

大数据证据民事司法应用的困境与出路

——基于176份民事判决书的实证分析

杨雅妮,石芬琴

(兰州大学法学院,甘肃兰州730000)

摘要:大数据证据是基于海量基础数据,通过算法分析技术进行“再加工”,进而挖掘数据背后规律所形成的证据。通过对176份民事判决书的分析发现,大数据证据作为数字时代的新产物,在法律地位、基础数据采集、算法分析技术审查以及大数据报告鉴定方面面临应用困境。为纾解此困境,进一步挖掘大数据证据的证明价值,应当明确大数据证据的独立法律地位,消解基础数据的采集障碍,加强对算法分析技术的审查判断,并提升大数据报告鉴定的质量。

关键词:大数据证据;民事司法;数据采集;算法分析

中图分类号:D912.29

文献标识码:A

文章编号:1008-7699(2025)01-0040-10

一、问题的提出

大数据证据是以海量电子数据凝练的规律性认识发挥证明作用的新的证据类型,主要以分析结果或报告的形式呈现。^[1]在民事司法中,大数据证据的应用不仅有利于提高诉讼效率、降低司法证明难度,而且能够推动事实认定的科学化,对于实现诉讼公正具有重要价值。2019年1月11日,司法部印发的《全面深化司法行政改革纲要(2018—2022年)》指出,要“以大数据应用和智能化建设为重点,加速推进业务的数据化、网络化、智能化,……形成‘大平台共享、大系统共治、大数据慧治’的信息化应用新格局。”随后,大数据技术不仅被广泛运用于智慧法院和智慧检务建设,而且法院、检察院等部门开始将大数据证据用作认定案件事实的依据。

伴随着大数据技术在司法领域的广泛应用,学界也对大数据证据的司法应用展开了研究,主要成果集中在两大领域。一是对大数据证据法律地位的研究。主要形成了“鉴定意见说”“电子数据说”“证人证言说”“书证说”“独立证据说”等不同观点。二是对大数据证据审查规则的研究。学者们既关注大数据证据的特殊性,也尊重证据审查规则的一般性,认为对于相关性的审查,应重点关注原始电子数据(基础数据)与案件的关联性、大数据证据与案件的关联性的统一^[2];对于真实性的审查,应从数据来源的真实性、机器算法的可信性以及分析结果的真实性和合法性等方面进行^[3];对于合法性的审查,应依据现有的证据合法性审查标准,从取证过程、取证手段和证据形式等角度进行^[4]。

以上研究,主要是对大数据证据法律地位及其审查规则的讨论,聚焦证据的内在属性及审查规则,但缺乏对大数据证据司法应用现状全面系统的分析,也未触及司法实践存在的问题。笔者以176份民事判决书为分析样本,梳理我国大数据证据司法应用现状,从大数据证据的法律地位、基础数据的采集、算法分析技术的运行以及大数据报告鉴定的质量等方面入手,对大数据证据民事司法应用的困境及其纾解路径进行研究。

收稿日期:2024-05-06

基金项目:国家社会科学基金西部项目(21XFX015)

作者简介:杨雅妮(1978—),女,甘肃宁县人,兰州大学法学院教授,博士。

二、大数据证据民事司法应用的现状

为全面掌握大数据证据的民事司法应用现状,笔者以“大数据”“数据分析”“数据比对”为关键词、以民事纠纷为案由,在威科先行·法律信息库进行全文检索,共检索出2015年1月1日至2023年11月1日的593份案例样本。除去案件当事人的工作性质涉及大数据、企业名称涉及大数据的判决、重复判决以及与本研究无关的判决,共筛选出176份有效案例(以下简称“样本案例”)。^①本部分主要以样本案例为研究对象,从大数据证据的采信情况、法院对大数据证据法律地位的认识以及法院不予采信的理由三个方面入手,对大数据证据民事司法应用的现状进行考察和类型化分析。

(一)大数据证据的采信情况

从样本案例来看,作为一种新生事物,法院对大数据证据在司法实践中的应用争议较大,该证据处于是否可以采信的“灰色地带”。在176件案件中,大数据证据被法院采信的有114件,占比65%;不予采信的有62件,占比35%,法院对应否采信该证据认识不一。

不仅如此,即使对同一个证据材料,在不同程序中,法院对其能否作为定案根据也存在较大分歧。以“广东加多宝公司与广州医药集团虚假宣传纠纷案”^②为例,在该案中,原告加多宝公司为了证明广告语“全国销量领先的红罐凉茶改名加多宝凉茶”是一种正常的商业宣传行为,在一审、二审和再审程序中都提交了《统计信息调查证明》《中国饮料行业状况分析报告》《消费品市场资讯报告》等证据材料,一审法院和二审法院均基于对法定证据类型的考虑而未予采信,而在再审环节,最高人民法院却大胆采信了该证据,并判决撤销原判。

