大数据时代个人信息的运作模式、理论困境及保护路径

薛悟娟

(中国政法大学 政治与公共管理学院,北京 100088)

摘要:大数据技术的快速发展和不断应用使个人信息能够被大规模收集以及重复循环再利用,为个人信息的运作模式带来了颠覆性的转变。在个人信息累积效应下,大数据能够推衍、萃取、洞见甚至连结出不为人知或不欲为人知的信息,其不当应用将产生侵害个人信息自决权和隐私权的风险,同时也可能与《个人信息保护法》所规定的保护原则形成冲突。为兼顾大数据发展与隐私保护义务,应当制定个人信息开放规范以促进大数据发展;采取牢固的匿名化技术以实现个人信息的"去连结性";将个人信息保护影响评估作为保护个人信息的有效工具;以"实质性参与"作为告知同意原则完善的路径取向。

关键词:个人信息;大数据;隐私权;《个人信息保护法》

中图分类号:D923 **文献标志码**:A **文章编号**:2096-028X(2024)02-0103-10

大数据,又称为"巨量资料",通过大规模收集、储存和分析资料,藉由复杂的算法或人工智能技术,以揭露其他方面尚未被确知的模式、连结、行为、趋势、身份与实用知识。① 大数据能快速收集和储存资料,预测、洞见未来趋势,为政府行政管理和治理提供工具,为社会发展预防风险,为国家政策的制定与执行提供参考依据。然而,以巨量资料为基础的大数据,凭借强大的算法撷取、推衍出尚不为人知或不欲为人知的个人信息以及个人隐私,对个人信息的保护形成不可预知的风险和挑战。如何在大数据时代开发利用个人信息的同时保障个人信息安全,已然成为时代难题。

为平衡大数据时代下个人信息的保护和利用,2020年5月28日颁布的《中华人民共和国民法典》(简称《民法典》)第四编第六章对个人信息保护作出了基础性规定。② 随后,2021年8月20日《中华人民共和国个人信息保护法》(简称《个人信息保护法》)出台,标志着中国个人信息保护制度的法律框架体系初步形成。然而,相继出台与实施的法律规范也无法完全避免大数据时代下个人信息遭受侵害。大数据的不当应用给《个人信息保护法》中规定的个人信息权益以及保护原则带来了挑战。鉴于此,通过分析大数据对个人信息的收集与处理模式,结合《个人信息保护法》的若干条款,指出个人信息保护面临的理论困境,探索大数据时代中国个人信息的保护路径。

一、大数据对个人信息的收集与处理模式

大数据的运作模式是对个人信息的大规模收集和重复循环再利用,通过大数据分析工具的扩散,能够轻松地跟踪、量化和交换人与人之间的信息,解开人类基因奥秘,解决城市生活问题,揭示出社会关系和文化偏好背后的隐藏模式。③ 大数据是个人信息的运作模式,个人信息也构成了大数据的实质内容。

(一)个人信息的大规模收集

大数据是一个抽象的概念,目前尚未有统一精确的定义,基于大数据覆盖范围的广泛性和类型的多样

收稿日期:2023-09-09

基金项目:2021年度国家社科基金一般项目"行政法上第三人的权利保护"(21BFX050)

作者简介:薛悟娟,女,法学博士,中国政法大学政治与公共管理学院博士后。

① Anita L. Allen, *Protecting One's Own Privacy in a Big Data Economy*, Harvard Law Review, Vol.130:71, p.71(2016).

② 参见赵宏:《〈民法典〉时代个人信息权的国家保护义务》,载《经贸法律评论》2021年第1期,第1-2页。

③ Karen E.C. Levy, *Relational Big Data*, Stanford Law Review, Vol.66:73, p.73(2013).

性,人们倾向于以“数据的数量”与“管理这些数据的能力”来定义大数据。[①] 大数据中含有个人信息和非个人信息。个人信息的含义体现在《个人信息保护法》第4条和《民法典》第1034条第2款,即“以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息”,或“能够单独或者与其他信息结合识别特定自然人的各种信息”,据此,“识别性”是识别个人信息的主要标准,也就是说,具有识别性的自然人信息,无论其以何种方式记录,都可被认定为个人信息。

大数据能够快速收集与储存各种来源的个人信息,其范围并不局限于互联网上的资料,还包括传统的数据集,部署在基础设施(例如通信网络、电网、全球定位卫星、道路和桥梁等)中的传感器收集的信息,[②]可见,大数据对个人信息的大规模收集已经远超于传统软件工具对个人信息的收集、储存,致使“所有的个人事务和行动都变成了数据,由网络数据完整描述个人人格”。[③] 同时,大数据通过算法等技术使得个人被信息化。算法作为一种自动分析工具,具有决策属性,在大数据的基础上进行快速、高效、精准的自动化决策。自动化决策对数据人格的塑造并非简单停留在数据记录或数据整合方面,其会更加深入地挖掘个人偏好、隐私、行为习惯等,这也不可避免地造成大数据对个人信息的过度收集或非法收集。例如,商家进行广告推送、精准营销、偏好记录以及再次营销。再如,不法分子利用个人信息谋求利益,甚至实施大型犯罪活动。在上述过程中,个人信息展现出巨大价值,收集数据能够获得非常强大的经济激励,以至于个人信息的收集变成了机会主义,而非是有目的的收集。换言之,只要有可能、有机会,数据就会被收集,即使没有具体的使用目的,也要进行收集。同样,个人信息的保留也具有了经济动机。个人信息会被尽可能地保存较长的时间,甚至远远超过最初的使用时间,从而反复使用。此外,大数据技术的应用并未形成明确的行业标准和统一的约束性规则,关于数据的采集、管理、共享、交易等缺乏技术标准予以规范,导致个人信息在被大数据收集的过程中面临着隐私受到侵犯的风险。

