

论 AI+校园欺凌治理的隐私保护

张 寒

(格乐大学 国际学院, 泰国 曼谷 10700)

摘要: 人工智能(AI)是提升校园欺凌治理效能的重要工具,校园欺凌治理亟需 AI 赋能。但是, AI 也存在缺陷,也会产生风险。在 AI+校园欺凌治理实践中, AI 衍生的隐私侵犯风险已经出现,散布于数据收集、使用、存储、传输等环节。隐私侵犯风险的根源在于传统校园欺凌治理无法快速适应 AI 给教育管理带来的挑战,具体表现为 AI 管理不力、隐私保护法治不健全、AI 的社会接受度较低等。应当秉持积极、包容、审慎的态度,大力推进 AI 伦理建设;加强隐私安全管理,构建多元协同的风险防控机制;完善 AI 防欺凌立法,健全隐私侵犯追责机制。通过综合防控,最大限度地降低、化解隐私侵犯风险,切实保护学生的合法权益。

关键词: 校园欺凌治理;AI+;隐私;风险;防控

中图分类号: G40-058.1

文献标识码: A

文章编号: 1672-1217(2026)02-0066-09

收稿日期: 2026-01-10

作者简介: 张 寒(1993-),女,山东曹县人,格乐大学国际学院博士研究生。

DOI:10.16284/j.cnki.cn37-1401/c.2026.02.003

联合国教科文组织 2019 年 8 月发布《北京共识——人工智能与教育》,其议题和决议主要围绕着人工智能应用于教育可能存在的弱势群体歧视、数据与算法偏见、隐私侵犯、透明度缺乏等风险。^①由此可见,教育人工智能中的隐私侵犯风险已经引起世界各国的高度重视,应当着力研究化解之道。

作为校园安全管理的重要组成部分,校园欺凌治理应当客观面对人工智能(AI)衍生的隐私侵犯风险,主动适应数智时代的新变化。在校园欺凌治理工作中,秉持积极、包容、审慎的态度,既要积极应用 AI 技术,更新治理理念、改进治理方法,又要增强风险意识,科学防控隐私侵犯风险,切实保护学生的合法权益。

一、AI+校园欺凌治理的实践和效能评价

校园欺凌治理是一个古老而又常新的课题。为有效防控校园欺凌,中小学校采取了多种常规治理手段,包括建章立制、日常监管、校园安全教育、心理健康辅导、多元主体协同等,取得了一定成效。但是,无须讳言,校园欺凌治理手段大多简单粗疏、更新滞后,甚至具有负面影响。^②整体来看,许多学校主要依靠人工防控校园欺凌,“人防”是主体,“技防”是辅助,且以事后处置为基本模式,导致校园欺凌治理面临事前预防滞后、事中干预乏力、事后处理不佳等一系列难题。面对校园欺凌治理常规模式存在的被动、滞后、人力资源紧张、信息沟通不畅、协同治理困难等“瓶颈”,紧跟数智时代的发展步伐,借助 AI 助力校园欺凌治理、提升治理实效,成为广大学校的不二选择。

(一)AI+校园欺凌治理的主要举措

随着 AI 技术的不断进步、行业接受度的不断提升, AI+校园欺凌治理的步伐日益加快。不少学校主动对接当地电信部门或 AI 研发企业,通过购买 AI 技术服务或开展校企合作、政企合作等方式,利用 AI 技术赋能校园欺凌防控,做出了诸多探索。

^①张慧:《规划人工智能时代的教育:引领与跨越——解读国际人工智能与教育大会成果文件〈北京共识〉》,《现代远程教育研究》2019 年第 3 期。

^②夏雨璐、胡隽:《行为惩罚转向关系修复:恢复性司法理念下的校园欺凌治理》,《天津市教科院学报》2022 年第 5 期。

1. 智能监测校园欺凌

目前，大部分学校利用智能视频监控技术监测校园欺凌，主要有两种方式。其一，“高清网络摄像头+智能视频分析技术”，由高清网络摄像头、智能视频分析系统、存储设备、监控中心等元素构成，利用先进的视频处理技术全天候、全方位监控校园，及时发现、处理校园欺凌。其二，“现有视频监控系统+人工智能”，以现有视频监控系统为基础，科学融入人工智能技术，与学校校园安全管理平台有机结合，辅助安保、管理人员值班值守。