(二)法院对大数据证据法律地位的认识

从样本案例来看,因大数据证据不属于我国《民事诉讼法》规定的八种法定证据类型,法院对大数据证据法律地位的认识并不一致,有将其视为电子数据的、书证的、鉴定意见的,还有将其视为一种单独证据类型的。

将大数据证据视为电子数据的案件最多,共有90件,占比51%。以“中国音像著作权集体管理协会与被告嘉峪关梦想王国餐饮店著作权侵权纠纷案”^③为例,该案原告为了证明被告存在侵害涉案音乐电视作品著作权的行为,通过录像公证的方式对现场点播歌曲全过程进行证据保全,并将以上视频文件复制到SD存储卡中作为电子数据在法庭上出示。

将大数据证据视为书证的案件共22件,占比13%。其中,以“新百伦公司和周某某商标权纠纷案”^④最为典型,该案原告周某某为证明被告新百伦公司的“New Balance”商标和原告的“新百伦”“百伦”商标进行捆绑销售的事实,向法庭出示了证明“新百伦”和“New Balance”二者的搜索指数曲线基本一致的浏览器搜索记录,该证据被法院认定为书证。

将大数据证据视为鉴定意见的案件共18件,占比10%。例如,在“北京菲斯特称公司与艾法史密斯菲斯特公司商标侵权纠纷案”^⑤中,原告艾法史密斯菲斯特公司为证明在被告北京菲斯特称公司成立及注册域名以前,“PFISTER 菲斯特”就已经在中国大陆以及水泥行业内拥有一定的知名度并被人们所熟知的事实,把两份鉴定机构的《检索报告》作为鉴定意见向法庭呈现。

将大数据视为单独证据的案件有46件,占比26%。在这类案件中,法官通常并不直接将该类证据表达为“大数据证据”,而是采用“情况说明”“分析报告”等表述方式,并结合其他证据一起证明案件事实。

^① 这176份样本涵盖最高人民法院以及地方三级法院的案件,具有一定的代表性。

^② 最高人民法院(2017)最高法民再151号民事判决书。

^③ 甘肃省嘉峪关市中级人民法院(2019)甘02民初150号民事判决书。

^④ 广东省高级人民法院(2015)粤高法民三终字444号民事判决书。

^⑤ 北京知识产权法院(2019)京73民终73号民事判决书。

例如,在“武汉力神动力电池系统科技有限公司、扬子江汽车集团有限公司买卖合同纠纷案”^①中,原告武汉力神动力电池系统科技有限公司为证明案涉车辆电池温差过大、均衡性很差的事实,曾委托第三方对车辆充电过程中的电池数据进行分析,并提交了《金控公司公交车辆充电大数据分析报告》,但最终该报告只是被视为一种单独的证据,未被纳入现有法定证据类型。

(三)法院不予采信的理由

在样本案例中,从不予采信大数据证据的62份判决书来看,法院不予采信的理由主要归因为大数据证据的法律地位不明确、基础数据采集不合规、算法分析技术准确性与真实性难以确认、大数据报告与案件的因果关系不强等四个方面。

一是,大数据证据的法律地位不明确。此类案件有8个,占比13%。如在“广州方达旅游管理咨询有限公司、杨某某等买卖合同纠纷案”^②中,法院认为上诉人提供的转账数据分析统计表不属于法定证据种类和形式,也无法证明待证事实,因此对该证据不予采信。

二是,基础数据采集不合规。此类案件有2个,占比3%。如在“叶某某诉陈某某等民间借贷纠纷案”^③中,被告陈某某为证明原告叶某某系职业放贷人,提交了一份“四川省民间借贷大数据报告”。对此,法院认为被告提交的大数据报告系从互联网上打印,不能证明其原始来源,采集对象的合规性难以保证。

三是,算法分析技术准确性与真实性难以确认。此类案件有32个,占比52%。在该类案件中,法院往往以算法运行的非公开性和非客观性可能导致大数据证据的真实性存疑而不予采纳。例如,在“许某某、杭州阿里妈妈软件服务有限公司网络服务合同纠纷案”^④中,被告为证明原告的推广数据存在“流量异常”现象,提交了一份《推广行为分析数据》,对此,法院以该大数据分析结果所依据的算法准确性及真实性均难以确认为由而未予采信。

四是,大数据报告与案件的因果关系较弱。此类案件有20个,占比32%。在这类案件中,法院主要以大数据报告与待证事实的因果关系不强为由不予采纳该类证据。例如,在“合肥新祈源商贸有限公司、深圳市力族电子有限公司侵害外观设计专利权纠纷案”^⑤中,法院认为原告提交的大数据报告鉴定结果既未能证明该公司耳机为本案的被告侵权产品,也未能证明该公司的耳机毛利率相同或者基本相同,即未能证明两者之间的对应性和相关性,而不予采信该证据。

