、第三方支付结算账户交易记录、通话记录、电子数据等证据，综合认定被害人人数及诈骗资金数额等犯罪事实。”

[17] 最高人民法院、最高人民检察院《关于利用网络云盘制作、复制、贩卖、传播淫秽电子信息牟利行为定罪量刑问题的批复》中提到：“鉴于网络云盘的特点，不应单纯考虑制作、复制、贩卖、传播淫秽电子信息的数量，还应充分考虑传播范围、违法所得、行为人一贯表现以及淫秽电子信息、传播对象是否涉及未成年人等情节，综合评估社会危害性，恰当裁量刑罚，确保罪责刑相适应。”

[18] 最高人民法院、最高人民检察院《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第 11 条第 3 款规定，“对批量公民个人信息的条数，根据查获的数量直接认定，但是有证据证明信息不真实或者重复的除外”。很多人理解这是将关于个人信息真实性和是否重复的证明责任移转至辩方。

上也提出了“底线证明法”<sup>[19]</sup>、“抽样取证法”<sup>[20]</sup>、“等约计量法”<sup>[21]</sup>等建议。若能直接从扩充证据范围的角度进行制度建设,即引入大数据以证明案情,将具有更为深远的意义。实践中,我国许多案件虽未将大数据列明为证据,但仍技巧性地发挥出它的证明作用。<sup>[22]</sup>这在很大程度上是受制于旧观念,却也反证了司法证明中存在着大量的需求。

最后,大数据用作证据具有独到的价值。每一份具体的传统证据反映的是案件中具体的人、事、物、时、空等信息;与之不同的是,大数据反映的是案件整体或作为其很大一部分的人、事、物、时、空等信息。这一差异在涉众型犯罪案件中尤为明显。调研发现,涉众型犯罪案件中难以通过传统证据证明的案件事实,都有可能通过大数据材料找到便捷的解题之策。比如,在证明涉众型犯罪案件的组织架构时,传统上调查工商注册材料等证明方法面临着工作量巨大、材料真假难辨的困扰,而海量资金流数据等能够像 DNA 一样将涉众型犯罪中不同涉案人员识别并关联起来,揭示不同人员在犯罪组织中的具体角色;<sup>[23]</sup>又如,在证明涉众型犯罪行为人为主观故意时,传统的证明方法主要是通过口供或者“通过客观行为推断主观明知”<sup>[24]</sup>,易遇到嫌疑人以不知情为由进行辩解的挑战,而利用数据挖掘方法则能发现行为人存在着规避打击的习惯做法,<sup>[25]</sup>从而以犯罪习惯证明其主观故意。目前这样的探索取得了初步成功,表明以大数据化解司法证明难题之路,值得进行实践推广和理论提升。

## (二) 大数据用作证据的方法论

大数据材料并非空中飘来之物,客观上就是一个个具体数据的聚合。当聚合到一定规模之后,将其中的数据规律揭示出来,就能够用于还原案件事实。如在一起涉税案件中,办案人员调取了涉案的全量资金数据——资金流水数据上亿条、账户数达上万个,聘请技术专家“建立了资金特征分析模型……在很短时间内完成可疑资金网络的刻画,利

[19] 底线证明是指按照法定的人罪和加重处罚两道“坎”,提供能用以定案的最基本的证据。参见刘品新:《网络犯罪证明简化论》,《中国刑事法杂志》2017 年第 6 期,第 35 页。

[20] 抽样取证是指办案人员从较大数量的物品中提取具有代表性的一定量的物品作为样本证据,并据此证明全体物品属性。参见万毅、纵博:《论刑事诉讼中的抽样取证》,《江苏行政学院学报》2014 年第 4 期,第 120 页。

[21] 等约计量是指对海量化的犯罪对象进行“估堆式”计量,作为认定犯罪数额的标准。参见罗猛、邓超:《从精确计量到等约计量:犯罪对象海量化下数额认定的困境及因应》,《预防青少年犯罪研究》2016 年第 2 期,第 40 页。

[22] 参见《宁波市江北区人民法院刑事判决书》(2017)浙 0205 刑初 411 号。在该案中,控方提交的宁波市社会保障管理局报案书等证据表明,社会保险管理部门通过医保大数据分析审查,发现励某某等参保人员异常就医、违规配药情况,有诈骗医疗保险金的嫌疑。在审判中,医保大数据分析意见没有标识为证据,其实质内容被吸收到书证等证据中。

参见《新疆维吾尔自治区乌鲁木齐市中级人民法院刑事判决书》(2017)新 01 刑初 200 号。在该案中,为了证明被告人普某某的行程轨迹,控方将被告人普某某的 1 张登机牌提交,证实与公安机关通过大数据平台分析的侦破本案的线索一致。在这里,大数据平台分析的意见仅仅用于说明书证“登机牌”的证明价值,未单独列为证据。

[23] 2015 年,公安部会同国家税务总局、人民银行用 8 个月时间查处了一起虚开增值税专用发票犯罪案件。办案人员收集到 10 亿多条资金流水,利用分析模型分析虚开公司资金流向,清楚地将涉案公司分为开票公司、收票公司、中转公司三层,根据资金流向区分出会员单位、二级单位等,绘制出资金流的闭环关系。这样的公司层级和会员关系,就揭示出涉案单位和人员的具体角色。