2. 智能识别校园欺凌

为了助力校园欺凌治理，一些 AI 研发机构着力研发智能识别技术，利用学校重点区域布设的监控摄像机，将视频流传输给 AI 智能分析服务器，自动识别监控区域出现的校园欺凌事件。某些 AI 防欺凌系统还能利用智能语音监测技术识别校园欺凌，一旦受害者发出呼救声，安装在隐蔽场所的报警器就会自动识别，实时报警。

3. 智能预警

利用 AI 推进智能预警的技术方案比较多，各有特色。有的智能平台利用 AI 智能视频分析系统、AI 智能分析网关，对视频信息进行实时分析、智能预警。有的平台可以进行区域划定，实现“人员聚集”“可疑人员逗留”等各类校园安全管理。一旦发生校园欺凌，平台会自动发出预警信号，值班人员可直接查看预警详情，并对告警位置进行三维场景定位。

4. 智能提取、保存、分析证据

证据是公平处理校园欺凌事件的基础，AI 为证据的提取、保存和分析提供了重要技术保障。有的智能安防系统可以联动监控设备，方便影像调取；联动巡护系统，实现全天候安全管控；支持局校统一部署，多维度数据统计。有的智能安防系统支持视频录像、存储、检索及回放，能够对监控区域实施 24 小时不间断录像；可以设置一定时间保存录像资料；可以通过录像资料快速锁定涉事人员、还原事发经过。某些智能安防系统还具备数据整合与管理功能，对监控视频、图片、报警记录等数据进行存储，支持多方式查询数据；能够分析、挖掘监控数据，发现潜在的校园欺凌隐患，与学校其他管理系统进行数据共享和交换。

（二）AI+校园欺凌治理的效能评价

近年来的实践探索表明，AI+校园欺凌治理不仅必要，而且可行，有利于事前预防，也有利于事中处置和事后处理。整体分析，AI+校园欺凌治理取得了不少成就，积累了一定经验，但是，也存在很多不足。

1. 成就

其一，AI+校园欺凌治理降低了预防的难度。大部分学校利用 AI 技术赋能既有的监控设备，或新安装 AI 监控探头，或布设 AI 语音报警终端，由人工监测逐步过渡到智能监测，提高了发现校园欺凌的效率。常规监测手段效率较低，无论是人工巡查还是常规监控探头，耗时长，漏检率高，AI 防欺凌系统通过智能监控、语音检测等技术极大提升了监测效能。常规监测手段的盲区多、死角多，导致部分欺凌行为难以被发现，AI 防欺凌系统有效填充了空白，突破性地实现了校园欺凌监测全覆盖。

其二，AI 技术有效拓展了处置方式，提升了处置效率。AI 防欺凌系统通过 AI 语音报警终端等智能设备进行语音警告或声光震慑，可以第一时间制止校园欺凌，实现了实时干预。实践证明，这一技术创新有效控制了绝大多数校园欺凌，降低了可能造成的危害。高效的信息传输是 AI 防欺凌系统的另一技术优势，借助于该优势，学校管理人员可以精准定位校园欺凌发生的地点，准确掌握其场景，为快速干预提供技术支持。

其三，AI 技术助力学校收集证据、固定证据，保证事后处理合规、有据；同时，部分 AI 防欺凌系统具备分析证据、提供决策参考的功效。目前，绝大多数 AI 防欺凌系统具备实时录音、录像、云存储等功能，不仅可以智能收集证据，而且可以智能固定证据、保存证据，为后续的公平处理奠定了坚实基础。人工智能有