三、大数据证据民事司法应用的困境

综上,虽然大数据证据已在民事司法中得以应用,但仍存在大数据证据的法律地位不明确、基础数据采集存在障碍、算法分析技术的运行风险难以规避以及大数据报告鉴定的质量难以保证等困境。

(一)大数据证据的法律地位不明确

前已述及,对于大数据证据的法律地位,学界认识并不一致,主要形成了“电子数据说”“书证说”“鉴定意见说”以及“独立证据说”四种观点。

持“电子数据说”的学者认为,大数据证据与电子数据都是基于电子化技术手段形成的客观资料,二者在形式与结构上都具有高度一致性。大数据证据是算法衍生的一系列分析数据,其最终表现形式仍然为电子数据^[5],应将大数据证据视为电子数据。以该观点为依据,只要是具备客观真实性的大数据,都应当属于“电子数据”的范畴。

① 湖北省武汉市中级人民法院(2017)鄂01民初4189号民事判决书。

② 广东省广州市中级人民法院(2022)粤01民终11640号民事判决书。

③ 四川省德阳市绵竹市人民法院(2019)川0683民初543号民事判决书。

④ 杭州互联网法院(2020)浙0192民初3081号民事判决书。

⑤ 广东省高级人民法院(2019)粤民终1841号民事判决书。

持“书证说”的学者认为,大数据的规律性是客观的,常常通过算法来反映,大数据与书证具有相同的属性,在司法实践中,大数据分析报告最终也是以书面材料形式出示的,因此,大数据证据是一种特殊类型的书证。^[6]该观点为大数据证据在法庭上的合法应用提供了依据,法院在司法过程中应用大数据证据时,应当遵循书证的有关规则。

持“鉴定意见说”的学者认为,大数据证据是鉴定意见的一种特殊形式,应将大数据分析纳入司法鉴定的范畴^[7],并修改相关规范以便于大数据建模等新兴技术得以合规合法地进入司法鉴定领域并体现价值^[8]。主要理由在于:一方面,大数据证据与鉴定意见都具有专业性和科学性,两者均以专业的分析结论来认定案件事实,因此在内容上具有一定的相似性;另一方面,我国有关鉴定意见的法律规范已经比较成熟,对于鉴定标准、鉴定机构、鉴定程序等都有非常明确的规定,在这种背景下,将大数据证据视为鉴定意见的一种特殊形式,并直接适用鉴定意见的审查规则,有利于提高民事司法的效率。

持“独立证据说”的学者认为,从本体与结构、法律性质以及种类归属三个角度出发,唯有承认大数据分析报告的证据属性并将其列为独立的证据类型,才能充分挖掘大数据分析报告的价值。^[9]理由在于:一方面,与其他法定类型的证据相比,大数据证据具有显著的独特性,无法将其纳入现有的法定证据种类;另一方面,大数据证据能够提高诉讼效率、降低司法证明的难度,推动事实认定的科学化,将其作为一种独立证据种类有利于实现司法证明的目的。

除以上外,国外学者还提出了“科学证据说”^①“机器证言说”^②等观点,这些观点均认可大数据用作证据的可行性。

(二)基础数据采集存在障碍

作为大数据证据的组成部分,基础数据的真实性与精确性直接关系到大数据分析报告的客观真实性。基础数据的质和量达不到要求,会影响到法院对大数据证据的效力认定。从样本案例来看,基础数据采集的障碍主要体现在以下三个方面。

1. 数据采集主体的力量单薄

在基础数据采集过程中,显著的问题是采集主体力量单薄。一方面,数据采集主体在取证能力上往往处于相对弱势的地位。多数情况下,这些数据存储在公权力主体自有的数据库或者由第三方数据存储平台管理,普通民众在访问这些数据时往往面临着烦琐的流程和信息不对等难题,他们通常只能依赖有限的渠道获取基础数据。另一方面,数据的提取和算法的设计需要专业知识和辅助设备,这不仅增加了数据采集的技术门槛,也使得个人在提取相关数据时面临着巨大的能力障碍。尤其对于普通民众而言,很难拥有满足数据采集要求的专业能力和先进设备。因此,基础数据的采集主体往往面临着难以获取数据的困境。

2. 数据采集过程存在合规风险

在信息化大背景下,个人信息、隐私、商业秘密等具有秘密性特征的数据极易暴露在大众视野中。在基础数据采集过程中,采集主体又不可避免地会接触到大量上述秘密信息。这些信息在未获得授权的情况下被任意地收集、整合时,可能会导致对他人个人信息、隐私以及商业秘密的侵犯,使得基础数据的采集过程面临合规风险。一旦采集主体未经授权进行基础数据采集,不仅法院会以该证据不具有合法性为由不予采信,而且主管部门也会依法追究采集主体的法律责任。

3. 数据筛选流程可能出现“数据偏见”

^① 即未设置单独的大数据证据类型,而是将这些需要运用科学技术、科学仪器等方法才能解读,并能证明案件事实的证据统称为科学证据,并通过长期的司法实践不断完善科学证据的采信规则。参见张吉豫. 大数据时代中国司法面临的主要挑战与机遇——兼论大数据时代司法对法学研究及人才培养的需求[J]. 法制与社会发展, 2016(6): 52-61.