(二)个人信息的重复循环再利用

传统个人信息的处理分析方式如下:基于人们提出的问题作出假设,收集个人信息进行分析并得出结论。其后,个人信息处理者[④]为了确保传统软件工具正常地、不超负荷地运转,对所收集的个人信息作出甄别并进行选择性剔除,仅对至关重要的个人信息予以保留。大数据时代,个人信息的处理分析发生了颠覆性转变,信息的存储已经不再是需要考量的事项,个人信息能够被全面完整地存储。信息处理技术也取得了重大进展,不需再考虑个人信息是否超过负荷或是否予以保留。例如,批处理技术能够将大量数据集中处理,实现数据的有效管理和信息的快速获取;流处理技术能够实时处理数据流、及时监测数据并获取数据的实时反馈,常用于实时分析和实时计算等场景;NoSQL技术(Not only SQL 非关系型数据库)能够在丰富的数据模型中支持高并发查询、数据分片处理等需求,提高了数据的扩展性和灵活性;数据挖掘技术能够通过使用数据挖掘算法,重复循环利用个人信息,发现潜在的趋势和模式,并预测未来趋势和方向。

过去,数据一般被基于特定的目的收集和一次性利用。在大数据时代,数据的潜在价值在收集时可能是不明确的,只有当数据基于相同或者不同的目的被重复利用时,才可能完全呈现出潜在的价值。[⑤] 数据价值不仅在于可以从中分析出更多的个人信息,更重要的是,通过与其他数据源相结合的方式,个人信息的利用得到极大提升。全面完整的数据不仅可以用来回答或者分析当前具体问题,还能激发新问题的提出,脱离结构化的数据库,在个人信息之间建立关联性,从而挖掘出隐匿的个人信息,甚至是个人隐私。同时,数据收集者能够实现个人信息的预期利用和非预期利用,甚至会将个人信息流转至第三方手中,脱离最初收集的目的进行再利用。例如,出行网站挖掘出旅客出行规律并共享给航空公司,平台将某群体的消费习惯转卖给其他商家等。而作为接收个人信息的第三方根据自己的需求继续交叉比对、重复使用和加工个人信息,实现个人

---

① Rick Swedloff, *Risk Classification's Big Data (R)evolution*, Connecticut Insurance Law Journal, Vol.21:339, p.339(2014).

② Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, Northwestern Journal of Technology and Intellectual Property, Vol.11:237, p.239(2013).

③ 参见王秀哲:《大数据时代个人信息法律保护制度之重构》,载《法学论坛》2018年第6期,第115-116页。

④ 《个人信息保护法》第4条第2款规定:“个人信息的处理包括个人信息的收集、存储、使用、加工、传输、提供、公开、删除等。”据此,个人信息处理者能够收集、存储、使用、加工、传输、提供、公开、删除个人信息。

⑤ Viktor Mayer-Schonberger & Yann Padova, *Regime Change? Enabling Big Data Through Europe's New Data Protection Regulation*, Columbia Science and Technology Law Review, Vol.17:315, p.315(2016).

信息的重复循环再利用。在这一过程中,不可避免地存在着“不知道何种个人信息经过技术处理后会导致个体的人格受到侵害”的情形,[①]而这也正是个人信息在大数据时代的本质风险和保护难点。

二、大数据时代个人信息保护面临的理论困境

大数据对个人信息的大规模收集和重复循环再利用能够证实最初的假设或者处理某些问题,甚至能“萃取或推衍出预想不到而可能有价值的资讯”,[②]获得个人信息收集的潜在价值。然而,基于某一特定目的或者特定情境下提供的个人信息若被大数据以完全意想不到或者不正当的方式运用,不仅可能侵害个人信息的自决权、隐私权,而且与《个人信息保护法》规定的原则相冲突。

(一)大数据的不当运用具有侵害个人信息自决权与隐私权的风险

大数据作为一种侵入性工具,能够详细记录人们的行为,汇编人们购买和消费的数据,[③]这类举动成为令人不安的监视来源,威胁着个人的隐私权。

1.大数据的再识别化攫取个人隐私

为了规范个人信息处理者的处理活动以实现保护个人信息的目的,《个人信息保护法》第 51 条[④]规定了个人信息处理者的义务,即应建立个人信息管理制度、作业流程,并进行分级分类管理,及采取加密、“去标识性”等安全技术措施。该条款表明个人信息处理者在处理个人信息活动过程中应当采取“去标识性”的技术措施。这一举措能够达到在处理个人信息过程中倘若不借助关联信息则无法识别到特定个人的目的。然而,大数据具有再识别个人信息的功能,能够通过对数据库中个人信息的汇整、统合,运用统计学与其他数据挖掘技术,将本来不具有识别性或“去标识性”的信息与其他额外信息进行关联、筛选、比对以及匹配,从而识别出特定个人信息。例如,个人信息使用者为了商业目的使用一些个人信息时,并不需要具体到个人身份,只需要识别网上 IP 信息即可。大数据技术的应用使得信息主体即使应用匿名化技术也难以完全实现隐匿的目的,有些信息表面上不是个人信息,但是经过大数据的处理仍然可以追溯到具体的个人,从而挖掘出相关联的其他信息。[⑤] 大数据的运作模式,能够更容易发现个人信息的关联性并建立起连结,通过确认、分析与预测个人人格、行为、兴趣与习惯,对个人进行剖析,用以协助自动化决策,[⑥]同时,大数据也能突破信息之间隐形因素无法被量化的瓶颈,[⑦]建构不相干信息之间的关联性,对个人信息进行再识别,在这一过程中,大数据对个人信息的获取可能转化为对个人隐私的攫取、吞噬,从而威胁着个人隐私安全。

2.大数据的自动化决策使个人信息自决权被虚置

《个人信息保护法》第 44 条[⑧]确立了个人信息自决权,即“信息主体对自身信息的控制与选择,即自我决定的权利”,[⑨]这也意味着个人享有自由决定其信息被收集、利用的权利。然而,大数据具有个人信息自动化决策功能。自动化决策,[⑩]又被称为“智能决策”,在大数据作为战略资源的当今社会,自动化决策成为决策作出的主要方式,凭借综合利用大量数据,有机结合各种模型,将海量数据汇集融合,发挥快速感知和认知能力及强大的分析与推理能力,最终从数据中提取知识,通过识别、判断,进而输出科学决策。自动化决策的计

---

① 参见张婉婷:《个人信息“合理利用”的规范分析》,载《法学评论》2023 年第 3 期,第 109 页。

② 参见翁清坤:《大数据对于个人资料保护之挑战与因应之道》,载《东吴法律学报》2020 年第 3 期,第 87 页。

③ Karen E. C. Levy, *Relational Big Data*, Stanford Law Review, Vol.66:73, p.73(2013).