[24] 参见刘福谦:《涉众型经济犯罪案件证据审查的几个问题》,《人民检察》2018 年第 4 期,第 13 页。

[25] 通过对海量资金流进行数据挖掘可以发现,许多嫌疑人存在着有规律性地瞬间转账账户来款、用一元钱给自己账户转账来测试账户是否正常、通过同一渠道将资金转出等规避打击的规律性反常行为。

用关系可视化技术清晰展现可疑资金的来源和去向,并且自动标注账号和主体的类别标签”。<sup>[26]</sup> 在这里,起到证明作用的就是海量数据凝练而成的规律性认识——具体呈现为专业分析报告。这是大数据用作证据的特色及魅力所在。

这样的分析报告类似于鉴定人或其他专家的意见,但也与之有明显不同,其专业化程度已经超越人类的专家。准确地说,其是借助于机器算法形成的判断。在前述 e 租宝案件中,北京检察机关曾经对海量的电子数据做了分类处理。其中,对于涉案人员的手机数据和产品介绍数据,主要是通过扣押、勘验、检验、远勘的方式形成扣押清单、勘验笔录、数据刻盘、远勘笔录等证据,这是以电子介质的载体和电子数据的内容来证明案件事实;对于相关的资金流水数据、OA 办公系统数据,主要是通过通知调取、扣押、勘验、委托鉴定的方式形成镜像数据、扣押清单、勘验笔录、鉴定意见书等证据,这是以转化性证据来证明案件事实。它们的共同之处是利用证据所蕴含的内容或信息。回看该案的办理过程,司法证明的思路在整体上仍然是将大数据材料转化为传统证据,即以转化性证据办案。如此办案不仅费时费力,而且也呈现出“固守”传统证据的陈旧观念。假如该案能够将基于海量资金数据的各种分析报告(其时主要是服务于侦查工作)提交法庭,那就跃升到一个更高级的司法证明层次,即将大数据用作证据证明的新境界。

### 三 大数据证据的法律地位

从技术原理上讲,大数据变身证据通常要经过三个环节:第一步是汇总数据并进行数据清洗,第二步是建构分析模型或机器算法,第三步是进行运算形成分析结论。这一过程结束后,就会产生大数据证据。为论述的便利,本文将“大数据证据”限定为基于海量电子数据形成的分析结果或报告。在这个定义中,海量的电子数据是“源”,分析得出的结果或报告是“果”。直接起证明作用的往往是“果”,即大数据分析结果或报告。当然,在特殊情况下,从海量电子数据中“淘”出来某一条或多条数据也可能起到证明作用。这一条或多条数据也可以被理解为“果”。<sup>[27]</sup> 对于如何为大数据证据赋予合理地位,目前有如下几种选择方案。

#### (一) 鉴定意见

一种主流观点认为,大数据证据应当作为鉴定意见使用。有论者针对“资金大数据分析结论”提出,“最好把资金大数据分析纳入司法鉴定范畴。这有利于司法实践的展开,在法律上也可以找到依据”。<sup>[28]</sup> 在组织传销案件、非法吸收公众存款案件、集资诈骗案件中,确实出现了很多这样的新型鉴定意见书。如在王某某等组织、领导传销活动罪一案中,司法鉴定机构对“××平台后台数据库”进行了检验分析,得出的鉴定意见为:该网站注册账户 1×××个,成功激活 1×××个;缴纳会费账户 1×××个,空单账户 1×

[26] 何家弘等:《大数据侦查给证据法带来的挑战》,《人民检察》2018年第1期,第54页。

[27] 参见《江苏省高级人民法院刑事裁定书》(2017)苏刑终135号。

[28] 何家弘等:《大数据侦查给证据法带来的挑战》,《人民检察》2018年第1期,第56页。

×个……根据推荐与被推荐的层级关系,共计 22 级;收取会员投资共计 26 × × × × × × 元;除被告人所有的账户外,还有 1 × × × 个会员账户共计 17 × × × × × × 元投资款没有返还”。<sup>[29]</sup>

然而,鉴定意见本意是由鉴定机构的鉴定人对专业问题做出的专业判断。将大数据证据纳入司法鉴定的范围,必然会遭遇诸多挑战。首先,法条上语焉不详,导致该做法的合法性存疑。我国公安部《公安机关鉴定规则》、司法部《司法鉴定执业分类规定(试行)》、最高人民检察院《人民检察院鉴定规则(试行)》均未将大数据鉴定明确纳入司法鉴定的业务范围。<sup>[30]</sup> 其次,在鉴定方法、鉴定主体等具体细节方面存有障碍。就鉴定方法来说,鉴定机构出具的现有“大数据分析鉴定意见书”显示,采取的鉴定方法通常是《电子物证数据搜索检验规程》<sup>[31]</sup>《数字化设备证据数据发现提取固定方法》<sup>[32]</sup>,而其实际检验过程同这两个方法并不完全对应;就鉴定主体来说,“大数据分析鉴定意见书”在很大程度上是由机器算法给出实质判断——不同于以往专家借助仪器设备做出判断,这对于以由专家做出判断的司法鉴定体制是一个过于超前的突破。