^② 该观点认为,作为一种新的技术,机器算法在找寻、分析电子数据发挥着无可替代的作用,大数据分析材料反映出的案件信息是一种“机器证言”。参见 ROTH A. Machine Testimony[J]. Yale Law Journal, 2017(1): 1972-2053.

在基础数据采集过程中,当缺乏明确的法律依据和技术规范指引时,采集主体可能以自身利益或预期目标为标准对数据进行有意或无意的筛选,从而引发了“数据偏见”现象。具体而言,由于缺少明确的法律依据和技术标准,采集主体在面对海量数据时,容易在数据筛选过程中受到主观因素的影响,选择性地采集对己方有利或符合特定期望的数据,而非客观、全面地获取数据。该情况会直接影响到基础数据的客观性,削弱大数据证据的可信度和证明力。

(三)算法分析技术的运行风险难以规避

大数据分析技术是通过基础数据的筛选、清洗以及分析,实现数据可视化的一种分析技术,其技术支撑点在于模型的建立,而算法本身又具有黑箱化运行、非公开性和非客观性等缺陷,这些缺陷使得算法是行风险难以规避,进而直接影响到大数据证据的真实性、客观性。

1. 算法运行的非公开性

算法分析技术具有黑箱化运行的特点,其非公开性导致大数据证据的真实性遭到质疑。一方面,对于非专业人员来说,算法相当于一个“黑箱”,他们无法了解结果的生成过程,无法完全理解和认可基于数据挖掘、算法逻辑而产生的机器证言,因此容易对算法的可靠性产生质疑。另一方面,算法具有保密性,一般情况下不予以公开。这是因为,算法是研发人员的智力成果,属于知识产权保护的范畴,具有财产权益的特征,有些算法甚至与国家、社会公共利益密切相关,如果强行公开会危及公共安全。因此,基于对创造性成果的尊重和保护以及对公共利益的维护,算法通常不会被公开,这种非公开性特征使得当事人难以对算法分析过程进行充分的质证,从而会影响到法院对大数据证据的采信。

2. 算法运行的非客观性

算法,容易受到研发者个人偏好的影响,无法做到完全的客观中立^[10],其主观性和非中立性使得大数据证据的真实性难以保证。一方面,由于每个研发主体的技术、知识、经验以及价值观念存在差别,所研发出的算法或多或少带有主观色彩,难以保证算法运行的客观性;另一方面,研发主体会将某些偏好(如性别、地域等)带入算法程序,使得经过算法分析得出的数据缺乏中立性,难以客观反映实际情况。

(四)大数据报告鉴定的质量难以保证

大数据证据的形成是学科交融的结果,不仅涉及统计学、数学以及计算机科学等知识的应用,而且涉及算法和程序的运行,具有专业性和指向性。^[11]大数据证据司法应用的特点之一是将数十页甚至上百页的证据材料浓缩为一份几页纸的分析报告。作为大数据证据的“最终呈现形式”,大数据报告不仅需要准确反映数据分析的结果,还必须经过严格的鉴定,以确保其准确性和科学性,因此,对大数据报告进行鉴定是极其重要的环节。

近年来,随着大数据和人工智能技术的不断发展,大数据证据的表现形式更为多样,内容也更为丰富,对鉴定设备、鉴定技术以及鉴定人员都提出了更高的要求,要求不仅具备更强的处理能力,还要能够适应相关技术发展带来的变化。与此同时,鉴定人员必须具备深厚的跨学科知识,能够理解并评估复杂的数据处理过程。但遗憾的是,我国不仅缺乏专门的大数据技术的鉴定机构,而且现有鉴定机构在对大数据报告鉴定时,所采取的方法、适用的标准也不统一,现有鉴定技术无法满足大数据证据鉴定的需要,大数据鉴定报告质量难以保证。

四、大数据证据民事司法应用困境的纾解

针对前述困境,应从明确大数据证据的法律地位、消解基础数据采集的障碍、加强算法分析技术的审查判断以及提升大数据鉴定报告质量四个方面入手,对大数据证据民事司法应用的困境进行纾解。