④ 《个人信息保护法》第 51 条规定:“个人信息处理者应当根据个人信息的处理目的、处理方式、个人信息的种类以及对个人权益的影响、可能存在的安全风险等,采取下列措施确保个人信息处理活动符合法律、行政法规的规定,并防止未经授权的访问以及个人信息泄露、篡改、丢失:(一)制定内部管理制度和操作规程;(二)对个人信息实行分类管理;(三)采取相应的加密、去标识化等安全技术措施;(四)合理确定个人信息处理的操作权限,并定期对从业人员进行安全教育和培训;(五)制定并组织实施个人信息安全事件应急预案;(六)法律、行政法规规定的其他措施。”

⑤ 参见金泓序、何畏:《大数据时代个人信息保护的挑战与对策研究》,载《情报科学》2022 年第 6 期,第 135 页。

⑥ 参见翁清坤:《大数据对于个人资料保护之挑战与因应之道》,载《东吴法律学报》2020 年第 3 期,第 96 页。

⑦ 参见高婴劢:《工业大数据价值挖掘路径》,载《中国工业评论》2015 年第 2 期,第 22 页。

⑧ 《个人信息保护法》第 44 条规定:“个人对其个人信息的处理享有知情权、决定权,有权限制或者拒绝他人对其个人信息进行处理;法律、行政法规另有规定的除外。”

⑨ 参见姚岳绒:《论信息自决权作为一项基本权利在我国的证成》,载《政治与法律》2012 年第 4 期,第 72 页。

⑩ 《个人信息保护法》第 73 条第 2 项规定:“自动化决策,是指通过计算机程序自动分析、评估个人的行为习惯、兴趣爱好或者经济、健康、信用状况等,并进行决策的活动。”

算范式是“从数据到知识,从知识到决策”,[①]通过挖掘、萃取以及分析数据,发现其中蕴藏的知识,再由知识转化为决策支持。[②] 自动化决策具有稳定性和精确性,其决策速度也展示出超越人类的强大效能。反过来,决策的实用性也体现了大数据的价值。在个人信息领域,大数据的自动化决策已是常态,大数据凭借自动化分析数据,萃取、关联、剖析、描绘,直至揭露个人信息,甚至是暴露个人隐私,这就导致信息主体对自身信息失控,个人信息自我决定的权利被侵害。自动化决策也能据此进行数字画像,即将每一个个人绘成数据的集合,并且将与画像相吻合的各种服务和信息推送给个人。个人被置于大数据所创造的不同片区中,并没有作为活生生的个体获得尊重,[③]个人信息自我决定的权利被虚置。

3.大数据推算匿名化信息或者不为人知的敏感信息

《个人信息保护法》第4条第1款[④]规定了个人信息不包括匿名化处理后的信息,换言之,匿名化信息不属于《个人信息保护法》保护的“个人信息”范围。然而,在大数据时代,即使是匿名化的数据也可以被重新识别并归因于特定的个体,[⑤]随着网络、物联网装置等先进科技手段和信息技术的广泛应用,个人信息的收集越发便捷与普及,大数据收集的巨量信息能够发挥再识别功能。匿名化信息通过大数据的再识别,与其他信息组合、比对后,将转化为具有特定个人识别性的信息,匿名化的个人信息因此变为显名化。毋庸置疑,大数据的再识别技术削弱了个人信息的匿名化,破坏了隐私政策的格局。[⑥] 大数据能够推论出不为人知的个人信息,其核心是使用人工智能挖掘和分析大量数据,目的是找到揭示新洞见或事实的“小模式”或相关性。[⑦] 即使某些个人信息并不在数据库中,但是通过对已收集或掌握的数据进行关联、萃取和推衍,知悉或洞见特定自然人并不困难。

(二)大数据的不当运用与个人信息保护原则形成冲突

大数据的运作模式是“对世界的某些方面进行建模”,并得出“预测未来可能发生的事件”的推论。[⑧] 在这一过程中,如果大数据运用不当,将可能与《个人信息保护法》确立的立法目的和原则相矛盾。

1.大数据与目的明确原则、最小化收集原则相冲突

《个人信息保护法》第6条[⑨]规定了个人信息的处理要遵循目的明确原则,以及个人信息的收集要遵循最小化收集原则。对于目的明确原则,大数据的运作模式常常是对个人信息的重复循环再利用,这就可能超出个人信息原始收集的目的,与个人信息处理的目的明确原则形成冲突。个人信息被再利用时,可能会进入多方流转的系统中,第一方收集者将个人信息流转至第三方机构,从而失去了对个人信息的控制权,[⑩]脱离控制的个人信息将存在被恶意滥用的风险。对于最小化收集原则,其要求个人信息的收集应当依据处理目的进行最小范围、最小限度的收集,为了保障个人信息安全,不得恣意扩大个人信息的收集范围。《个人信息保护法》第47条[⑪]规定的删除权也体现了这一原则。个人信息用于特定的处理目的,该处理目的一旦实现,则必须删除该信息,以规避信息被滥用、泄漏、遗失之可能。[⑫] 按照最小化收集原则,处理个人信息时也

---

① 参见陈纯、庄越:《大数据智能:从数据到知识与决策》,载《中国科技财富》2017年第8期,第49页。

② 参见杨善林、周开乐:《大数据中的管理问题:基于大数据的资源观》,载《管理科学学报》2015年第5期,第3页。

③ 参加陈林林、严书元:《自动化决策中数据处理者的合理分析义务》,载《吉首大学学报(社会科学版)》2022年第6期,第20页。

④ 《个人信息保护法》第4条第1款规定:“个人信息是以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息,不包括匿名化处理后的信息。”

⑤ Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, Stanford Law Review, Vol.64:63, p.63(2011—2012).