## (二) 专家辅助人意见

2016 年最高人民法院、最高人民检察院、公安部印发《关于办理刑事案件收集提取和审查判断电子数据若干问题的规定》第 16 条规定:“对扣押的原始存储介质或者提取的电子数据,可以通过恢复、破解、统计、关联、比对等方式进行检查。”这一法律条文规制的范围已超越司法鉴定工作,而更多的是专家辅助办案活动。其中的关键词“统计”、“关联”、“比对”等通常理解可用于大数据分析。2018 年最高人民检察院发布司法解释,明确了各级人民检察院可以指派、聘请有专门知识的人,运用专门知识参与办案活动,协助解决专门性问题或者提出意见。<sup>[33]</sup> 专家辅助人意见是一种比鉴定意见更加宽泛的概念。任何有专门知识的人参与办案,提出的意见都是专家辅助人意见。那么,大数据证据可否作为“专家辅助人意见”(或“有专门知识的人意见”)<sup>[34]</sup> 提交呢?

这种观点比较接近于前述“鉴定意见说”,但在很大程度上还属于一种学术观点,因为诉讼法中并不存在“专家辅助人意见”这一证据形式,故若延伸至司法层面仍需要进一步定位。最高人民法院《关于适用〈中华人民共和国民事诉讼法〉的解释》第 122 条第 2 款规定:“具有专门知识的人在法庭上就专业问题提出的意见,视为当事人的陈述。”<sup>[35]</sup> 《关于适用〈中华人民共和国民事诉讼法〉的解释》第 87 条第 1 款规定:“对案件中的专门性问题需要鉴定,但没有法定司法鉴定机构,或者法律、司法解释规定可以进行检验的,可

[29] 参见《山东省济南市中级人民法院刑事裁定书》(2017)鲁 01 刑终 88 号。

[30] 一些学者认为,可以类比电子数据鉴定开展大数据鉴定。例如,根据《人民检察院电子证据鉴定程序(试行)》第 4 条规定,可以将大数据鉴定纳入电子证据鉴定的范围。

[31] 《电子物证数据搜索检验规程》规定的是电子物证检验中数据搜索检验的方法。

[32] 《数字化设备证据数据发现提取固定方法》规定的是从数字化设备发现提取固定证据数据,并保证证据数据原始性和证据数据完整性的方法。

[33] 参见《最高人民检察院关于指派、聘请有专门知识的人参与办案若干问题的规定(试行)》第 2 条及第 7 条。

[34] 这里所说的专家辅助人,三大诉讼法中的正式术语是“有专门知识的人”。

[35] 在司法实践中,法官确实习惯于这么处理。参见 2018 年最高人民检察院第八批指导性案例第 28 号,法庭的审理表明:“专家意见(对鉴定意见或专业问题)经质证,可以作为认定事实的根据。”

以指派、聘请有专门知识的人进行检验,检验报告可以作为定罪量刑的参考。”这说明,大数据证据相应地可以作为“准当事人陈述”或“定罪量刑的参考”。当然,关于该问题现阶段还存在着广泛争议,我国有必要通过立法明确“专家辅助人意见”的证据地位。<sup>[36]</sup>

### (三) 证人证言

美国学者洛斯教授指出,机器在解决法律争议事实中正扮演着越来越重要的角色,机器传递出的一些信息可作为“机器证言”。<sup>[37]</sup>她列举的“机器证言”包括专家系统的意见等。她还解释,与人类证言一样,机器证言的可信性取决于其来源的可信性;而类似于“传闻的危险”潜藏在人类主张中,“黑箱危险”(人类或机器的错误导致机器被错误的设计)也潜藏在机器传达的信息中。这一“机器证言说”,理论上可适用于大数据证据。但我国还普遍不接受机器证言这类概念,其在法律中也不是法定证据形式,故对大数据证据套用“机器证言说”仍存在较大困难。

不过,在我国司法实践中,已经出现以大数据证据辅助或质疑证人证言的例子。在程某某贪污罪一案中,湖北省住建厅2016年通过大数据对比有重大案情发现,控方在庭审中将此内容纳入证人朱某的证言予以举证。<sup>[38]</sup>本案中大数据证据转化为证人证言使用。在赵某某组织、领导传销活动罪一案中,为质疑证人胡某,控方询问“通过大数据查找你的下线人数为291人,你怎么解释?”胡某回答“我一共发展了10多个,剩下的那些下线都是多个人发展的,具体多少数我不知道。”<sup>[39]</sup>这则是将大数据证据作为一种质证方式。

