Advances in Social Sciences 社会科学前沿, 2025, 14(12), 516-524
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2025.14121120

Hans汉斯

AI数据过度收集风险的驱动因素与治理策略

陈　瑜

大连海洋大学海洋法律与人文学院，辽宁　大连

收稿日期：2025年10月29日；录用日期：2025年12月10日；发布日期：2025年12月22日

摘　要

AI应用不仅带来大量的社会红利，也带来一定的社会风险，其中AI应用数据过度收集问题就是需要提前预防技术风险之一。为了探究AI应用数据过度收集风险的驱动因素，研究基于中国场景案例，运用扎根理论识别AI应用数据过度收集风险的驱动因素。研究结果表明中国场景存在四类驱动因素：超越业务范围收集数据的技术因素，未经许可或征求意见而收集、使用或公开信息数据的价值因素，政府政策过度引导的政策因素，政策制定缺乏技术伦理考量制度因素。

关键词

AI，数据收集，技术风险，扎根研究

The Risk Sources and Governance Strategies of Excessive AI Data Collection

Yu Chen

School of Marine Law and Humanities, Dalian Ocean University, Dalian Liaoning

Received: October 29, 2025; accepted: December 10, 2025; published: December 22, 2025

Abstract

The Application of AI not only generates substantial social benefits but also introduces certain societal risks. Among these, the excessive data collection by AI applications is a technical risk that requires proactive prevention. This study investigates the driving factors behind the risk of excessive data collection in AI applications. Based on case studies within the Chinese context and employing grounded theory, the research identifies the key drivers of this risk. The findings reveal four categories of driving factors in the Chinese context: technological drivers related to data collection beyond necessary business boundaries; value-based drivers involving the collection, use, or disclosure of data without permission or consultation; policy-driven drivers stemming from excessive

文章引用: 陈瑜. AI 数据过度收集风险的驱动因素与治理策略[J]. 社会科学前沿, 2025, 14(12): 516-524.
DOI: 10.12677/ass.2025.14121120

governmental policy guidance; and institutional drivers arising from a lack of consideration for technology ethics in policy formulation.

Keywords

AI, Data Collection, Technical Risk, Rooted Research

1. 引言

AI 作为全球第四次工业革命的重要驱动力，已经在政府管理、公共交通、医疗服务、消费零售等多个领域发挥了巨大贡献。伴随着技术高速发展给社会带来的红利，人们也需要对技术可能产生的社会风险加以预防，其中 AI 数据过度收集问题就是产生的技术风险之一[1]。数据过度收集是人工智能技术运营商凭借先进、隐蔽的技术算法，过度采集、存储与使用用户信息数据，准确捕捉用户个性特征，精准描绘用户画像、定位用户需求并预测其未来行为选择，以此满足自己商业营销目的的行为过程。例如新浪微博修改用户使用协议案例引发了用户数据所有权的争议。中国人脸识别第一案引发了企业数据收集边界的争论。实践中，中国政府已经意识到此类问题带来的社会风险并出台相应的管理政策[2]，但是技术应用仍然存在此类情况，例如 2021 年滴滴顺风过度收集个人信息案例等。所以，技术政策制定步伐很难与技术创新速度保持一致，一直是政府回应社会需求时面临的管理难题[1]。

国内外学者关于人工智能技术数据收集风险展开了相关研究。部分研究探讨了人工智能技术数据过度收集会带来信息泄露等技术安全[2]，对数据收集进行了总体画像描绘[3]。部分研究将数据收集进行了初步分类，包括多渠道收集、在线跟踪及预测行为和销售数据等，并探讨数据收集对消费者隐私影响、个性化营销与消费者反应、消费者选择加入或退出人工智能技术应用影响因素等[4]。遗憾的是，目前有关讨论人工智能技术数据过度收集的文献中，提炼数据过度收集风险的驱动因素的研究非常有限，尚不能为优化人工智能技术数据过度收集风险提供清晰的框架图景。因此，本文将研究问题凝练为：AI 数据过度收集风险的驱动因素是什么？将中国案例数据过度收集风险的驱动因素总结出来，不仅为中国政府制定优化 AI 风险政策提供启示与借鉴意义，同时为新兴经济体中企业如何在动态情境中实现技术良性发展提供现实指导。