⑥ Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, UCLA Law Review, Vol.57:1701, p.1701(2010).

⑦ 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.1(2018).

⑧ 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.1(2018).

⑨ 《个人信息保护法》第6条第1款规定:“处理个人信息应当具有明确、合理的目的,并应当与处理目的直接相关,采取对个人权益影响最小的方式。”第6条第2款规定:“收集个人信息,应当限于实现处理目的的最小范围,不得过度收集个人信息。”

⑩ 参见范为:《大数据时代个人信息保护的路径重构》,载《环球法律评论》2016年第5期,第94页。

⑪ 《个人信息保护法》第47条第1款规定:“有下列情形之一的,个人信息处理者应当主动删除个人信息;个人信息处理者未删除的,个人有权请求删除:(一)处理目的已实现、无法实现或者为实现处理目的不再必要;(二)个人信息处理者停止提供产品或者服务,或者保存期限已届满;(三)个人撤回同意;(四)个人信息处理者违反法律、行政法规或者违反约定处理个人信息;(五)法律、行政法规规定的其他情形。”第47条第2款规定:“法律、行政法规规定的保存期限未届满,或者删除个人信息从技术上难以实现的,个人信息处理者应当停止除存储和采取必要的安全保护措施之外的处理。”

⑫ 参见申卫星:《论个人信息权的构建及其体系化》,载《比较法研究》2021年第5期,第11-12页。

应当具有合理的目的并限于实现处理目的的最小范围。[①] 然而,大数据的运作模式实现了个人信息的大规模收集和无限次利用,挑战着个人信息最小化收集原则。

2.大数据冲击告知同意原则的传统架构

告知同意原则是信息主体与信息处理者之间所形成的权利及义务的合同规则,[②]《民法典》第 1035 条[③]确立了个人信息处理的告知同意原则,形成了“个人信息权益的基本保护模式”。[④]《个人信息保护法》第 14 条[⑤]和第 17 条[⑥]确立了以告知同意原则为核心的个人信息保护制度。虽然中国在立法上确立了告知同意原则的法律地位,但是在实施过程中大数据技术的运用导致其实际效果遭受严重质疑。[⑦]

第一,大数据时代下越来越多的个人信息在信息主体不知情的境况下被收集,例如,随身佩戴的手表、手机,公共场所的摄像头以及互联网设备等随时随地地收集个人信息,在收集过程中并未提供让信息主体知悉或者同意的方式。

第二,在实践中对于个人信息收集的隐私权政策通知,一般情况下信息主体并不会予以阅读和理解,加之隐私政策以技术性、法律性的语言为基础,呈现出冗长复杂、晦涩难懂的特征,信息主体对此不愿意耗费大量的时间,[⑧]因此,告知同意原则成为一种摆设。此外,信息主体对于隐私政策的声明只能被迫选择同意,倘若不同意则意味着放弃了使用权。

第三,虽然法律赋予个人信息主体知情同意权,但是,个人很难充分掌握数据协议的内涵,这也体现出数据处理者凭借大数据、人工智能的技术赋能或赋权,在事实上与个人信息主体之间形成一种非对称的权力结构,数据处理者的权力与信息主体的权利呈现出非均衡性,显然数据处理者占据优势地位。[⑨] 基于权力与权利的非均衡性事实,个人信息得不到有效保护。

第四,大数据对个人信息的无限次利用往往会脱离最初的收集目的,原先在告知同意原则下取得的合法授权,随着大数据的非预期运用和再利用将失去法律效力,重新取得信息主体的授权则往往不现实,这就导致告知同意原则被架空。

3.大数据的算法分析与个人信息透明化原则相矛盾

《个人信息保护法》第 7 条[⑩]规定了处理个人信息时应当遵循透明化原则。这一原则是个人信息处理的前提,是保障个人信息权益的基础,有助于增进个人信息主体的安全感和信任感,提升信息主体对个人信息处理的接受度。然而,大数据的不当运用与个人信息透明化原则相冲突。快速和复杂的数据分析能力应用到巨大的数据集中,对个人信息予以分析,这一过程被称为大数据分析。人工智能学习数据,智能地响应新的数据并且随时调整其输出,以达至预测和洞见未来。在这一过程中,人工智能使用复杂的数学算法来处理数据并基于数据作出决策,这些算法一般是非透明的,将产生所谓的“黑箱效应”。[⑪] 算法黑箱是一个建模系统,通常会智能地作出连续性的常规动作,突破数据本身,一旦涉及个人信息,可能产生系统化和机制化的侵

---

① 参见万方:《个人信息处理中的“同意”与“同意撤回”》,载《中国法学》2021 年第 1 期,第 169 页。

② 参见赵婧薇、尹伟民:《个人信息保护中告知同意规则的立法纾困》,载《内蒙古社会科学》2022 年第 2 期,第 85 页。

③ 《民法典》第 1035 条规定:“处理个人信息的,应当遵循合法、正当、必要原则,不得过度处理,并符合下列条件:(一)征得该自然人或者其监护人同意,但是法律、行政法规另有规定的除外;……”

④ 参见万方:《个人信息处理中的“同意”与“同意撤回”》,载《中国法学》2021 年第 1 期,第 167 页。

⑤ 《个人信息保护法》第 14 条第 1 款规定:“基于个人同意处理个人信息的,该同意应当由个人在充分知情的前提下自愿、明确作出。法律、行政法规规定处理个人信息应当取得个人单独同意或者书面同意的,从其规定。”第 14 条第 2 款规定:“个人信息的处理目的、处理方式和处理的个人信息种类发生变更的,应当重新取得个人同意。”