(一)明确大数据证据的独立法律地位

针对当前大数据证据法律地位不明确的问题,目前最合理的做法是把大数据证据定性为一种独立的证据类型,原因如下。

1. 大数据证据难以归入现有证据种类

(1) 大数据证据不是电子数据

电子数据属于“原生证据”,而大数据证据属于“加工证据”,电子数据无法完全囊括大数据证据,两者存在以下差异。其一,两者的分析工具不同。大数据证据基础数据总量大、种类多,数据非结构性明显,并且其价值密度低,无法直接用于证明,需要运用通过算法分析形成的大数据分析报告,因此大数据证据的意思表达高度依赖算法;而电子数据的形成并非一定依据算法分析技术。其二,两者的分析对象不同。大数据证据分析复杂、非结构性的数据,而电子数据分析简单、结构性的数据。其三,两者证明侧重点不同。大数据证据侧重证明数据与案件之间的相关关系,而电子数据侧重证明数据与待证事实之间的因果关系。综上,二者在分析工具、分析对象及证明侧重点方面均存在差异,不宜将大数据证据归入电子数据的范畴。

(2) 大数据证据不是书证

将大数据证据认定为书证,也具有一定的局限性。一方面,两者的呈现方式不同。书证主要是通过文字、符号、图形等方式呈现结论,而大数据证据的呈现方式不仅限于文字、图像、图标、柱状图等可视化方式,还包括视频、音频等电子化方式。另一方面,两者的形成时间不同。书证是在案件发生过程中形成的证据,而大数据证据是对案件发生过程中的数据进行后期分析、处理、加工后形成的证据。综上,二者在呈现方式和形成时间上均存在差异,将大数据证据认定为书证过于局限,不利于发挥大数据证据的特殊价值。

(3) 大数据证据不是鉴定意见

大数据证据与鉴定意见的形成虽然都必须经历一个判断过程,但二者之间的差异显著。一方面,二者的形成过程不同。鉴定意见是鉴定人依据其专业知识作出的专业判断,机器在鉴定意见形成过程中发挥的主要是工具性作用;而大数据证据的形成则主要依赖于机器的判断,基本不需要专家参与。另一方面,二者所依据的工具不同。鉴定意见主要是借助科学仪器以及科学知识对鉴定材料的真实性进行判断;大数据证据则主要依赖算法模型,对基础数据所呈现的规律性进行判断。综上,二者存在本质差异,不宜将大数据证据归入鉴定意见的范畴。

2. 大数据证据是一种独立的证据类型

基于前述分析,应将大数据证据界定为一种独立的证据类型,理由如下。

(1) 大数据证据的内涵具有独特性

对于大数据证据内涵的界定,学界主要有三种思路:一是将大数据证据仅仅归纳为大数据分析报告^[1];二是强调大数据技术的作用,混淆大数据证据和大数据技术的边界^[7];三是认为大数据证据是一种可以产生某种实用意义上增值价值的方法^[12]。以上三种思路均存在缺陷,是对大数据证据内涵的误识。其一,大数据分析报告无法完全囊括大数据证据,分析报告更倾向于一种结果,但结果是否完整和准确,还需要对形成过程进行审查,若缺乏对基础数据和算法技术的审查则易引发算法的黑箱化运行风险;其二,运用算法分析技术收集的证据不同于大数据证据,前者更加强调算法的工具性作用,最终算法不用于法庭的举证和质证;而后者强调证据的收集必须经过算法模型处理这一“二次加工”,算法技术仍属于大数据证据的外延;其三,大数据证据不应界定为一种“方法”,而是应作为认定事实的依据,“方法”对应的实为大数据证明而非大数据证据。

因此,厘清大数据证据的内涵需要把握以下三个重点。第一,大数据证据的基础数据具有海量性。大数据证据的基础数据是由海量的、无序的电子数据所组成,其规模大且种类丰富,远远超出传统证据所包含的数据体量和类别。^[13]第二,大数据证据对算法具有依赖性。因大数据证据信息量巨大,需要利用如数据挖掘、数据分析等算法对基础数据进行收集、清洗和整理,帮助司法主体在大量信息中迅速找到与案件相关的关键信息。第三,大数据分析报告与案件具有相关性。区别于传统证据更追求与待证事实的

因果关系,大数据更擅长进行相关性分析^[14],大数据分析报告反映数据与待证事实间的量的关系,从而定性把握因果关系分析。

综上,与传统证据相比,大数据证据具有独特的内涵。应当采用整体性思维对大数据证据的内涵进行界定,将其界定为基于海量基础数据,通过算法对数据进行汇总、筛选、建模分析等再加工,挖掘出数据背后的规律,进而形成的一份可视化分析报告。大数据证据由基础数据、大数据算法技术以及大数据分析报告构成,三者之间相互区别且相互关联,具有明显的“复合型”特征,若将其纳入任何一种法定的证据种类,都失之偏颇。