2. 国内外研究综述

数据收集是近几年伴随着 AI 技术兴起带来的关于数据如何安全运用的政府企业或个体行为。郭海等(2019)认为数字化企业会收集和利用大量用户数据，并在收集和使用数据过程中违规使用[5]。这些数据并不仅仅限于我国《网络安全法》和《民法》所定义的识别自然人个人身份的各种信息，还包括用户与网络交互过程中产生的浏览、加购、收藏、搜索等一系列数据。林海(2022)进一步认为数据过度收集是网络经营者违反《中华人民共和国个人信息保护法》关于“信息收集者收集用户个人信息应当获得用户同

[1]人民网. APP 越界索权 呼唤规范治理[EB/OL]. [2018-03-22]. http://scitech.people.com.cn/n1/2018/0322/c1007-29881581.html
[2]网信办秘书局, 工业和信息化部办公厅, 公安部办公厅, 市场监管总局办公厅.网络安全实践指南——移动互联网应用基本业务功能必要信息规范[EB/OL]. [2021-03-12]. http://www.gov.cn/zhengce/zhengceku/2021-03/23/content_5595088.htm
国务院新闻办公室. App 违法违规收集使用个人信息行为认定方法[EB/OL]. [2019-12-20]. http:www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/42311/44109/xgzc44115/Document/1691066/1691066.htm

意，并且不得过度收集个人信息”规定的数据收集；是信息收集者强制读取用户的隐私数据(如个人定位、通讯录等)，或者技术平台在用户不知情情况下偷偷运行，收集与其提供服务无关的数据[3]。

文献中数据过度收集风险体现在五个方面：第一，多渠道收集数据，公众不具备控制数据的能力。研究者认为包括 3 类利用 AI 收集数据的渠道：一是技术运营商通过 APP、WIFI 和基站 3 种定位方式，能够实现持续、精准收集用户信息[6]。二是公共部门利用公共场所视频监控收集公众信息数据，这些数据被数据挖掘技术整合成隐私信息[7]。三是公众会主动上传隐私信息，公众将微博、微信、抖音等各种社交平台作为表达价值判断、关系亲疏和融入社会关系的平台，并向社交媒体泄漏一定的隐私且具有合理的逻辑前提[6]。以上三种收集数据渠道并没有主动赋予公众控制数据的能力，在遭遇非法收集或网络盗窃情况下，公众信息安全将受到侵害。第二，超越业务范围收集数据，用户信息存在被泄漏风险。此类数据收集表现为技术运营商不仅收集业务功能所需要的用户数据(例如浏览时间、收藏、加购数据等)，还会收集业务功能之外的用户数据[8] (例如位置、联络人、相册存储数据等)。知情同意原则对数据收集者约束力薄弱、数据存储和使用未受体系化监管以及数据跨境流通规范过于模糊是数据安全风险的主要成因[9]。如果这些用户数据被用于业务功能之外，或者进行了非法交易，用户信息将存在被泄露的风险[10]。第三，在线跟踪及预测数据将窥视到用户隐私。研究表明部分技术运营商会持续跟踪收集用户观看、浏览等数据，并推测出这些数据背后隐藏的社交联系[11]、种族[12]、甚至是身份证号码[13]。部分技术运营商将多个 APP 程序收集的用户数据进行拼接，预测用户偏好、消费倾向并向其进行个性化推荐，同时为广告商提供用户行为定位[14]。例如 Bleier A，Eisenbeiss M (2015)研究了某视频公司利用收集的数据推断出用户种族，向其推送特定广告，被这些种族群体认为窥探其隐私[12]。第四，销售数据将侵犯用户隐私。研究者发现技术运营商通过 3 种形式销售用户数据：一是部分运营商为了获利出售用户信息，同时部分运营商为了获得用户需求重金购买数据的互惠需求情形[10]。这会造成用户数据被用于技术运营商业务功能之外的目的，极易威胁到用户的个人隐私[15]。二是技术运营商和第三方平台合作，技术运营商 APP 程序中包含第三方 SDK 程序，第三方以嵌入代码、插件等形式收集和处理平台用户信息[16]。三是违规收集数据训练 AI 大模型面临版权风险，包括直接侵犯版权、衍生作品版权归属模糊、版权链断裂、法律风险及合规难题[15]。