⑥ 《个人信息保护法》第 17 条第 1 款规定:“个人信息处理者在处理个人信息前,应当以显著方式、清晰易懂的语言真实、准确、完整地向个人告知下列事项:(一)个人信息处理者的名称或者姓名和联系方式;(二)个人信息的处理目的、处理方式,处理的个人信息种类、保存期限;(三)个人行使本法规定权利的方式和程序;(四)法律、行政法规规定应当告知的其他事项。”第 17 条第 2 款规定:“前款规定事项发生变更的,应当将变更部分告知个人。”

⑦ 参见冯健鹏:《个人信息保护制度中告知同意原则的法理阐释与规范建构》,载《法治研究》2022 年第 3 期,第 32 页。

⑧ 有研究表明,用户一年中阅读使用的网络服务的隐私声明要花费 244 小时,参见范为:《大数据时代个人信息保护的路径重构》,载《环球法律评论》2016 年第 5 期,第 93 页。

⑨ 参见陈林林、严书元:《自动化决策中数据处理者的合理分析义务》,载《吉首大学学报(社会科学版)》2022 年第 6 期,第 20 页。

⑩ 《个人信息保护法》第 7 条规定:“处理个人信息应当遵循公开、透明原则,公开个人信息处理规则,明示处理的目的、方式和范围。”

⑪ 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.1(2018).

权后果。[①] 算法黑箱的成因是,第一,算法本身具有保密性质,算法研发者或控制者缺乏将其对外披露的意愿;第二,算法技术壁垒的客观存在,使得非专业人员难以理解算法的运行逻辑与真实内涵;第三,算法的自主学习特性导致其运行过程变得极其复杂,甚至超出算法设计人员所能感知的范畴。[②] 从成因来看,算法黑箱可能是算法设计师主观为之,也可能是由于技术原因客观存在。在"黑箱效应"下,要求处理个人信息时公开个人信息处理规则,明示处理的目的、方式和范围,显然是强人所难的。

三、大数据时代下个人信息的保护路径

大数据时代,《个人信息保护法》所规定的个人信息权益面临前所未有的风险,但绝不能为了保护个人信息而一味地钳制与约束大数据的发展。大数据的创新应用是社会的驱动力与发展力,能产生巨大的经济效应和社会效应,不仅关系到个人的福祉,也关系着国家的未来。[③] 顺应科技发展,同时立足于个人信息保护作出应对政策,兼顾个人隐私保护和数据发展,实现隐私保护与数据价值开发的共赢,是大数据时代的发展需求。

(一)制定个人信息开放规范以促进大数据发展

《个人信息保护法》以个人对信息的控制为前提,以个人信息最小化收集和目的明确、目的限制等原则为基础,以保护个人信息权益与合理利用为目的。然而,大数据的运作模式极易侵害个人信息自决权与隐私权,尽量减少个人信息收集看似是个人隐私保护的一个实际方法,但其并不可行,因为个人信息作为大数据的构成部分有着巨大的潜能,其不仅能够带动产业创新,还有助于发展社会经济。[④] 因此,个人隐私和个人信息保护必须与公共健康、国家安全和执法、环境保护和经济效率等附加社会价值相平衡。在保护个人信息的同时必须兼顾大数据的发展。一味地钳制大数据并不利于国家经济发展,因此,转向制定个人信息开放规范,对个人信息进行适度松绑,反而能够兼顾二者的共同发展。一方面,此举改变了巨量的个人信息被少数处理者所掌握的局面与境地,降低了垄断造成的危害;另一方面,开放数据有助于提升个人信息透明化并创造出新的价值。例如,在经济领域出现的开放银行要求银行开放客户资料,改善银行垄断金融资料的现象,允许第三方合作伙伴在获得客户授权后存取账户资料,将金融资料主导权交还消费者,使消费者获得更多元的金融服务。[⑤]

有鉴于个人信息很大部分被政府部门掌握,政府开放数据对经济增长、包容性发展和改善公众参与的价值意义重大。出于为数据赋能的目的,应降低政府数据供公众利用的难度。匿名化信息的共用在提升国家竞争力和科技创新发展方面发挥着重要作用。目前,中国现行政府数据开放规范多是各地方政府制定的规范性文件,如《浙江省公共数据开放与安全管理暂行办法》《青岛市公共数据开放管理办法》《重庆市公共数据开放管理暂行办法》等。以地方规范性文件规范数据开放行为虽然能推动数据开放,但是存在科学性、统一性不足的弊端,[⑥]因此,中国应当制定政府数据开放的相关法律。制定时应当遵循《政府信息公开条例》所规定的政府信息公开的基本原则——"公开为原则,不公开为例外",坚持政府数据"开放为原则,不开放为例外"的开放路径,以最大限度地保证数据生产要素的供给。[⑦] 当然,"开放为原则,不开放为例外"的适用前提是,政府数据开放行为并不损害公共利益和私人利益。当为了实现公共利益而必须让渡私人利益时,应受到比例原则[⑧]的限制,例如,个人隐私利益应当向基于安全保障的公共利益让渡,而政府基于公共利益开放

---

① 参见林洹民:《自动决策算法的风险识别与区分规制》,载《比较法研究》2022年第2期,第189-190页。

② 参见李欣、曹艺萱:《我国算法风险及其治理研究综述》,载《信息安全研究》2024年第2期,第114页。

③ 参见池建新:《个人信息保护政策的国际比较研究》,东南大学出版社2021年版,第35页。

④ 参见申卫星:《大数据时代个人信息保护的中国路径》,载《探索与争鸣》2020年第11期,第8页。

⑤ 参见蔡鹏程:《赋能实体,开放银行的另一条道路》,载搜狐网2022年7月28日,https://www.sohu.com/a/572480954_116132。