### (四) 其他证据

司法实践中,还存在着将大数据证据作如下处理的情形:(1)作为物证、书证。如在孙某某、周某某掩饰、隐瞒犯罪所得、犯罪所得收益、信用卡诈骗、传授犯罪方法罪一案中,控方举出“关于手机号187××××8422的大数据查询”结果,证明该手机机主“自2016年12月份以来连续多天××网咖上网”。<sup>[40]</sup>且裁定书中明确将大数据查询结果标示为定案的书证、物证。(2)作为“破案经过材料”。在陈某某等盗窃罪一案中,控方举出“本案的侦破经过”,证实“通过大数据分析等技侦手段,确定被告人陈某某手机运行轨迹与案发地在时间上、空间上相吻合”。<sup>[41]</sup>(3)作为单列出来的报告。例如,在唐某某操纵“航天动力”股票案中,为证明唐某某是19个账户的实际控制人,证监会提交了如下证据:19个账户的Mac地址、IP地址具有高度重合性的数据分析(实质上属于大数据证

[36] 在制定《最高人民法院人民检察院关于指派、聘请有专门知识的人参与办案若干问题的规定(试行)》时,对有专门知识的人给出意见的证据地位曾经征求过不同部门、部分知名专家学者的意见。最高人民法院、最高人民检察院侦查庭和部分专家认为,不宜作为证据使用,建议写为“以作为诉讼材料使用”;最高人民检察院刑事申诉厅建议从“可以作为定罪量刑的参考”角度规定;最高人民检察院反贪四局和部分专家认可“意见”的证据资格,并建议明确具体类别;人大法工委、公安部和部分专家则建议再慎重考虑。

[37] Andrea Roth, Machine Testimony, 126 Yale L. J. 1972 (2017).

[38] 参见《湖北省麻城市人民法院刑事判决书》(2017)鄂1181刑初273号。

[39] 参见《吉林省松原市宁江区人民法院刑事判决书》(2017)吉0702刑初460号。

[40] 参见《山东省菏泽市中级人民法院刑事裁定书》(2018)鲁17刑终430号。

[41] 参见《湖北省安陆市中级人民法院刑事判决书》(2017)鄂0982刑初13号。

据);19个账户交易股票品种具有共同性的数据分析;19个账户交易行为具有一致性的数据分析。法院最终也据此认定了这些账户由唐某某操纵。

前述各种做法均体现了现阶段的实务探索或理论思考。在具体案件的语境中,每一种处理均有一定的合理性;在具体探讨的学理上,每一种观点均能自圆其说。经分析可知,上述探索和思考呈现出一定的共性规律。剖析其中的规律,能够为大数据证据的合理定位提供一份答案。

问题之一,单独讨论大数据证据的证据地位还是结合其他证据一起考虑。当下的实践中是两种情况并存的。但笔者认为,考虑大数据证据的地位,归根到底,还是要讨论其单独发挥作用时的定位问题,故应当限于前一种情况。在第二种情况下,大数据证据就是简单用作其他证据的支持或反驳材料,这只能算为讨论问题提供了素材,其处理结果并不具有参考价值。

问题之二,在现有证据形式的法律框架内还是不限于法律框架进行讨论。理想的学术研究是应然层面的讨论,相应地对大数据证据的定位也就不能局限于我国法定的八种证据形式。换言之,既可以在物证、书证、证人证言等法定证据的框架内将大数据证据对号入座,也可以呼吁未来将大数据证据单列出来作为一种新的证据形式。而后一种选择有赖于司法案例和实践经验的丰富,更有赖于学术理念和司法观念的与时俱进。我国学者张建伟指出,“未来的证据法当中,大数据分析有必要单列出来作为独立的证据种类”。<sup>[42]</sup>这种未来之策是有道理的。而现实的研究从实然层面讨论,多数人认为大数据证据的定位应该放入我国现行的法定证据形式。这也是实务部门容易接受的做法,如有些裁判文书将大数据证据列入物证、书证、鉴定意见的范围,有的将其转化为证人证言。就我国现有法律框架而言,实务中最常见的做法、学理上最流行的观点当属“鉴定意见论”。

笔者也认同现阶段的这一便宜选择。大数据证据无疑是一种专业性或科学性很强的证据,其结论部分的意见表达是普通人无从凭借常识就能理解的。而在我国现行法律框架中,只有鉴定意见与之相符。前述少数案例中裁判文书将大数据证据列为物证、书证或证人证言等,多少有些生搬硬套的感觉。将大数据证据纳入鉴定意见之列,既具有形式上的亲缘性、可比性,也满足司法追求经济性的原则,司法人员可以相对容易地援引鉴定意见的审查判断规则去处理大数据证据的效力。诚然,严格地讲,大数据证据同一般意义上的鉴定意见也存在前述一些差异,这有待于设立其采用规则时进行必要的调整。

## 四 大数据证据规则的建设

大数据证据并不是普通司法人员能够解读的,甚至不是具有专门知识的人能够辨析的,其极致形态是机器学习的产物。大数据证据如何适用证据规则?回答这个问题,要厘清传统证据规则用于规范大数据证据的审查判断方面已经和可能遇到的挑战,应当全面