因此，就 AI 数据过度收集风险单独研究成果较为丰富，为我们即将开展的研究奠定了宝贵的基础。但是将中国情境中 AI 数据过度收集风险的驱动因素挖掘出来的文献还较为鲜见。基于此考虑，本文在已有成果基础上，研究中国案例 AI 应用数据过度收集风险的内在驱动因素，为优化 AI 应用数据过度收集风险提供理论参考。

3. AI 数据过度收集风险的驱动因素

本文采用扎根理论研究方法，按照开放式、主轴式、选择式三级编码程序对原始资料系统性梳理与归纳[17]。本文质性材料来自 3 种类型：一是媒体报道，包括主流媒体报纸(人民网、南方日报等电子版)、政府官方网站、主流媒体网站(人民网、新华社、新浪网、腾讯网等)等；二是访谈，包括科学社区(知乎等)、公众微博等；三是文献，包括 SSCI、CSSCI 等数据库，实现数据的三角验证。由于 AI 应用引发的社会风险多出现于 2010 年之后，因此本文筛选出 2010~2022 年间符合本文主题的 9 个案例作为研究对象，包括：2015 年新浪微博诉脉脉(案例 1)；2016 年大众点评诉百度案例(案例 2)；2017 年新浪微博修改用户协议案例(案例 3)；2017 年腾讯诉华为 Magic 手机案例(案例 4)；2019 年中国人脸识别第一案(案例 5)；2019 年监测头环进校园案例(案例 6)；2019 年以“码”管理案例(案例 7)；2020 年外卖骑手被困算法案例(案例 8)；2021 年滴滴顺丰等 APP 手机软件案例(案例 9)。案例可以分为企业过度收集数据案例和政府过度收集数据案例两种类型，基本涵盖了社会实践中数据过度收集行为的情景，共获得质性文本资料

共 134,612 字，当检索内容无法提供新的有效信息时，表明数据检索已达到饱和[16] [18] [19]。

经过三级编码过程，本文归纳出四大核心风险驱动因素，即技术驱动因素、价值驱动因素、政策驱动因素、制度驱动因素(研究过程示例如附录表 A1、表 A2 所示)。

3.1. 技术驱动因素

技术驱动因素表现为技术系统超业务范围收集数据。大数据是 AI 应用的基石，企业为了发展和更新自身技术应用，以主动和自动化的数据挖掘技术收集大量数据。虽然中国《个人信息安全法》(2021)第 5~7 条、第 13~17 条和《网络安全法》(2016)第 41~46 条规定收集个人信息需要主体明确同意和授权，但是企业如果对每一种采集都征得主体同意，不仅降低企业效率，在现实技术处理中也存在很大困难。所以，企业数据挖掘技术“只是将早已拟就的、规避法律风险的知情同意文件呈现在用户面前”[20]。如果用户不同意其中一种行为或者未来每个数据收集行为，那么用户将不得不放弃使用该应用的整套服务[21]，所以企业自然在用户“同意”下“合法”的尽可能地收集数据。另外，绝大多数知情同意政策存在阅读困难、模糊性语言表述、未清晰表述第三方共享机构信息等问题[22]。例如案例 1 表现为脉脉软件与新浪微博停止合作后，仍然抓取新浪微博用户头像、教育信息及用户自定义标签等信息，脉脉软件将这些非用户信息与自身用户信息相关联，扩大自身软件数据范围与影响力。案例 5 表现为杭州野生动物园针对申领年卡游客要求采集指纹和人脸信息，动物园初期将指纹作为年卡识别条件，后期又自主扩展为人脸信息。这两个案例都体现了企业利用“知情同意”政策获取用户授权后，自行决定如何处理用户数据，由此引发的技术风险用户或非用户信息数据被迫公开，危害了用户信息安全。