⑥ 参见王东方:《政府数据开放规范的精细化构建——基于政府数据开放与政府信息公开的关系视角》,载《电子政务》2021年第10期,第29-30页。

⑦ 参见王东方:《政府数据开放规范的精细化构建——基于政府数据开放与政府信息公开的关系视角》,载《电子政务》2021年第10期,第35页。

⑧ 比例原则是指行政主体实施行政行为时,应当兼顾行政目标的实现和行政相对人权益的保护,如果为实现行政目标可能对行政相对人权益造成某种不利影响,那么应使这种不利影响限制在尽可能小的范围和限度内,保持二者处于适度的比例。参见莫于川主编:《案例行政法学》,中央广播电视大学出版社2009年版,第34页。

数据时则不能损害个人利益。有学者将其总结为,在政府数据开放的具体情境中,基于公共利益对个人利益进行限制属于法律适用,应结合具体情境动态判断公共利益的内容和比例原则的适用。① 那么,如果开放政府数据不能实现公共利益,同时却对私人利益造成损害,此类政府数据不应予以开放。总之,在制定开放规范时所开放的个人信息应当相应地满足信息主体隐私期待并使个人信息受到尊重。② 另外,应进一步将个人隐私保护政策具体化,如设置专门的隐私保护机构以提供隐私政策咨询和支持,任命隐私保护专业人员以负责可能涉及的隐私问题,在开放数据时进行隐私影响评估,③构建完善的隐私风险管理体系,细化风险识别机制、风险评估机制和风险控制机制,从而对数据的隐私风险进行量化、分析、减轻。④ 由此,平衡好个人信息“利用”和“保护”二者之间的关系,确保个人信息利用的合法化与合理化。

(二)采取牢固的匿名化技术以实现个人信息的“去连结性”

防止大数据侵害个人隐私权的有效措施是将个人信息进行匿名化以实现个人信息的“去识别性”⑤和“去连结性”⑥,然而,《个人信息保护法》对于“去识别性”的规定过于简陋,对匿名化方法和技术也未作规定。《个人信息保护法》第 4 条将“可识别性”作为个人信息保护的界限,即受《个人信息保护法》保护的个人信息须具有“识别性”,以“识别性”作为个人信息保护该当性的要件,失去了识别性的信息被排除在《个人信息保护法》保护的范围之外。而《个人信息保护法》第 73 条第 4 项仅规定了匿名化的含义。⑦ 因此,有必要采取“去识别性”的匿名化技术,建构与规范具体的匿名化制度,包括匿名化的程度、程序和监督机制。

个人信息匿名化是平衡数据价值与主体权益的重要手段,在技术层面,可以采用去中心化的匿名化方法和个性化匿名化方法。去中心化的匿名化方法⑧是基于区块链系统,通过智能合约进行信息交互。相较于中心化的集中式结构,去中心化使得数据并不集中在任何一个中心节点或实体手中,而是建构了更扁平、更平等、更分散的结构,从而解决了单点故障和传统匿名化技术中数据共享双方的信任问题,有效保护了信息主体的隐私。个性化匿名化方法是针对现实世界中不同的隐私需求,允许信息主体自己控制和定义隐私数据的用途。现实中,每个主体的隐私保护需求、个人信息利用敏感程度等是不同的,更注重隐私的主体一般认为默认的匿名化等级难以满足隐私保护需求,而另一些人则可能会觉得默认的匿名化等级过高,因此,统一的匿名化等级无法满足不同主体的隐私保护需求。而个性化匿名化方法根据不同主体提供了不同的隐私级别,由信息主体自己控制数据发布的匿名化级别,以此满足个人隐私需求和差异化保护。如此一来,既尊重了个人隐私偏好,又最大限度地发挥了数据的可用性。

在匿名化制度层面,值得借鉴参考的是日本《个人信息保护法》中的匿名加工制度,该法的第 2 条、第 36 条至第 38 条和第 53 条规定了匿名加工信息的制作、利用时应遵守的义务。具体而言,日本《个人信息保护法》第 36 条第 1 项⑨规定,由个人信息保护委员会订立规则和确定标准,按照《个人信息保护委员会规则》所规定的一般性最小限度的加工方法予以匿名加工。为了防止匿名化制度过于僵化,匿名加工方法不宜采用统一方式和划定一致标准。不同业界匿名加工方法应考量其处理个人信息的内容、利用目的,因此,日本《个人信息保护法》第 53 条第 1 项⑩规定实践中各业界运用匿名加工方法,可以再委任个人信息保护认证组

---

① 参见吴亚光:《政府数据开放中个人隐私信息的公开界限》,载《图书馆学研究》2020 年第 22 期,第 48 页。

② 参见翁清坤:《大数据对于个人资料保护之挑战与因应之道》,载《东吴法律学报》2020 年第 3 期,第 136-137 页。

③ 参见黄如花、温芳芳:《我国政府数据开放共享的政策框架与内容:国家层面政策文本的内容分析》,载《图书情报工作》2017 年第 20 期,第 12 页。

④ 参见梁乙凯、陈美:《美国隐私影响评估制度及其启示》,载《情报资料工作》2022 年第 5 期,第 68 页。

⑤ 所谓“去识别性”,是指数据保有者采用技术手段对其所保有的数据信息进行集中的筛查,将其中能够识别特定个人身份的数据信息予以删改的过程。参见张勇:《个人信息去识别化的刑法应对》,载《国家检察官学院学报》2018 年第 4 期,第 96 页。

⑥ 所谓“去连结性”,是指通过个人信息匿名化的技术措施有效削弱和去除信息与特定主体之间的关联性。参见刘晓春、刘瑾:《个人信息匿名化标准的实践和优化》,载《中国对外贸易》2023 年第 8 期,第 31 页。

⑦ 《个人信息保护法》第 73 条第 4 项规定:“匿名化,是指个人信息经过处理无法识别特定自然人且不能复原的过程。”

⑧ Romana Talat & Mohammad S. Obaidat, et al., *A Decentralised Approach to Privacy Preserving Trajectory Mining*, Future Generation Computer Systems, Vol.102:382, p.382(2020).