[42] 何家弘等:《大数据侦查给证据法带来的挑战》,《人民检察》2018年第1期,第56页。



审视“大数据证据+证据规则”的问题与出路,特别是聚焦大数据证据在真实性与关联性的审查判断方面出现的特殊难题,并进行特色证据规则的创新。

### (一)创建以“大”真实性为主的大数据来源真实性规则

人们对大数据真实性的质疑大体上包括数据本身与分析结果两个方面。在此先讨论数据本身的真实性问题,即如何对海量数据进行逐一核实?通常的疑问包括:海量数据中是否存在不实的、误导的数据?是否存在着重复的、过时的数据?是否存在其他不为人所知的错误呢?例如,在一起侵害商标权纠纷案件中,原告举出了“淘宝指数”和“百度指数”,拟证明被告使用相似的商标标识造成相关公众混淆。<sup>[43]</sup>如果对诸如此类的指数所依据的数据产生真实性的疑问,从表面上看针对的是“淘宝指数”和“百度指数”的公信力,实质上则针对的是“淘宝指数”和“百度指数”的数据来源及质量。

从学理上讲,数据源的真实性审查包括整体数据的真实性和具体数据的真实性。前一层含义是指大数据作为一个整体数据集的真实性;也就是说,赖以做出大数据分析报告的数据(通常呈现为一份或多份数据库文件)必须是客观形成及收集的真实文件。至于该份或多份数据库文件中每一条记录或信息是否属实,则是后一层含义。简言之,两者分别指向宏观真实性与微观真实性。

按照这样的思路,司法实务中对数据来源真实性的审查判断通常要分两步走。以“兴中天案(涉嫌传销犯罪)为例,第一步是宏观真实性的审查。兴中天公司在被查处之前已经将传销网站资金盘停盘,公安机关介入后调取的是该公司为稳定会员情绪开发的网上商城数据。此时需要审查数据在从传销网站迁移到商城网站的过程中是否经过篡改。办案人员调查发现了解这一情况的,只有一名掌握核心技术的公司技术人员。经向该证人取证,其承认“该数据系其从传销网站迁徙到商城网站的原始数据,未进行删改”。这就说清了该数据集的来源。第二步是微观真实性的审查。办案人员一是调取部分传销参与人员的银行交易明细,同数据库中150万条排单打款记录及90万条收款记录进行比对,确认能够相互印证;二是选择该传销组织的一个县分支机构进行查证,部分印证了该数据层级关系和会员数量的准确性;三是对每一个犯罪嫌疑人发展的下线进行抽样取证,连取三级,每一级选择两名会员调取证据,对会员数量和层级进行验证,以完成微观核实的工作。

大数据证据的宏观真实性方面能够、也必须做到查证无疑。然而,实务中这一点却往往被忽略。在一起组织传销犯罪案件中,控方指控的证据主要是上海××计算机司法鉴定所提取该公司会员系统后台数据的鉴定意见及有关会计师事务所做的会计报告。其中,该鉴定意见书表明,鉴定人的主要工作是远程登录相关网站提取某某公司会员系统的

[43] 参见《广东省高级人民法院民事判决书》(2015)粤高法民三终字第444号。在该案中,原告所举的“淘宝指数”和“百度指数”具体表述为:179882号《公证书》显示,在淘宝指数搜索栏输入“新百伦”时,出现“新百伦男鞋”、“新百伦女鞋”等结果;在淘宝指数上搜索“新百伦”和“new balance”时,显示二者的搜索指数曲线基本一致。179823号《公证书》显示,在百度指数搜索栏输入“新百伦”和“new balance”进行对比显示,发现对“新百伦”的搜索量远远高于对“new balance”的搜索量。最后,法庭综合多份证据,认定了造成相关公众混淆的事实。

数据。<sup>[44]</sup> 然而,该次鉴定选择的技术标准错误,且关于鉴定实施过程的录像被“掐头去尾”,使得整个鉴定过程在关键环节上无法查清;在涉案网站已经无法访问的情况下,无从对涉案数据集的真实性进行核实。之后,会计师事务所据此得出关于吸收会员层级、人数、资金数的判断,其真实性就必然大打折扣。

大数据证据在微观真实性方面很难做到查证无疑,一般要求进行必要的验证。英国学者维克托指出,“‘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”<sup>[45]</sup> “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。数据不可能是完全对或者完全错的。当数据的规模以数量级增加时,这些混乱也就算不上问题了”<sup>[46]</sup> 这里说的是,大数据中只要在整体上达到一定规模的具体数据属实,对案件事实的认定就不会产生实质性影响。如在一起组织传销犯罪案件中,控辩双方围绕是否存在一人多号、虚拟账户、上级会员为下级会员垫付投资款的情况产生了争议。二审法院经审理后认为“由于涉案人数众多,且存在不少会员进行匿名登记的情况,本案无法逐一核实会员身份是正常的。但一审已经基于相关鉴定意见,对空单账户进行了排除,目前也没有其他相反证据,据以对鉴定的会员人数和参与传销资金数额进行扣减,并且,鉴定的会员人数和收取的参与传销资金数额,远远超过司法解释规定的‘情节严重’认定标准,因此,一审依据鉴定意见认定有关犯罪数额,并据此认定各被告人犯罪属于情节严重也无不当”<sup>[47]</sup> 在这里,法院对如何看待“网站后台数据库信息中每个账户、会员、人员”信息的真实性,是有道理的。可见,大数据中只要确保真实的具体数据在整体上达到一定规模,那就不会对案件事实的认定产生实质性影响。