助于教育部门建立数据平台，基于人工智能分析后的有关信息进行数据决策，^①个别学校上新的高端 AI 防欺凌系统具备数据分析、辅助决策功能，可以智能分析平台收集、存储的相关信息，提供可行性校园欺凌治理建议。

2. 不足

在智能识别方面，AI 防欺凌系统依然存在短板。从目前投入使用的 AI 语音报警终端、校园智能监控系统来看，对于较为隐形的、或形式特别的校园欺凌行为，其识别能力还比较弱。个别 AI 防欺凌系统识别校园欺凌的精准度欠佳，某些欺凌行为没有被准确识别出来。

在智能干预方面，干预的方式、方法、力度等有待改进。目前，大部分 AI 防欺凌系统不具有实时干预功能，通常方式为平台向管理人员发出预警信息，相关人员赶赴现场进行人工干预。部分平台通过前端监测设备或者联网校园广播，可以实施语音警告或声光震慑等干预措施，但是，干预的方式、方法较为单一，制止校园欺凌的力度比较弱。

在辅助决策方面，AI 防欺凌系统较好解决了证据保全问题，但是，证据分析及辅助决策的水平一般。大部分 AI 防欺凌系统的数据分析能力比较弱，难以满足校园欺凌治理工作多样化的需求；部分系统虽然具备一定的数据分析能力，但是，算力有限，仅能解决一些简单问题。对于较为复杂的校园欺凌治理宏观分析、远景设计、风险预测、治理决策等，无力提供具有较高参考价值的方案或建议。

二、AI+校园欺凌治理衍生的隐私侵犯风险

智能化技术在优化产业结构促进经济发展同时也给普通大众带来隐私保护、数据控制、知识产权以及道德伦理等新类型的风险。^②在 AI+校园欺凌治理领域，存在同样的问题。随着 AI+校园欺凌治理的不断深入，其衍生的隐私侵犯风险已经开始显现。

AI+校园欺凌治理以大数据为技术支撑，在 AI 防欺凌模型构建、校园欺凌防控等实际工作中，需要采集、分析海量的学生信息。“人工智能系统对数据的大规模收集和分析可能侵犯个人隐私权，用户在不知情或未明确授权的情况下，其敏感信息可能被用于商业目的等，侵犯其自主控制个人信息的权利。”^③

从技术层面分析，在 AI+校园欺凌治理工作中，隐私侵犯主要呈现出如下表征：过度采集与无边界监控、数据的永久性存储与滥用、算法黑箱与不可解释性、生物识别技术的强制应用等。这些表征既反映了技术本身的特性，也暴露了技术应用中的伦理失范，对于广大学生的生物识别信息、个人身份与联络信息、心理与行为数据、学习与学业信息、社交关系与通信隐私等构成严重威胁。近年来，侵犯学生隐私事件时有发生，主要分散在数据收集、使用、存储、传输等多个环节。

（一）数据收集环节，过度收集学生信息

在数据收集方面，AI 防欺凌系统的正常运行需要收集大量的学生个人信息，包括学生的面部特征、语音信息、行为数据等，以实现对欺凌行为的监测和分析。个别学校存在过度收集数据、不尊重学生知情同意权等问题，客观上侵犯了学生隐私。

有的学校在学生宿舍、厕所等区域广泛安装 AI 语音监控设备，全天候、无选择地采集各种声音数据，导致场景覆盖泛化，极易侵犯学生隐私，也引发学生对“如厕对话被记录”的担忧。有的学校超范围采集学生数据，不仅收集学生的基本信息，还通过 AI 系统采集学生的生物特征数据，如指纹、声纹、人脸等。这些生物特征数据具有唯一性和不可更改性，一旦泄露，将对学生的隐私和安全构成威胁。2021 年，广东珠海某小学与当地一家智慧校园服务商合作，强制性收集全校师生的人脸识别信息，学生进出校园需经人脸扫