⑨ 日本《个人信息保护法》第 36 条第 1 项规定:“个人信息处理经营者在制作匿名化信息(仅限于匿名化信息数据库等的组成信息,下同)时,应按照《个人信息保护委员会规则》规定的标准处理个人信息,以使其无法识别特定个人且无法恢复制作所用的个人信息。”

⑩ 日本《个人信息保护法》第 53 条第 1 项规定:“为确保其服务范围内的经营者妥当处理个人信息等,针对利用目的说明、安全管理措施、披露等要求等的应对程序以及匿名化信息的制作方式、安全管理措施等个人信息相关事项,个人信息保护认证组织应征求消费者代表或其他相关人员的意见,努力制定符合本法规定的目的的指引(以下简称《个人信息保护指引》)。”

织制定《个人信息保护指引》作为团体自律性规范。为了避免个人信息保护认证组织在制定时过于片面,该条还规定了听取消费者代表及其他相关人员的意见,[①]在此基础上制定出匿名加工规则,建立互信的匿名加工制度,完全切断特定主体与个人信息内容的连结性,减少大数据对个人信息和个人隐私的侵害。通过匿名化技术的应用以及具体的匿名化制度,共同构建牢固的个人信息保护路径,有效削弱和去除信息与特定主体之间的关联性,实现个人信息的"去连结性"。

(三)将个人信息保护影响评估作为保护个人信息的有效工具

大数据时代个人信息保护影响评估作为事前防范机制比以往任何时候都更加重要。[②] 事前性的合规评估和风险评估程序能够预测个人信息处理活动的结果,提前为个人信息安全提供预防性保护措施,降低个人信息安全风险。为此,2020年中国出版了《信息安全技术 个人信息安全影响评估指南(GB/T39335—2020)》(简称《指南》),该指南建构了一个框架,旨在为个人信息安全影响评估提供基本原理和操作步骤等,并且明确了个人信息安全影响评估的国家标准,更加具体地帮助评估组织识别和减轻个人信息处理的相关风险,为个人信息保护提供实质性的工具。随后,2021年《个人信息保护法》第55条[③]规定了个人信息保护影响评估的"适用情形",第56条第1款[④]规定了个人信息保护影响评估的"具体内容",这两项主要条款标志着个人信息保护影响评估制度初步确立,以及该制度从推荐性要求上升为信息处理者的法定义务,这也揭示了个人信息保护影响评估旨在防止或最大限度减轻不利影响,并形成事前预测,根据评估结果制定针对性的应对方案,在处理个人信息活动时预防风险的发生和不利后果。[⑤]

《个人信息保护法》只是初步回答了哪些情形应当予以个人信息保护影响评估,以及评估涉及哪些内容。为了使评估结果更加客观、有效,理应科学合理地设定个人信息保护影响评估程序,在此可以参考《指南》中规定的个人信息安全影响评估实施的九个步骤,即(1)评估必要性分析;(2)评估准备工作;(3)数据映射分析;(4)风险源识别;(5)个人权益影响分析;(6)安全风险综合分析;(7)评估报告;(8)风险处置和持续改进;(9)制定报告发布策略。其中,(3)至(8)是评估的主体内容,数据映射分析是评估的基础性工作,通过此步骤确定评估对象。在这一过程中可以采用问卷、调研、走访等方式,多维度地梳理个人信息处理活动,形成数据清单及数据映射图表,从而确定评估内容和范围;此步骤之后需要对所涉风险进行识别和分析,主要从《指南》中规定的安全事件和个人权益影响两个维度进行分析,前者侧重于识别可能面临的危险源及是否采取安全措施,后者侧重于识别对个人权益造成的不利影响;接下来是综合评估,在前述步骤基础上综合衡量并得出个人信息处理活动的风险等级,根据风险等级进一步采取相应措施,进行风险管控和处置,确保风险始终在可控范围之内。与此同时,整个评估程序应当特别重视"参与程序、复审程序、事先咨询程序和公开程序"。[⑥]

(四)以"实质性参与"作为告知同意原则的完善路径

无论是在未提供告知下毫不知情地收集个人信息,还是告知同意原则囿于晦涩难懂的技术性语言成为摆设,以及个人信息脱离最初收集目的而无法回溯获得授权,都将导致告知同意原则被架空。当下,告知同意原则已经备受质疑,甚至被认为已然失效。控制个人信息的利用环节可能是更为有效的方式,这也更符合用户的隐私偏好与期待。[⑦] 然而,也有学者认为,告知同意原则以人格尊严为基础,以尊重人本身的判断和选择为前提,因此,在设置告知同意时,应将人视为目的而非手段、尊重个人用户的理性判断和选择。[⑧] 告知同意原则源于宪法对人格尊严的保护,不能否定该原则在个人信息保护中的重要地位。除《个人信息保护

① 参见范姜真媺:《大数据时代下个人资料范围之再检讨——以日本为借镜》,载《东吴法律学报》2017年第2期,第16-18页。

② 参见梁乙凯、陈美:《英国数据保护影响评估制度及其启示》,载《情报理论与实践》2022年第7期,第202页。

③ 《个人信息保护法》第55条规定:"有下列情形之一的,个人信息处理者应当事前进行个人信息保护影响评估,并对处理情况进行记录:(一)处理敏感个人信息;(二)利用个人信息进行自动化决策;(三)委托处理个人信息、向其他个人信息处理者提供个人信息、公开个人信息;(四)向境外提供个人信息;(五)其他对个人权益有重大影响的个人信息处理活动。"

④ 《个人信息保护法》第56条第1款规定:"个人信息保护影响评估应当包括下列内容:(一)个人信息的处理目的、处理方式等是否合法、正当、必要;(二)对个人权益的影响及安全风险;(三)所采取的保护措施是否合法、有效并与风险程度相适应。"