形象地说,宏观真实性可称为“大”真实性,微观真实性可称为“小”真实性。大数据证据的真实性规则应当是以“大”真实性为主,兼涉“小”真实性的规则。“大”真实性审查要严格,“小”真实性审查则不能僵化,两个部分合为一个整体。

## (二) 构建针对机器算法是否可信的大数据分析结果真实性规则

关于大数据分析结果的质疑,主要指向结论的可靠性。人类经验表明,数据越多越大,得出结论越可靠。我国有学者指出,“数据之大,充分抵消了少数个体的特殊性,类似于在众多裁判中去掉最高分和最低分。显然,裁判越多,打分的结果就越接近运动员的真实水平。”<sup>[48]</sup> 但这里的特殊性在于,大数据分析结果是由机器算法得出的,有的机器算法还涉及数据模型。人们提出此类质疑,实质针对的是机器算法等分析方案的可信度。当下许多设计者将机器算法与数据模型视为核心的商业秘密,并不愿公开细节。由此,人们对机器算法之参数是否正确及数据模型之逻辑合理与否,难免心怀疑虑。

[44] 参见《上海××计算机司法鉴定所司法鉴定检验报告书》[2017]鉴字第69号。

[45] [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页。

[46] [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240页。

[47] 参见《山东省济南市中级人民法院刑事裁定书》(2017)鲁01刑终88号。

[48] 吕乃基:《大数据与认识论》,《中国软科学》2014年第9期,第36页。

一种观念认为,既然基于机器学习的大数据证据被作为证据提交法庭,那就应当要求设计者公开技术方案。如果不公开,大数据证据就难以满足真实性标准。反过来,“如果设计者能够保证机器算法的透明度,那么该工具就更容易被法庭接受,而且这些计算的透明度也会同时提高刑事司法系统的透明度”。<sup>[49]</sup> 这一看法过于武断。洛斯教授指出,“像基于复杂、不透明算法的谷歌地球卫星图像分析报告,往往就被认为含有相当准确的信息而具有高度的证明力。持异议的当事人不能仅仅说,自己无从获取有关的源代码,而要求法庭对该证据予以排除”。<sup>[50]</sup> 在我国,人们也不会仅仅因为不了解技术原理而否认各种基于 GPS、基站、IP 地址的轨迹分析报告的证据价值。

即便要公开机器算法,“算法透明也不能保证软件得到有效的审查”。<sup>[51]</sup> 由于部分机器算法是机器学习的产物,人类的专家并不能有效解释或说明。“源代码是冗长的,像 TrueAllele<sup>[52]</sup> 软件有 170,000 行代码。只要案件中持异议的当事人(或公众)可以自由不受限地接触、测试该软件,只要该软件客观上是扎实开发并经过标准检测的,那么获得软件源代码本身就不那么紧要了”。<sup>[53]</sup> 况且,司法人员绝大多数都是技术外行,如何进行关于数据模型与机器算法的准确裁断? 又如何进一步去评价其中哪些因素可能影响到机器结论?

笔者认为,解决大数据分析结果可靠性难题,需要法律界与科技界的协同探索。当前实践中试行的一种技术方案是,由法庭组织司法鉴定机构或行业组织的专业人员进行黑箱测试。黑箱测试是在已知产品所应具有功能的前提下,通过测试来检测每个功能是否正常的一种测试方法。在测试时,把软件程序看作一个不能打开的黑盒子,在完全不考虑其内部结构和内部特性的情况下,在软件程序接口进行测试。<sup>[54]</sup> 例如,笔者所在团队在协助办理一起组织传销犯罪案件中,为借助某个大数据分析平台对后台数据库中海量的会员数据、资金数据等进行分析,通过模拟数据反复启动对该大数据分析平台的多轮黑箱测试,抽样验证过关后再用于检验鉴定,得出了传销犯罪的组织层级、成员关系等。<sup>[55]</sup>

另一种是值得推行的法律方案,是由法庭组织控辩双方进行对抗。“就像法官判断一个新的、复杂的科学流程是否满足多伯特标准<sup>[56]</sup>,需要组织听证会听取各方专家意见一样,法官在评估一个(机器)算法的方法论与可靠性时,同样可以召集听证会听取双方专家的作证意见。”<sup>[57]</sup> 为了使得这样的庭上对抗富有效率,法庭还应给出庭发表意见的双

[49] Ric Simmons, Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System, 2016 *Mich. St. L. Rev.* 947, 951 (2016).

[50] Andrea Roth, Machine Testimony, 126 *Yale L. J.* 1972, 2030 (2017).

[51] Andrea Roth, Machine Testimony, 126 *Yale L. J.* 1972, 2034 (2017).

[52] TrueAllele 是一家公司开发的非常规 DNA 检测工具。其利用复杂的数学公式,从统计学的角度来测定案件中 DNA 样本是来自于某人的可能性。

[53] Andrea Roth, Machine Testimony, 126 *Yale L. J.* 1972, 2035 (2017).

[54] 参见田培:《软件系统的黑箱测试》,《计算机教育》2005 年第 8 期,第 61 页。

[55] 参见《××××大学物证技术鉴定中心司法鉴定意见书》××鉴定[2018]电子第 14 号。

[56] Daubert v. Merrell Dow Pharm., 509 U.S. 579, 597 (1993).