①杨晓哲、任友群：《教育人工智能的下一步——应用场景与推进策略》，《中国电化教育》2021 年第 1 期。

②宋华健：《论生成式人工智能的法律风险与治理路径》，《北京理工大学学报(社会科学版)》2024 年第 3 期。

③张莉莉、华志宇：《生成式人工智能的法律风险及其规制》，《行政与法》2024 年第 4 期。

描，引发了家长对学生隐私安全的担忧。个别学校的 AI 系统不仅分析校园监控视频，还整合学生社交媒体动态，收集与欺凌无关的私人聊天记录，超出校园欺凌治理的客观需要。

在数据收集过程中，部分学校未充分保障学生的知情同意权。根据《个人信息保护法》，收集个人信息应当向个人信息主体明示收集、使用信息的目的、方式和范围，并取得其同意。在实际操作中，不少学校未提前告知学生数据收集的目的、范围、方式、存储期限以及可能的共享对象等信息，也未取得学生的明示同意。有的学校仅在校园公告栏、学生家长群简单告知相关信息，难以保障学生的知情权，也不等于获得了其明示同意，不仅侵犯了学生的知情权、隐私权，还容易导致学生及家长对学校产生信任危机。

（二）数据使用环节，滥用学生信息

AI 防欺凌模型研发是 AI+校园欺凌治理的一项重要基础工作，AI 研发机构、AI 服务商需要使用大量的学生个人数据，用于模型设计、训练及验证。个别研发机构和服务商为了降低成本、快出成果，漠视法律规定，无视学生的正当权益，利用与学校合作之便利，将自己掌握的学生个人信息（包括未脱敏数据）直接用于模型训练，侵犯了学生隐私。

有的 AI 服务商未经学校及学生授权，将学生隐私信息共享给第三方机构或个人，甚至用于与校园欺凌治理无关的广告推送、市场调研等商业活动，背离了数据收集的初衷。某 AI 科技公司将从学校获取的学生学习数据和行为数据出售给第三方——某教育培训机构，该机构获取学生隐私信息后频繁地向学生及其家长推送培训课程广告，有针对性地拓展生源、对接学生，不仅干扰学生和家长的正常生活，而且严重侵犯学生隐私。

（三）数据存储环节，安全防护不到位

在数据存储方面，存储设备的物理安全至关重要。如果存储设备被盗或损坏，很容易导致学生隐私信息丢失或泄露。目前，某些学校和 AI 服务商使用的存储设备安全性较低，容易受到病毒攻击、黑客入侵和物理损坏，学生隐私信息泄露的安全风险较高。

一些学校和 AI 服务商的数据安全防护措施不够完善，导致学生隐私信息面临泄露风险。有的学校将学生隐私信息存储在安全性较低的普通服务器上，未进行必要的加密和防护处理，一旦服务器受到攻击，黑客很容易获取学生隐私，隐私安全面临极大风险。有的学校对数据存储权限管理不当，未经授权的人员可以轻而易举地访问、获取学生信息，客观上增加了学生隐私泄露风险。如，某中学数据存储权限设置不够合理，安全级别较低，普通工作人员即可查看、下载学生信息，包括学生的敏感信息，严重威胁学生隐私安全。

（四）数据传输环节，网络安全性不佳

在 AI+校园欺凌治理过程中，数据传输存在诸多安全隐患。数据传输有可能受到网络攻击，如数据劫持、中间人攻击等，导致学生隐私信息被窃取或篡改。

数据传输离不开网络，网络的安全性、稳定性关乎学生信息安全。如果数据传输使用的网络不够稳定，有可能导致数据传输中断或数据丢失，危及学生信息安全。有的学校或 AI 服务商不注重数据传输过程中的网络安全防护，安全漏洞较多，给恶意软件入侵留下了可乘之机，容易造成学生隐私信息泄露。一些学校和 AI 服务商在传输学生数据时未采用加密传输方式，使学生隐私信息在传输过程中被窃取或篡改。如，某中学使用 HTTP 协议传输学生隐私信息，该类协议不具备加密功能，属于不安全的网络协议，使用该类协议传输的相关数据容易被黑客拦截、篡改。