⑤ 参见陈朝兵、郝文强:《作为政府工具的隐私影响评估:缘起、价值、实施与启示》,载《中国行政管理》2020年第2期,第146页。

⑥ 参见刘权:《论个人信息保护影响评估——以〈个人信息保护法〉第55、56条为中心》,载《上海交通大学学报(哲学社会科学版)》2022年第5期,第46页。

⑦ 参见范为:《大数据时代个人信息保护的路径重构》,载《环球法律评论》2016年第5期,第109页。

⑧ 参见冯健鹏:《个人信息保护制度中告知同意原则的法理阐释与规范建构》,载《法治研究》2022年第3期,第38页。

法》建构的告知同意原则的基本规范体系外,还应当塑造精细化的、实质性参与的告知同意原则,充分实现信息主体的知情权与同意权。①

具体而言,在告知层面,对告知协议的形式和内容作出限定,尽可能以简短、精炼的语言进行告知。网络平台提供告知时,可以藉由 code 或装置本身提供隐私通知。不阅读用户隐私通知政策、隐私政策通知的冗长晦涩以及选择“不同意”将失去使用的权利,其实质均是没有选择权。对于这种情形,为了充分实现信息主体的知情权,保障其充分参与,隐私通知应当清楚简化。欧盟《通用数据保护条例》(*General Data Protection Regulation*)鼓励“资讯须简洁,易于取得及理解,用语清楚简明,且可适时视觉化”。关于没有提供告知的情形或较难提供告知时,应提供创新性的隐私通知方式,通过多样化的方式充分保障信息主体的知情权。信息处理者的告知义务对应着信息主体的知情权,以公共场所的视频监控为例,出于对知情权的保护,应当作出监控设备的提示。告知方式的创新,有利于信息主体实质性地参与个人信息收集与处理的过程,有助于大数据时代下个人隐私保护意识的提高。如果大数据对个人信息的非预期利用使得个人信息脱离最初收集目的,信息处理者有义务更新隐私通知,持续、动态地征求信息主体同意,确保信息主体知悉并且作出是否同意个人信息被新目的利用的决定。总之,信息主体的实质性参与是保障知情同意原则实施的根本。创新隐私通知政策,为信息主体提供富有实效意义的隐私通知,尽早告知信息利用的目的,保障个人信息主体的权益,是大数据时代应坚守之道。

四、结语

大数据技术的飞速发展对个人信息保护提出了新的要求。一方面,大数据快速收集与储存各种来源的个人信息,其范围不仅局限于互联网,还包括越来越多传统的数据集。另一方面,通过信息处理技术,个人信息之间被不断地建立关联、重复利用,甚至脱离最初的收集目的,被流转至第三方手中。尽管大数据时代下个人信息权益保护遭遇了前所未有的挑战,然而,如果为了保护个人信息而制约大数据发展,其所带来的负面影响将辐射整个社会。大数据的创新应用是社会发展的驱动力,因此,个人信息保护与大数据发展必须同时兼顾。《个人信息保护法》的立法目的强调了个人信息权益保护和个人信息合理利用的平衡,二者不可偏废。个人信息的安全价值和利用效率价值同等重要,因此,“大数据时代,个人信息法律保护不是单纯的权利实现,而是要在个人信息有效利用与权利行使之间寻找平衡”,②为兼顾大数据发展与隐私保护义务,应当制定个人信息开放规范以促进大数据发展,采取牢固的匿名化技术以实现个人信息的“去连结性”,将个人信息保护影响评估作为保护个人信息的有效工具,以“实质性参与”作为告知同意原则完善的路径取向,由此实现个人信息保护和利用的双重目的。

① 参见翁清坤:《大数据对于个人资料保护之挑战与因应之道》,载《东吴法律学报》2020 年第 3 期,第 139-140 页。
② 王秀哲:《大数据时代个人信息法律保护制度之重构》,载《法学论坛》2018 年第 6 期,第 121 页。

The Operation Mode, Theoretical Dilemma and Protection Path of Personal Information in the Era of Big Data

XUE Wujuan

(School of Politics and Public Administration, China University of Political Science and Law, Beijing 100088, China)

Abstract: The rapid development and continuous application of big data technology have brought about a disruptive shift in the mode of operation of personal information, that is, personal information can be collected in huge quantities and recycled. On the one hand, big data rapidly collects and stores personal information from a variety of sources, not limited to the Internet, but also including data sets. On the other hand, through information processing technologies, personal information is constantly linked to each other and is not only used to answer or analyze the specific question, but also reused out of structured databases. In some cases, it is even transferred to a third party and recycled for purposes other than those for which it was originally collected. In this process, personal information is subject to the infringement of an individual's personality, which is precisely the nature of the risk and the difficulty of protecting personal information in the era of big data. With the cumulative effect of personal information, big data can correlate unknown information, and its inappropriate application poses the risk of violating the right to self-determination and the right to privacy of personal information. On the one hand, the identification of big data can capture personal privacy, and the related automated decision-making can void the right to personal self-determination, and big data can also derive anonymous information or sensitive information. On the other hand, improper use of big data may conflict with the principles of protection set forth in the *Personal Information Protection Law*. For example, the principle of clarity of purpose, the principle of minimization, the principle of informed consent, and the principle of transparency of personal information. The rights and interests of personal information in the era of big data are facing unprecedented challenges, but if the development of big data is restricted in order to protect personal information, the negative impact will radiate to the whole society and the country. The innovative application of big data is the driving force, which can produce huge economic and social effects. Therefore, the protection of personal information and the development of big data must be taken into account at the same time. Firstly, the "open norms" of personal information should be formulated. The fact that most of the personal information is held by the governmental departments, and the government should adhere to the principle of "openness as a principle and no openness as an exception" and open data for public use, especially anonymous information, which plays an important role in enhancing national competitiveness and the development of science and technolgy. Secondly, solid anonymous techniques are adopted to de-anonymize personal information. For example, decentralized approach and personalized anonymity approach. Thirdly, we should utilize personal information protection impact assessment as an effective tool for protecting personal information, so that the results of personal information processing activities can be predicted and the risk of personal information security can be reduced through prior compliance assessment and risk assessment procedures. Lastly, the principle of informed consent should be improved through "substantive participation", and the form and content of the informed agreement should be limited, so that substantive information is provided in as short and concise a language as possible.

Key words: personal information; big data; right to privacy; *Personal Information Protection Law*