3.2. 价值驱动因素

价值驱动因素表现为未经许可或征求意见而使用、公开信息数据。对于政府或企业而言，公共价值应该是将公众意见、社会利益融入技术政策或技术决策之中。发展型国家政府倾向于放松监管为科技创新营造更加宽容自由的环境[23]，这种政策管理理念带动着各地方政府配套出台各自的 AI 发展政策。各类 AI 企业在政策引导下，做出有利于自身的技术决策。例如案例 1、案例 2、案例 5 都是企业单方面决定数据如何使用，完全把产生数据的公众排除在外，企业代替用户、利益相关者意愿做决策，是价值后置考量的表现。由此引发的技术风险是限制消费者选择，误解消费者；损害同行经济利益，构成不正常竞争。案例 7 是苏州政府计划强行升级渐变健康码、扩展场景应用范围，地方政府利用 AI 为自己“加权”，未征求公众意见前提下代表公众做出决策，也是公共价值后置考量的表现，引发的社会风险是政府公信力大大减少。而案例 9 则表明中央政府根据社会舆论反馈的 AI 数据收集问题之后，出台了相应的处罚文件，是公共价值考量的体现。

3.3. 政策驱动因素

政策驱动因素表现为政府政策过度引导。中国政府对利用 AI 提升政府治理效率非常感兴趣[24]，特别是抗疫时期政府利用 AI 快速筛查病毒感染者的行程轨迹，极大地提高了政府疫情防控效率。各地方政府也竞相成为“智能政府”的先锋，各地开始对“健康码”、“行程码”进行扩展与延伸，其中苏州“文明码”就是地方政府试图以量化的文明提升市民文明管理效率。在政府政策引导下，AI 应用商也会制定有利于自身的技术使用政策。例如案例 3 表现为《微博用户使用协议》存在霸王条款，单方面规定用户无权利使用已经发布在微博的内容，并强制用户授权给微博法律诉讼权。腾讯诉华为案例表现为华为想调用微信用户数据，但腾讯以侵犯用户隐私为由拒绝提供，双方企业将用户排除在外展开的数据使用纠纷。案例 5 同样如此。由此引发的技术风险是侵犯用户许可使用权、索赔权、公民隐私权。同时，公民

缺乏有效的救济渠道，减弱公众对技术政策的遵从。

3.4. 制度驱动因素

制度驱动因素表现为政策制定缺乏技术伦理考量。责任是一个法律概念，用来判定某个主体违背法律义务行为时应承受的不良后果[25]。责任确定可以是国家政府以法律、法规形式颁布的“硬法”，也可以是群体间认定的符合某种道德标准的行为准则等“软法”。中国政府管理网络信息采取了工信部统一监管和相关职能部门在各自职责范围内监管相结合的形式，但在国家颁布的《互联网信息服务算法推荐管理规定》中对于算法分级分类、算法违法行为、算法机制机理和模型的官方验证等方面尚缺少详细的规定。AI 运营商缺乏技术伦理考量体现于三方面：一是 AI 算法本身缺少技术伦理约束。案例 8 体现了算法本身没有考虑路面、天气和人员车辆等伦理因素，即便面对社会舆论质疑，技术运营商仍不愿改变算法，而是让消费者承担外卖员的时间成本。二是为了实现精准营销，AI 运营商出于各种原因将用户信息数据共享给客服外包、第三方软件开发商(ISV)或者软件开发包(SDK)等也存在伦理约束缺失[16]。第三方企业利用复杂技术算法追踪、监控、预测消费者活动，导致用户的“整合型隐私”频频泄露。例如 2020 年“新华视点”披露网络公众人脸信息遭泄露，非法黑产从业者将公众人脸照片修改为人脸识别视频后用于解封微信或支付宝账号等 [3]。三是 AI 产品应用缺乏伦理考量，案例 6 是 AI 产品销售商逃避主管部门监管，与校方直接建立使用意向，严重缺少对未成年人的伦理道德考虑。引发的技术风险是侵犯未成年人人格，引发社会公众强烈谴责。