三、隐私侵犯风险的成因

AI+校园欺凌治理衍生隐私侵犯风险的根源在于传统校园欺凌治理无法快速适应 AI 给教育管理带来的新变化，具体表现为 AI 管理不力、隐私保护法治不健全、AI 的社会接受度较低等多个方面。

（一）AI 管理不力

1. AI 管理能力欠佳

有的学校虽然引入了 AI 防欺凌系统，但是，对于 AI 技术的复杂性和潜在风险认识不足，缺乏有针对性的管理策略，依然沿用传统的管理手段、管理模式。由于没有重视 AI 技术的特殊性，造成 AI 防欺凌系统难以融入校园欺凌治理格局，导致校园欺凌治理工作漏洞频出，为学生隐私保护工作埋下隐患。

在学生个人信息收集、存储、使用、共享等方面，部分学校缺乏有效的数据管理措施，主要表现为数据管理规范缺失、数据监管手段落后。有的学校违规收集学生信息，既不明确告知收集数据的目的，也不采取应有的数据保护措施。个别学校管理人员甚至将学生个人信息转手给第三方，造成学生敏感数据泄露，侵犯学生隐私。

在 AI 防欺凌系统的运维管理方面，大部分中小学校缺乏专业技术人员和完善的运维制度。一旦 AI 防欺凌系统出现技术问题，故障往往难以得到及时修复，系统安全面临重大威胁，有时甚至可能陷于瘫痪，给学生隐私安全带来极大风险。

2. 隐私保护责任意识淡薄

一些学校对于 AI 防欺凌系统可能引发的隐私侵犯风险认识不足，重视不够。在实际工作中，漠视风险隐患，忽略风险预防。个别 AI 服务商无视企业应当承担的法律责任、社会责任，超范围使用学生敏感信息，甚至将其转入商业领域。

在责任承担方面，个别学校和 AI 服务商不敢担责、不愿担责。当 AI 防欺凌系统出现隐私侵犯等问题时，学校和服务商没有把解决问题的重点放在如何保护学生隐私、改进管理上，而是拘泥于摆脱责任、保全自身。有时，双方各执己见，相互推诿于己不利的法律后果，不仅损害学生合法权益，也降低了 AI+校园欺凌治理的社会公信力。

（二）隐私保护法治不健全

1. 相关法律法规不够完善

针对人工智能，我国尚未颁布专门的法律，而是由《网络安全法》《数据安全法》《个人信息保护法》《科学技术进步法》《互联网信息服务深度合成管理规定》等法律法规构成监管规则体系。^①具体到 AI+校园欺凌治理领域的学生隐私保护，近年来颁布实施了《人脸识别技术应用安全管理规定》等法律文件，但是，总体来看，依然不够健全。

对于 AI+校园欺凌治理中的隐私保护，现有法律法规为其确立了基本的法律遵循。但是，只能作为原则性指导，难以细化为具有较强针对性的法律规范。因此，面对 AI 技术在校园欺凌治理中的复杂应用场景，现有法律法规存在不少模糊地带，在实践中暴露出许多漏洞和不足，难以有效规范 AI 技术应用，无法充分保护学生的隐私权。

2. 政府部门监管不力

人工智能带来了类别和程度不同的风险，监管人工智能也要做出相应的分类，并采取相应的对策。^②由于 AI+校园欺凌治理尚处于起步阶段，政府主管部门对 AI 在校园欺凌治理中的应用尚未实施有针对性的监管，导致隐私侵犯等风险难以得到有效遏制。