(2)大数据证据的性质具有独特性

在性质上,不同于传统证据,大数据证据更具有技术性、思维性、弱关联性和预测性。其一,大数据证据具有技术性。大数据证据以数据和算法为核心,分析报告的解读需要非常高的专业知识和技术能力,非专业人士无法进行。其二,大数据证据具有思维性。在大数据证据生成及证明作用发挥过程中,不仅需要融合诉讼认知主体的能动性,还要依托算法、人工智能技术等挖掘海量数据与案件事实之间的客观性、逻辑性关系,这一过程是科学性和思维性的统一。其三,大数据证据具有弱关联性。大数据证据体现的是相关关系而非因果关系,与待证事实之间具有弱关联性。其四,大数据证据具有预测性。大数据证据的独特价值在于通过大数据报告能够总结出事物发展的规律,可预测事物的未来发展。

(3)大数据证据的收集方法具有独特性

与传统证据的收集方法不同,大数据证据的收集具有独特性。一方面,大数据证据的原始数据具有海量性和种类多的特点;另一方面,大数据证据的关联性分析更多体现为相关性,并不过度关注与案件之间的因果关系。基于此,在原始数据收集过程中,应突破传统证据强相关性甚至因果性的证据聚焦范围以及传统七要素收集方法的限制,尽可能多地吸纳相关数据和模糊数据,提高大数据证据的可靠性和价值。

(4)大数据证据的证据调查方式具有独特性

在大陆法系国家或地区民事诉讼法中关于证据的规定中,法定证据类型是以证据调查形式为判断标准,而非以证据方法或证据材料为表征。^[15]具言之,书证是对文书的内容进行调查,人证是对其所陈述的证言进行调查,鉴定是对鉴定人所陈述的判断意见进行调查,勘验是对物体的性质和外状进行调查等。而对于大数据证据,法官不仅需要调查最终呈现的大数据分析报告的内容,还要求证据提供方提交采集的基础数据、运用的算法等分析技术,必要时可能对算法运行过程进行当庭测验。基于此,在证据调查方式方面,大数据证据与书证、鉴定意见等证据的取得存在显著不同,现有的证据类型无法代替或完全涵盖大数据证据。

(二)消解基础数据采集存在的障碍

数据是与土地、劳动力、资本、技术并列的五大生产要素之一,是一种基础、战略性资源。^①针对前述基础数据采集方面存在的障碍,应从以下方面入手,消解基础数据采集障碍。

1. 提升采集主体的数据采集力量

在基础数据采集过程中,应采取以下措施提升采集主体的采集力量:其一,赋予采集主体对公权力主体自有数据库的访问权,促进数据的合理利用。数据访问权也被部分学者称为“数字开示”,即在合理和正当的情况下,当事人有权经申请获得与其案件相关的基础数据。^[16]该权利的赋予有助于采集主体更全面地了解数据,为大数据证据的形成提供全面可靠的基础数据支持。其二,加强对第三方数据存储平台维护数据的管理。应建立更为规范、透明的数据管理机制,通过对数据存储、更新、共享等方面的精细管理,

^① 中共中央、国务院《关于构建更加完善的要素市场化配置体制机制的意见》提出:推进土地要素市场化配置、引导劳动力要素合理畅通有序流动、推进资本要素市场化配置、加快发展技术要素市场、加快培育数据要素市场,数据已成为五大生产要素之一。

确保基础数据的完整性和可信度,确保采集主体能够有效利用第三方数据存储平台维护的数据。

2. 保障数据采集过程的合规性

个人信息的公共属性为大数据范围的选取提供了理论基础。^[17]针对基础数据采集过程存在的合规风险,应采取以下措施。一方面,对涉及个人信息的数据进行分类分级保护。依据数据的来源、重要性以及用途为标准对数据进行分类,并通过风险评估对涉及公共安全以及个人隐私等关键数据进行重点标识,从而实现对数据的限制使用和特别规制。另一方面,应进一步完善数据保护的配套措施,如对具有识别性的个人信息收集,必须具有目的上的正当性和合法性;对敏感个人信息收集,需要取得相关授权,以“知情+同意”为原则;对高风险数据,需要采取数据脱敏等技术,实现对数据所含隐私信息的保护。

3. 规范数据筛选的流程

在数据筛选流程方面,应明确数据筛选的原则,并建立统一的规范,以确保数据筛选的合法性、准确性与及时性。其一,数据筛选的原则至关重要,数据筛选必须遵循三大原则,即及时性、准确性和合法性。及时性原则是确保数据筛选能够在合适的时间内完成,以反映当前真实的情况;准确性原则要求筛选的数据具有高度的符合性,以避免信息误导和错误;合法性原则是确保数据筛选过程符合法律规定,目的是保障数据的合法获取和使用。其二,为了确保数据筛选的规范性,应制定具体的数据筛选标准和技术规范,将标准量化,并将其转化为法律问题,从而更有效地确保数据筛选的真实性和可靠性。需要注意的是,在数据筛选和结果审查时,不仅要进行形式上的审查,如是否明确了筛选方法、分析过程等信息,以及是否有数据筛选主体的签字盖章等,还应进行实质审查,如筛选过程和方法是否符合相关专业的规范要求等。此外,还应注意数据筛选流程的公开性,透明的筛选流程不仅能够使数据采集过程经得起法律的审查,也有助于提高大数据证据的可信度。