因此，本文归纳出 AI 应用数据过度收集风险的驱动因素(如图 1 所示)。

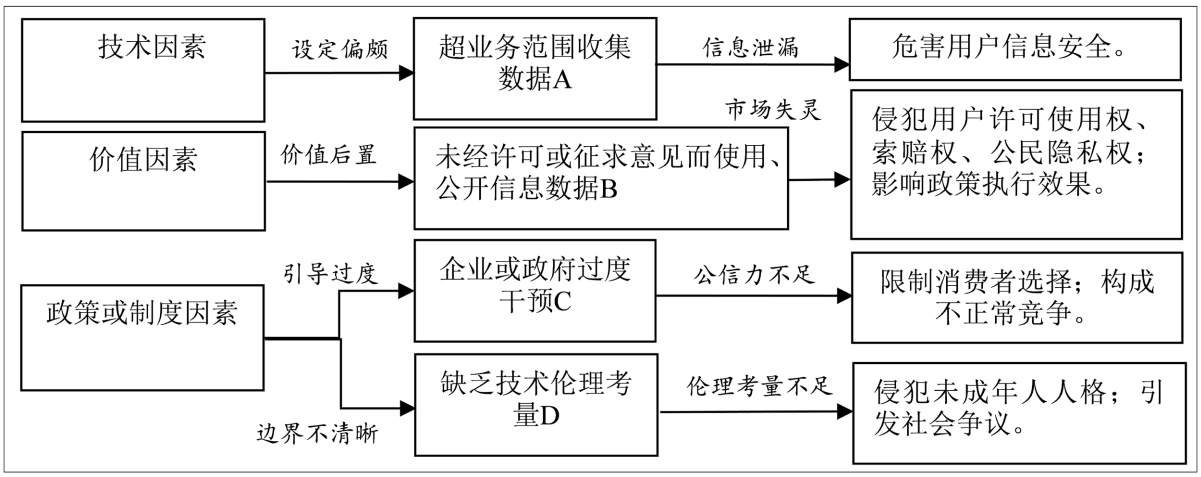

Figure 1. Drivers of excessive data collection in AI applications
图 1. AI 数据过度收集风险的驱动因素

4. AI 数据过度收集风险的治理策略

4.1. 出台违法违规使用数据行为细则及处罚办法

针对技术驱动因素引发的风险，建议在《App 违法违规收集使用个人信息行为认定方法》之下出台违法违规使用数据行为细则及处罚办法。首先，对每类违法行为进行明确定义和列举，确保违法行为得到清晰的界定和覆盖。其次，规范技术应用商知情同意条款，明确政策条款必须规定的事项，以及审查

[3]央视网. 0.5 元一份！谁在出卖我们的人脸信息[EB/OL]. [2020-07-14]. http://news.cctv.com/2020/07/14/ARTIIcp7s37vUIBn3Dx93WOO200714.shtml

剔除条款涵义不清晰、霸王条款等。然后，细化处罚办法，包括细化处罚手段、处罚金额等，确保 AI 运营商违法违规行为能够得到及时监管和纠正。最后，建议除了运营商可以提供自己没有过错的信息证明外，从上到下的数据供应链各行为主体，都需要承担一定的个人敏感信息被泄露和侵犯的法律责任[26]。

4.2. 建立 AI 数据利用社会监管机制

针对价值驱动因素引发的风险，企业可以建立 AI 数据社会监督机制。AI 创新过程是一个迭代过程，从设计(定义问题)到开发(数据采集、编程)再到实施(AI 实践并监测其性能)，任何阶段发现问题可以重新改进或弥补设计系统[27]。这个迭代过程会涉及各种社会问题，政府和企业应将关注点需要转移到公众的道德判断上，建立社会共同监管机制[28]。该机制允许企业灵活选择社会共同监管形式，一方面要求企业利用产品平台与公众进行直接互动，为公众提供表达、反馈产品意见的空间，并邀请那些可能受到负面影响的公众参与，目的是达成广泛可接受的技术改进建议[29]。另一方面要求企业与专家进行互动，针对平台反馈的社会问题，企业与专家之间需要定期或不定期召开会议论坛，与会者应能获得有关社会问题的所有必要信息和所提议的解决办法及后果。优化技术风险可以通过专家们投票来实现，要保证大多数专家意见一致，还要保证适当考虑反对专家的意见。