政府部门监管职责划分不明晰。按照属地管理原则，地方政府下设的教育、公安、网信、质监等部门均有权对 AI+校园欺凌治理实施监管。但是，对于其中的隐私保护，现行法律法规并未明确设定具体的监管部门，造成部门之间分工不清、实际监管不到位。

监管标准和评估体系缺失。针对 AI 防欺凌系统的可靠性、安全性，政府部门尚未制定、出台明确的监

^①徐磊：《发展与安全并重：生成式人工智能风险的包容审慎监管》，《理论与改革》2024 年第 4 期。

^②谢新水、袁汝兵：《论人工智能监管底线的设置：基于主体性的探究》，《学习论坛》2020 年第 8 期。

管标准和评估方法。学校在引入、使用 AI 防欺凌系统时,无法判断该系统是否符合法定要求,难以对其进行科学、全面的评估,给隐私侵犯等风险留下了可乘之机。

监管手段落后,制约了政府部门对 AI 技术风险的有效监管。当前, AI 技术快速迭代升级,新的应用场景、新的技术形态不断涌现。但是,政府部门的监管手段难以与时俱进,导致政府监管落后于 AI 技术进步。人工检查、文件审查等依然是政府监管的主要方式,无法对 AI 防欺凌系统的运行过程进行实时、动态监管,客观上放纵了隐私侵犯行为的发生。

(三)AI 的社会接受度较低

1. 传统文化不利因素的影响

我国的传统文化一直重视发展教育,强调立德树人,关注学生健康成长。传统文化与 AI+校园欺凌治理的价值取向具有高度一致性。但是,这并不意味着传统文化对校园欺凌治理的影响都是积极的,某些落后思想、价值追求等具有消极作用,一定程度上催生了隐私侵犯等风险。

传统文化重视教育,但是,在德、智、体、美、劳等方面,更加侧重于智育。这种价值取向不利于 AI+校园欺凌治理工作,导致某些学校不重视 AI+校园欺凌治理,人、财、物资源的投入较少,隐私侵犯等风险随之增加。

在传统文化中,集体主义、权威主义等观念一直备受尊崇。但是,如果在 AI+校园欺凌治理工作中过于强调这些观念,一定程度上会阻碍学生个体合法权益的保护。在校园欺凌事件处置方面,某些学校没有把保护学生隐私权和身心健康放在第一位,而是将维护学校声誉和形象放在优先位置,隐私侵犯事件再次发生的几率随之上升。

2. 社会认知的偏差

社会各界对于 AI 技术的认知和态度直接影响公众对于 AI+校园欺凌治理的接受度。目前,存在两种较为片面的认知,影响了公众接受度,无谓地增加了隐私侵犯等风险。

一些社会成员过于崇拜、迷信 AI 技术,认为 AI 防欺凌系统法力无边,能够解决校园欺凌的所有问题,忽视 AI+校园欺凌治理可能衍生的隐私侵犯等风险。个别人盲目相信 AI 技术的准确性、可靠性,将校园欺凌治理的所有希望寄托于 AI,否定常规治理手段的重要性,忽视人的主导作用。

与之相反,部分公众对 AI 技术知之甚少,接受度较低,对 AI+校园欺凌治理疑虑过重。有的担心 AI 技术侵犯学生隐私, AI 应用于校园欺凌治理会导致学生的一举一动都暴露在监控之下。部分家长、学生对于学校引入 AI 防欺凌系统难以接受,客观上成为排斥 AI+校园欺凌治理的阻力。

3. 公众参与度较低

AI+校园欺凌治理离不开学生、家长、社会各界的参与和支持,尤其在事关学生隐私权等集体利益的事项上,应当广泛听取社会各界的意见,集思广益,民主决策,充分保护学生的合法权益。