(三) 加强算法分析技术的审查判断

2021年11月出台的《互联网信息服务算法推荐管理规定》虽然对算法滥用、算法歧视等问题进行了规制,但我国对于算法的技术性审查仍处于起步阶段。针对前述算法分析技术存在的运行风险,应从以下方面加强对算法分析技术的审查判断。

1. 强化算法的验证和源代码的审查

一方面,可以通过“黑箱测试”的方式来验证算法的稳健性,同时可以结合白箱测试、灰箱测试等多种测试手段,以确保测试结果的准确性和可信度;另一方面,在源代码的审查中,美国的多伯特案件^①提供了相应的标准,以此为据,应重点审查是否存在科学性错误代码和技术性错误代码^[18],具体通过重复性测试、同行评议、公开错误率以及相关领域普遍接受度等方式进行审查。除以上外,还可在前端对算法进行动态评估和监督,并由国家互联网管理部门牵头进行算法影响评估服务体系建设,支持开展影响评估和认证服务。^[19]

2. 借助区块链技术对算法可靠性进行审查

区块链技术具有去中心化的特点,能使得上链的基础数据具有不可篡改性和安全稳定性,其主要通过哈希值加密的方式实现。在借助区块链技术对算法可靠性进行审查时,可将大数据证据的基础数据通过哈希值进行加密处理,并将哈希值储存在区块链上,以保证基础数据的真实性。同时,可通过区块链技术建立数据的溯源机制,将数据的来源、采集、处理等流程信息进行记录,实现数据流通全流程的可验证性和可追溯性。

3. 利用专家辅助人出庭协助质证

由于算法分析技术的专业化程度高,往往导致证据质证难以进行,因此,在涉及算法分析技术的案件

^① 美国最高法院在1993年作出的裁决,确立了联邦法院在审查科学证据时应采用的标准,即法官应充当“守门人”角色,判断证据是否具有科学可靠性和相关性。参见肯尼斯·R·福斯特,彼得·W·休伯,对科学证据的认定[M].王增森,译.北京:法律出版社,2001:6.

中,通常需要专家辅助人出庭协助质证。要有效发挥专家辅助人的作用,应确立以下规则:一是当事人应聘请熟悉算法技术的专业人员作为专家辅助人,例如计算机工程师等等;二是在质证过程中,专家辅助人仅能对案件相关的源代码发表意见,对于非开源的源代码负有保密义务;三是专家辅助人应当对算法的训练集,模型以及算法进行审查,并对形成的报告中存在的风险进行说明。

(四)提升大数据报告鉴定的质量

针对前述大数据报告鉴定质量存在的问题,应通过培养专业的大数据报告鉴定队伍、制定具体的大数据报告鉴定标准、配备先进的鉴定设备等措施,提高大数据报告鉴定的专业性水平,进而保证大数据报告鉴定的质量。

1. 培养专业的大数据报告鉴定队伍

大数据证据具有跨学科性和专业性,既涉及数学、统计学、计算机科学等多个学科,也涉及算法-人工智能等专业技术的运用,因此,高质量大数据报告的鉴定离不开一支专业的鉴定队伍。为此,一方面,应对大数据报告鉴定人员的专业背景和知识结构作出明确规定,将跨学科的、懂技术的专业人员吸纳进鉴定人员队伍;另一方面,针对现有鉴定人员在大数据报告鉴定方面存在的不足,可通过各种形式的在职培训,帮助其了解和掌握大数据技术,提高其对大数据证据的鉴定能力。

2. 制定具体的大数据报告鉴定标准

对大数据报告鉴定时,需要考虑不同类型大数据证据的特殊性,如结构化数据、非结构化数据等,制定具体的大数据报告鉴定标准。此过程中,应遵循技术自治理念,一方面,通过事前报备的方式赋予鉴定机构自行制定有关技术规范的权利,对于已经制定的技术规范,应进行事后的监督和评估;另一方面,加快行业规范和国家规范的制定步伐,针对不同类型的大数据证据,分别制定具体的鉴定标准,为大数据报告鉴定提供专业指导。

3. 配备先进的大数据报告鉴定设备

为保证大数据报告鉴定的顺利进行,大数据报告鉴定机构还应配备先进的数据分析设备、安全可靠的数据存储系统等设备。一方面,大数据报告鉴定机构应当配备高性能计算机硬件平台和先进的数据分析软件,并根据需要及时对相关设备进行升级和扩展;另一方面,大数据报告鉴定机构应当配备安全可靠的数据存储系统,既要存储容量大、稳定性强,又要加强对数据的保护,利用强大的加密技术,确保数据的存储安全。