[57] Ric Simmons, Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System, 2016 *Mich. St. L. Rev.* 947, 998 (2016).

方专家提供必要的条件。具体来说,“控辩双方都可以审查可公开的机器(程序)代码、设计不同的输入指令、机器运行的不同参数,特别是测试与案件中当事人说法吻合的材料看看机器会产生什么样的报告”。<sup>[58]</sup> 这说明,面对大数据证据,司法领域的经典对抗制仍是一把金钥匙,可以帮助人们解开评价大数据证据之机器算法是否可靠的难题。

这里试举笔者调研的许某某诉淘宝(中国)软件有限公司、杭州阿里科技有限公司网络服务合同纠纷一案进行说明。为了维护淘宝客推广秩序,被告(淘宝)开发出反作弊系统,它可以通过抓取异常推广数据识别淘宝客的不正当推广行为。被告(淘宝)指称,原告的淘宝客 PID 在 2013 年 3 月、4 月、5 月的推广数据被反作弊系统抓取存在明显异常,三个月内共劫持淘宝流量 626 次、天猫流量 4952 次。法庭上,原告对被告(淘宝)通过机器算法——反作弊系统抓取淘宝客异常推广行为的做法表示强烈不服,双方产生了激烈的争议。那么,怎么处理这一争议呢? 对抗式庭审是必不可少的,专业测试也极具价值。在该案中,被告(淘宝)提供了上海××计算机司法鉴定所出具的司法鉴定意见书一份,用以证明淘宝客反作弊系统异常检测方法的科学性、合理性及劫持流量的方法、统计依据的事实。<sup>[59]</sup> 该案虽然回避了公布淘宝反作弊算法模型的问题,但同样通过庭上对抗的方式解决了大数据分析结果的真实性认定。当然,假如该鉴定所能够进行技术上的黑箱测试,则司法裁判更具有公信力。

### (三)探索超越人类经验判断的大数据证据关联性规则

关联性是证据发挥证明作用的自然属性。只有对案件事实的证实与证否具有实质意义的证据才具有关联性。大数据产生之初,早期的研究就提出,“在小数据世界中,相关关系也是有用的,但在大数据的背景下,相关关系大放异彩。”<sup>[60]</sup> 但是,基于大数据分析得出的相关性,是否可以纳入证据法的相关性概念范畴? 这是一个值得深究的问题。

一种观念认为,“大数据相关性与证据相关性在概念上相差不大,核心差异在于面向的问题域,以及是否有人参与因素”。<sup>[61]</sup> 相应地,许多人主张,大数据证据完全可以援引证据关联性的传统规则或标准进行审查。“传统的关联性判断标准还是可以用的,这个标准可以分为两个具体标准:一是指向标准,即这些数据和相关分析是不是指向案件的争点问题;二是功能标准,即这些数据和相关分析得出的结论对于证明对象有没有证明作用或者证明价值,亦即这个材料的存在使争议的问题变得更有可能还是更无可能。”<sup>[62]</sup> 传统的判断方法对大数据证据的关联性审查具有应用价值。

实践表明,部分大数据分析所揭示的关联性结论,很可能是人类依靠经验一时无法确

[58] Andrea Roth, Machine Testimony, 126 *Yale L. J.* 1972, 2028 (2017).

[59] 参见《浙江省余杭区人民法院民事判决书》(2014)杭余民初字第 3 号。在该案中,鉴定意见书的结论包括:1、系统可以找出推广数据异常的淘宝客;2、系统所使用的异常检测方法是建立在大规模数据分布式计算基础之上的,有成熟的概率学理论作为支撑,通过该方法检测异常淘宝客并判定其推广行为为不正当推广行为,是科学合理的。

[60] [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 71 页。

[61] 周蔚:《大数据在事实认定中作用机制分析》,《中国政法大学学报》2015 年第 6 期,第 82 页。

[62] 何家弘等:《大数据侦查给证据法带来的挑战》,《人民检察》2018 年第 1 期,第 57 页。

认的。例如,沃尔玛公司通过分析其历史交易记录的庞大数据库发现,每当在季节性飓风来临之前,不仅手电筒销售量增加了,而且 POP-Tarts 蛋挞(美式含糖早餐零食)的销量也增加了。<sup>[63]</sup> 英国学者维克托评价说,“认识论意义上大数据分析运用的相关性理论是超脱于人类经验判断的‘数据经验’……”<sup>[64]</sup> 我国有学者将这种现象称为产生了一种不同于人类经验的新经验。“大数据分析方法让我们看到了瞬间大批量处理非结构化信息的可能性,同时大数据分析方法能够弥补人类对庞大数据分析理解上的不足,为事实认定者提供了基于数据的‘数据经验’或者‘特殊经验’。”<sup>[65]</sup> 而这就是大数据带给证据关联性规则的关键性挑战。