目前,在校园欺凌治理实践中,学生的主体价值尚未得到充分认识,导致学生在校内欺凌风险治理中成为“沉默的一群人”。^①在防控隐私侵犯风险方面,一些学校不重视征求学生的意见、倾听学生的呼声,认为学生仅仅是校园欺凌治理的对象,无所谓参与权。学生对 AI 防欺凌系统的感受和需求难以得到学校的足够重视,学生提出的关于隐私保护的意见建议往往被忽视。

作为学生的监护人,家长有权利、有义务保护孩子的隐私。但是,在实践中,家长的参与渠道和机会比较少。大部分学校引入 AI 防欺凌系统时,没有征求过学生家长的意见。个别学校根本不通知家长,更不让家长参与,导致不少家长对学校的 AI 防欺凌系统一无所知。在学生隐私保护方面,不少家长根本没有机会提出合理化建议。

^①张旭光:《落差与纠偏:从风险思维视角探解校园欺凌治理之难》,《内蒙古师范大学学报(教育科学版)》2024年第4期。

四、防控隐私侵犯风险的主要路径

2017年,我国教育部等十一部门联合印发《加强中小学生欺凌综合治理方案》,提出应对校园欺凌的四大基本原则,即“坚持教育为先,坚持预防为主,坚持保护为要,坚持法治为基。”^①在AI+校园欺凌治理工作中,防控隐私侵犯风险是重要一环,应当从管理、法治、伦理等多角度综合施策。

(一)加强隐私安全管理,构建多元协同风险防控机制

1. 政府部门协同

教育部门在AI+校园欺凌治理工作中起到统筹、协调作用,根据本地实际,制定AI技术在校园欺凌治理中的应用规范,明确数据采集范围,避免过度采集学生个人信息,保护学生隐私。各地公安机关大都设立了“一键报警”平台,对接校园AI预警系统,确保第一时间予以响应,及时制止欺凌行为,保护学生隐私。网信部门负责监管学校的网络安全、信息安全,通过日常监督检查等方式,保证学校依法、依规开展AI+校园欺凌治理,有效防范隐私侵犯、网络入侵等风险。司法机关在法治宣传方面具有独特优势,可以组织法学专家、律师等法律专业人员进学校、进教室,通过法治讲座、法治公开课等形式,宣传、普及校园欺凌防控方面的法律知识,帮助学生认清校园欺凌的危害,增强自我保护个人隐私的意识。

上述各部门分工与协作并举,各司其职、各负其责,共同为学生织就一张宏大的隐私保护网。在监管路径上保持包容审慎的态度,实施分类分级监管,制定相应规则和指引,确保不同级别的服务得到适当的监管与指导。^②

2. 家、校联动

消除校园欺凌风险隐患仅靠政府、学校的力量远远不够,必须打破单纯依靠政府和学校的校园欺凌防控路线,激发多元主体参与校园欺凌风险治理的内生动力,形成异质社会资源优势互补的“多元协同”风险治理共同体。^③在AI+校园欺凌治理工作中,家、校联动可以整合学校、家庭的资源,有效防控隐私侵犯等风险,共同守护学生身心健康。

学校是学生学习的主要场所,应当有计划地组织、开展隐私保护教育,提高学生自我防范意识。同时,学校还应当建立较为完善的快速响应机制,一旦发生隐私泄露事件,学校能够第一时间启动应急预案,及时为学生提供心理辅导和支持,切实保护学生的合法权益。学校应当加强与家长的沟通与合作,定期交流学生的学习、生活情况,共同关注学生的心理和行为变化,及时发现、化解隐私侵犯风险。

家庭是孩子生活的港湾,家长应当通过言传身教等方式对孩子进行道德引领和心理健康辅导,帮助孩子树立正确的人生观、价值观,培养孩子的同情心、责任感,不断增强自我保护意识。对于孩子的异常行为、情绪变化,家长应当多加关注,及时与学校沟通。一旦发现隐私侵犯,双方应当共同研究干预方案和解决问题的办法。