4.3. 完善个人信息行政公益诉讼法律途径

针对政策驱动因素引发的风险，政府需要完善个人信息行政公益诉讼法律途径。大数据时代，所有社会主体应该是“平等”“自由”地参与网络活动，但现实情况是只有政府和大型数据企业有技术和能力去控制数据资源[30]。面对政府或企业的某些网络强制行为，现行的公法保护救济制度仍难以确保社会公众权益得到保证[31]。对于大规模或情节严重的信息侵权行为，行政执法重点往往聚焦于危害国家和社会的重大信息安全事件，对个人信息保护关注不足[32]；同时个人信息领域多采取联合执法方式，存在对违法企业追责不及时、缺乏主动执法的弱点。建议完善个人信息行政公益诉讼法律途径，行政公益诉讼可以填补公法的制度空白。通过赋予检察机关或其他社会组织维护公共利益的权利来解决“公法失灵”问题，向监管部门提起行政公益诉讼可以起到重要的威慑作用，不仅可以维护个人信息公共秩序，还可以督促相关行政执法部门依法主动履职[33]。

4.4. 建立 AI 数据利用责任伦理制度

针对制度驱动因素引发的风险，建立 AI 数据利用责任伦理制度。德国哲学家约纳斯提出了著名的“责任原则”，认为科技时代无法清晰确定危害“肇事者”的身份及具体过失，需要人们自觉地意识到自己行动直接或间接导致的后果，并为此承担一种“前瞻性责任”[24]。责任伦理就是强调必须考虑行为的结果，在行动之前就需要“事先”考虑可能的后果[34]。建议政府应该强制 AI 企业建立数据利用责任伦理制度，该制度要求企业在收集数据之前、数据使用过程中、数据使用之后全流程都考虑责任伦理。责任伦理制度要嵌入到技术人员的“实地”工作和具体技术应用中[35]。将企业责任伦理制度作为工程师技术要求的一部分，作为企业技术决策提供伦理标准，将该制度置于和法规等同地位。要求企业责任伦理制度可以转化为具体的 AI 系统行为，AI 系统行为最终也应该取决最初的人为设定。例如设定超过一定存储时限后自动删除数据、限制第三方访问特定类别的数据。AI 责任伦理制度可以帮助企业设定不道德行为的关键标准，提高企业的伦理意识，有助于培育企业的伦理生态系统和文化。

基金项目

本文系 2020 年度辽宁省社会科学规划基金项目。人工智能技术社会风险预警的前瞻性政策工具研究

(重点项目，L20AGL016)。项目主持人陈瑜。

参考文献

[1] 王俊豪. 中国特色政府监管理论体系: 需求分析、构建导向与整体框架[J]. 管理世界, 2021(2): 148-164+184.

[2] 蒋洁. 人脸识别技术应用的侵权风险与控制策略[J]. 图书与情报, 2019(5): 58-64.

[3] 林伟. AI 数据安全风险及应对[J]. 情报杂志, 2022, 41(10): 105-111.

[4] Krafft, M., Arden, C.M. and Verhoef, P.C. (2017) Permission Marketing and Privacy Concerns—Why Do Customers (not) Grant Permissions? *Journal of Interactive Marketing*, **39**, 39-54. https://doi.org/10.1016/j.intmar.2017.03.001

[5] 郭海, 李永慧. 数字经济背景下政府与平台的合作监管模式研究[J]. 中国行政管理, 2019(10): 56-61.

[6] 刘素华. 论手机自动记录用户行动轨迹与个人信息保护[J]. 法学评论, 2020(5): 101-111.

[7] 顾理平. 大数据时代隐私信息安全的四重困境[J]. 社会科学辑刊, 2019(1): 96-101.

[8] Mothersbaugh, D.L., Foxx, W.K., Beatty, S.E. and Wang, S. (2012) Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information. *Journal of Service Research*, **15**, 76-98. https://doi.org/10.1177/1094670511424924

[9] 张先贵, 邱炳晟. 自动驾驶汽车数据安全风险及其治理[J]. 安徽师范大学学报, 2025, 53(5): 125-135.