五、结语

综上所述,大数据证据作为一种新型证据类型,在法律地位、基础数据采集、算法分析技术审查及大数据报告鉴定方面仍存在应用困境,亟需进一步优化。在大数据时代,面对实践中出现的大数据证据,我国的民事证据制度也应作出积极回应,探索和思考其在民事案件中的应用,尤其是在突破证据法定形式主义原则、放弃将证据种类作为证据门槛的限制、构建更加灵活的证据采纳标准等方面开展研究。^[20]未来,深化大数据技术的研究与开发,进一步加强法官和诉讼代理人的大数据技术培训,进而推动大数据证据在民事司法中的广泛应用,助力证据法在新时代实现创新与发展。

参考文献:

- [1] 刘品新.论大数据证据[J].环球法律评论,2019(1):21-34.
- [2] 张吉喜,孔德伦.论刑事诉讼中的大数据证据[J].贵州大学学报(社会科学版),2020(4):82-88.
- [3] 吴才毓.大数据公共安全治理的法治化路径:算法伦理、数据隐私及大数据证据规则[J].政法学刊,2020(5):33-41.
- [4] 张全涛.从实践应用到证据法定:大数据证据的理论证成与规则建构[J].四川轻化工大学学报(社会科学版),2023(3):12-22.

- [5] 罗文华. 大数据证据之实践与思考[J]. 中国刑事警察, 2019(5):18-20.
- [6] 胡铭, 龚中航. 大数据侦查的基本定位与法律规制[J]. 浙江社会科学, 2019(12):12-20+155.
- [7] 何家弘, 邓昌智, 张桂勇, 等. 大数据侦查给证据法带来的挑战[J]. 人民检察, 2018(1):54-57.
- [8] 谢君泽. 论大数据证明[J]. 中国刑事法杂志, 2020(2):125-137.
- [9] 周慕涵. 论大数据证据的法律地位[J]. 法律科学(西北政法大学学报), 2023(4):104-114.
- [10] 杨子飞. “第三洞穴”与“数据主义”——论大数据社会科学的内在矛盾[J]. 自然辩证法研究, 2016(8):63-67.
- [11] 维克托·迈尔·舍恩伯格, 周涛. 大数据时代生活、工作与思维的大变革[J]. 人力资源管理, 2013(3):174.
- [12] 谢君泽. 论大数据证明[J]. 中国刑事法杂志, 2020(2):125-137.
- [13] 董少平, 张则超. 大数据证据的证据属性及种类认定[J]. 铁道警察学院学报, 2022(4):91-97.
- [14] 黄健. 刑事司法证明中大数据相关关系的局限作用论[J]. 清华法学, 2023(2):22-39.
- [15] 占善刚. 证据法定与法定证据——兼对我国《民事诉讼法》第63条之检讨[J]. 法律科学(西北政法大学学报), 2010(1):137-142.
- [16] 裴炜. 论刑事数字辩护:以有效辩护为视角[J]. 法治现代化研究, 2020(4):124-142.
- [17] 李晓东. 大数据证据规则之体系构建研究[J]. 政法学刊, 2022(4):122-128.
- [18] 弗朗西斯卡·帕米奥托. 审判黑箱:算法不透明性对刑事司法公正审判权的影响[C]//马丁·艾波斯, 玛尔塔·坎泰罗·伽米托. 算法治理——法律和道德挑战. 姚前, 冯蕾, 译. 北京:中国金融出版社, 2022:55.
- [19] 张欣. 算法影响评估制度的构建机理与中国方案[J]. 法商研究, 2021(2):102-115.
- [20] 郑飞. 证据种类法定主义的反思与重构[J]. 中国法学, 2024(1):105-123.

Dilemmas and Solutions to the Application of Big Data Evidence in Civil Judicial Cases: An Empirical Analysis Based on 176 Civil Judgments

YANG Yani, SHI Fenqin

(Law School, Lanzhou University, Lanzhou 730000, China)

Abstract: Big data evidence refers to a visual report that is formed based on massive basic data, utilizing algorithmic analysis technology for “re-processing”, in order to mine the laws inherent within the data. An analysis of 176 civil judgments reveals that, although big data evidence has been applied in some civil cases, it faces several application challenges, notably the ambiguity of its legal status, obstacles in basic data collection, and the intricacies of mitigating operational risks associated with algorithmic analysis while ensuring the quality of the resulting big data report. In order to further explore the value of judicial application of big data evidence, it is expected to address these dilemmas. This involves acknowledging the independent legal status of big data evidence, overcoming barriers to the collection of basic data, bolstering the scrutiny and evaluation of algorithmic analysis technology, and enhancing the overall quality of the big data reports produced. By tackling these issues head-on, we can further explore the invaluable contributions that big data evidence can make to the judicial system.

Key words: big data evidence; civil justice; data collection; algorithmic analysis

(责任编辑:董兴佩)