化解挑战须从整体数据与具体数据两个方面去构建规则。如果整个数据集同案件中的人事物时空没有关联,那么大数据证据的关联性就没有搭建起来。<sup>[66]</sup> 至于具体数据的关联性判断,人们要学会认可机器的关联性判断,即由机器识别的关联性意见转化为人类接受和认同的因果关系认识。“传统证据的因果关系是一种强关联关系,而大数据的相关关系是一种弱关联关系。实际上,大数据的相关性是基于机器逻辑的相关关系。”<sup>[67]</sup> “在大数据时代,因果关系不仅不会退居次于相关关系的地位,而且相反,作为因果派生关系,相关关系的当然基础是因果关系。”<sup>[68]</sup>

人们对这一规律的认识需要一个过程。我国在以前治理网络犯罪,面对网络犯罪呈现的“事一机一人”特殊规律,经常遭遇无法依靠电子证据证明实际作案人的难题。从人类的因果逻辑认识上看,电子证据是由机器产生的,根据其中的信息只能追查 to 涉案机器,根本追不到作案人。但是,随着轨迹大数据分析<sup>[69]</sup> 技术出现,人们可以将得出的机器轨迹同行为人的轨迹进行匹配,实现关联性证明方面的认识跃升。如在徐某某、卢某走私、贩卖、运输、制造毒品一案中,控方举出的证据既包括几名被告人的住宿证明,也包括反映其驾驶机动车活动轨迹的智能轨迹分析材料。<sup>[70]</sup> 两者相一致,就将贩毒的犯罪行为与具体的被告人关联起来了。

## 五 余 论

“凡是过去,皆为序曲”。莎士比亚道出了历史与当下的关系。大数据证据是电子证

[63] 参见[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73页。

[64] 周蔚:《大数据在事实认定中作用机制分析》,《中国政法大学学报》2015年第6期,第70页。

[65] 周蔚:《大数据在事实认定中作用机制分析》,《中国政法大学学报》2015年第6期,第82页。

[66] 笔者调研的一起非法经营罪案件中,被告人被指控利用境外网站经营期货业务,指控证据是一份没有标明出处的期货交易清单。然而,由于有关证据不能反映出是来自于涉案网站的特定账户,以及该账户确实属于被告人实施期货交易所单独使用,亦无法排除系模拟交易服务的记录截屏,因此无法建立证据的关联性锁链。

[67] 刘品新:《电子证据的关联性》,《法学研究》2016年第6期,第189页。

[68] 王天思:《大数据中的因果关系及其哲学内涵》,《中国社会科学》2016年第5期,第37页。

[69] 大数据轨迹分析意指根据海量的话单基站数据、车辆运行数据、出入境数据、航班车辆使用数据、GPS定位数据等,确定行为人的历史活动轨迹。

[70] 参见《云南省高级人民法院二审刑事裁定书》(2015)云高刑终字第226号。

据迭代的产物。相比早期一条一条的电子数据而言,大数据证据则以数据量大为新特色,且可以隐藏其中的规律来证明案件事实。这已然是一种质的变化。现阶段通过大数据证据化解各种证明难题已初露端倪,且还会有更长远的发展。在司法实践探索的基础上,承认大数据证据的法律地位并进行证据规则建设,就成为当务之急。

与大数据证据相关的配套制度建设还会接踵而至。其一,如何对大数据进行高效率采集。具体包括梳理可适用于大数据收集、保全的法定取证措施,设计将社会大数据平台接入司法机关证据平台协助办案的机制,以及采集数据证据的规范化、法治化问题等。其二,如何开展大数据鉴定的规范化建设。具体包括为司法鉴定行业划定清晰的大数据鉴定范围,确立机器算法做出专业判断时的角色和限度,出台基于机器学习的主要司法鉴定方法等。其三,如何对大数据证据开展技术性审查。具体包括设置针对大数据证据的抽样验证标准,完善借助专家辅助人出庭质疑大数据证据的规则等。

[本文为作者主持的 2016 年度国家社会科学基金项目“大数据时代电子文件的证据规则与管理法制建设研究”(16BFX033)的研究成果。]

---

---

[ **Abstract** ] Big data is transforming the evidentiary legal system. The use of all kinds of big data as evidence in litigation is an unavoidable trend of development in the current judicial practice. There are important epistemological and methodological bases behind this phenomenon. In short, big data evidence plays the role of proof through the knowledge about certain regularity extracted from huge amount of electronic data and mainly presents itself in the form of analysis results or reports. As for the question of how to legally classify this kind of evidence, there are different theoretical and practical arguments in different countries, which mainly classify such evidence as appraisal opinion, as expert opinion, as witness testimony and so on. In view of the professional and scientific characteristics of big data evidence, the proper choice for China is to treat it as appraisal opinion. As for the question of how to review big data evidence, judges should focus on the innovative construction of authenticity and relevance rules; to ensure the truthfulness of the huge amount of electronic data itself, it is necessary to establish a rule based mainly on “big” authenticity; to ensure the truthfulness of the big data analysis results, it is necessary to construct a rule based on the reliability of machine algorithm; and to ensure the relevance of big data analysis results, it is necessary to construct a double-relevance rule for the data as a whole and for specific data, especially a specialized relevance rule that transcends human empirical judgment.

---

---

(责任编辑:郑佳)