[10] 和军, 李江涛. 人工智能数据风险及其治理[J]. 中国特色社会主义研究, 2024(6): 42-51.

[11] Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D. and Kleinberg, J. (2010) Inferring Social Ties from Geographic Coincidences. *Proceedings of the National Academy of Sciences*, **107**, 22436-22441. https://doi.org/10.1073/pnas.1006155107

[12] Bleier, A. and Eisenbeiss, M. (2015) Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where. *Marketing Science*, **34**, 669-688. https://doi.org/10.1287/mksc.2015.0930

[13] Acquisti, A. and Gross, R. (2009) Predicting Social Security Numbers from Public Data. *Proceedings of the National Academy of Sciences*, **106**, 10975-10980. https://doi.org/10.1073/pnas.0904891106

[14] Rafieian, O. and Yoganarasimhan, H. (2020) Targeting and Privacy in Mobile Advertising. *Marketing Science*, **40**, 193-218. https://doi.org/10.1287/mksc.2020.1235

[15] 黄蒙苏. AI 大模型训练数据的版权风险与治理路径[J]. 湖北大学学报, 2025, 52(5): 185-193.

[16] 刘裕, 周毅, 农颜清. 网络信息服务平台用户个人信息安全风险及其治理——基于 117 个 APP 隐私政策文本的内容分析[J]. 图书情报工作, 2022, 66(5): 33-43.

[17] 阮荣彬, 陈菀. 企业科技向善: 内涵、量表开发与检验[J]. 科学学研究, 2023, 41(3): 511-520.

[18] Boivie, S., Withers, M.C., Graffin, S.D. and Corley, K.G. (2021) Corporate Directors' Implicit Theories of the Roles and Duties of Boards. *Strategic Management Journal*, **42**, 1662-1695. https://doi.org/10.1002/smj.3320

[19] Koppenjan, J.F.M. and Klijn, E.H. (2004) Managing Uncertainties in Networks: A Network Approach to Problem Solving and Decision Making. Routledge.

[20] 张新宝. 个人信息收集: 告知同意原则适用的限制[J]. 比较法研究, 2019(6): 1-20.

[21] Tang, C.M. (2022) Privacy Protection Dilemma and Improved Algorithm Construction Based on Deep Learning in the Era of AI. *Security and Communication Networks*, **2022**, 1-9. https://doi.org/10.1155/2022/8711962

[22] 贺小石. 大数据背景下公民信息安全保障体系构建——兼论隐私政策的规制原理及其本土化议题[J]. 中国特色社会主义研究, 2021(6): 100-109.

[23] 张海柱. 新兴科技风险、责任伦理与国家监管——以人类基因编辑风险为例[J]. 人文杂志, 2021(8): 114-121.

[24] 任蓉. 算法嵌入政府治理的风险及其防控[J]. 电子政务, 2021(7): 31-41.

[25] 张勇, 冯明显. 数据安全刑事合规的责任伦理[J]. 河南社会科学, 2022, 30(8): 105-114.

[26] Wang, C., Zhang, J., Lassi, N. and Zhang, X. (2022) Privacy Protection in Using Artificial Intelligence for Healthcare: Chinese Regulation in Comparative Perspective. *Healthcare*, **10**, Article 1878. https://doi.org/10.3390/healthcare10101878

[27] Saltz, J.S. and Dewar, N. (2019) Data Science Ethical Considerations: A Systematic Literature Review and Proposed Project Framework. *Ethics and Information Technology*, **21**, 197-208. https://doi.org/10.1007/s10676-019-09502-5

[28] Buhmann, A. and Fieseler, C. (2022) Deep Learning Meets Deep Democracy: Deliberative Governance and Responsible Innovation in Artificial Intelligence. *Business Ethics Quarterly*, **33**, 146-179. https://doi.org/10.1017/beq.2021.42

[29] 梅傲, 李坤佳. 日本数据安全治理制度述评及其启示[J]. 情报理论与实践, 2023, 46(7): 195-200.

[30] 宋保振. 数字时代信息公平失衡的类型化规制[J]. 法治研究, 2021(6): 80-92.

[31] 丁晓东. 个人信息私法保护的困境与出路[J]. 法学研究, 2018(6): 195.

[32] 蒋都都, 杨解君. 大数据时代的信息公益诉讼探讨——以公众的个人信息保护为聚焦[J]. 广西社会科学, 2019(5): 107-115.

[33] 陈奇伟, 聂琳峰. 技术+法律: 区块链时代个人信息权的法律保护[J]. 江西社会科学, 2020, 40(6): 166-175.

[34] [德]马克斯·韦伯. 学术与政治[M]. 北京: 三联书店, 1998: 116.

[35] Raab, C.D. (2020) Information Privacy, Impact Assessment, and the Place of Ethics. *Computer Law & Security Review*, **37**, Article 105404. https://doi.org/10.1016/j.clsr.2020.105404

附　录

Table A1. Examples of open coding for AI data over-collection
表 A1. AI 数据过度收集的开放式编码示例

初始范畴	原始语句例证	对应的案例
非法抓取、使用用户信息数据	非法抓取、使用新浪微博用户信息，包括头像、名称(昵称)、职业、教育信息及用户自定义标签。	2015 年新浪微博诉脉脉(案例 1)；
超出合理范围收集用户数据	中国消费者协会早前曾做过一个调查，结果显示大量应用收集的个人信息与其实现的产品功能并没有明确关联，甚至明显超出合理范围。	2019 年中国人脸识别第一案(案例 5)；
超必要原则采集用户敏感信息	法院审理认为，在办理年卡时合同约定以指纹识别方式入园，野生动物世界采集照片信息，超出了法律意义上的必要原则，不具有正当性，且欲将其已收集的照片激活处理为人脸识别信息，超出事前收集目的，违反了目的限制原则，所以应当删除郭兵办卡时提交的包括照片在内的面部特征信息。	2021 年滴滴顺丰等 APP 手机软件案例(案例 9)。
……	……	……

Table A2. Coding of AI data over-collection behaviors
表 A2. AI 数据过度收集行为的编码结果

选择式编码	主范畴	副范畴	初始范畴	引发的技术风险
技术驱动因素	超业务范围收集数据行为 A	违法、违规收集数据 A_1 收集基本业务功能之外数据 A_2	非法抓取、使用用户信息数据 超出合理范围收集用户数据 超必要原则采集用户敏感信息	用户或非用户信息数据被迫公开，危害用户信息安全。
制度或政策驱动因素	未经许可或征求意见而使用、公开信息数据行为 B	过度公开用户及非用户信息 B_1 未经许可、未经征求意见或自行认定用户同意使用数据 B_2 不告知公众技术规范和应用目的 B_3	网络公开用户手机通讯录联系人及其相关职业、教育等信息 网络公开未经许可抄袭、复制的相关信息 未经征求意见默认用户同意收集敏感信息 未向公众充分征求技术应用意见并收集信息 用户敏感信息收集程序不规范	限制消费者选择，误解消费者；损害同行经济利益，构成不正常竞争。
	企业或政府过度干预行为 C	擅自变更收集信息技术手段、变更合同 C_1 企业干预用户自主决策数据；霸王条款 C_2 政府过度干预行为 C_3	企业强制要求公众遵守其技术规则 要求用户不得自行或授权第三方使用用户自身数据 要求用户无偿授权；企业强制公众同意隐私政策 政府自行改变技术使用规则；扩展技术场景应用范围 政府强制采集公众生物特征信息 地方政府回应方式简单粗暴	将侵犯用户许可使用权、索赔权、公民隐私权；影响政策执行效果。
价值驱动因素	缺乏技术伦理考量行为 D	技术算法 缺乏伦理考量 D_1 技术应用 缺乏伦理考量 D_2	未考虑路面情况、天气影响、人员车辆等情况的技术“精准”优化计算 技术时刻监控未成年人并转化为数据 技术应用前未能进行质量评定	侵犯未成年人人格；引发社会公众强烈谴责。