(二)完善AI防欺凌立法,健全隐私侵犯追责机制

长期以来,我国校园欺凌治理多依靠教育主管部门的指导意见或要求,多涉及校园安全管理或加强学生思想道德教育等宽泛问题,缺乏有针对性的校园欺凌专项立法以明确欺凌与暴力的边界及相对应的惩治办法。^④为有效防控隐私侵犯等风险、提升AI+校园欺凌治理实效,制定专项法律文件、完善追责机制,实属必要。

1. 制定AI防欺凌专项法律文件

①《教育部等十一部门联合印发〈加强中小学生欺凌综合治理方案〉》,《中国应急管理》2017年第12期。

②徐伟、李文敏:《生成式人工智能法律风险的源头治理》,《行政与法》2025年第2期。

③付玉媛、韩映雄:《多元主体参与:英国校园欺凌治理实践与启示》,《比较教育学报》2021年第4期。

④李青:《“生态治理”:新西兰校园欺凌治理行动探赜》,《新生代》2024年第1期。

目前,我国已经制定实施了一系列与 AI+校园欺凌治理相关的法律、法规、规章,如《网络安全法》《数据安全法》《个人信息保护法》《未成年人保护法》《未成年人网络保护条例》《生成式人工智能服务管理暂行办法》《人脸识别技术应用安全管理办法》等,为 AI+校园欺凌治理奠定了坚实的法律基础。但是,鉴于 AI+校园欺凌治理工作刚刚起步,立法机关尚未将 AI 防欺凌立法纳入立法规划。

为有效防控 AI+校园欺凌治理衍生的各类风险,制定专项的 AI 防欺凌法律文件已经成为必要。立法机关应当未雨绸缪,以问题为导向,启动 AI+校园欺凌治理专项立法调研,厘清实际工作中存在的主要问题及风险,尽快制定《AI 防校园欺凌管理条例》。

《AI 防校园欺凌管理条例》应当科学设定 AI 技术在校园欺凌治理中的应用边界,重点化解某些学校强制学生刷脸、滥用智能监控设备等引发的隐私侵犯风险。在数据采集与使用方面,严格落实《个人信息保护法》,明确 AI 防欺凌数据采集应当遵循合法、正当、必要原则,且需获得学生的明示同意。在数据存储方面,要求学校、AI 服务商采用数据加密技术,严防学生隐私泄露。对于采集的学生音频、视频数据,应当采用国密算法进行加密,切实降低隐私泄露风险。

2. 健全隐私侵犯追责机制

隐私侵犯法律责任的归属应当坚持过错原则,谁对隐私侵犯存在主观过错(包括故意和过失),谁就应当承担于己不利的法律后果。作为数据采集、使用、管理的主体,学校应当对于隐私侵犯承担主要责任。作为数据的处理者,如果 AI 服务商提供的 AI 防欺凌系统存在安全漏洞,或未履行应有的安全保障义务,导致学生隐私被侵犯,AI 服务商也应当承担相应的法律责任。

应当加大对隐私侵犯的惩处力度,确保法律责任落到实处。追责的重点有两个方面:一是学校、AI 服务商因为管理不善导致隐私侵犯,造成学生权益受损,而引发的法律责任;二是政府主管部门怠于监管、失职失责而引发的法律责任。前者主要表现为民事赔偿责任;后者表现为行政责任、监察责任,根据案件性质,由行政机关或监察机关依法、依规给予责任人行政处分或政务处分。

(三)大力推进 AI 伦理建设,平衡技术效能与权利保护

1. 加强 AI 伦理意识培养

科学构建由学生、教师等多方参与的伦理意识培养体系,意义重大。通过对教师、学生等参与方的教育,可以全面提升社会各界对 AI 伦理的认知水平,引导各方理性对待 AI 技术在校园欺凌治理中的应用,有效平衡技术效能与权利保护。

教师是 AI+校园欺凌治理的参与者、实施者,应当通过专题讲座、继续教育等方式对教师开展 AI 伦理教育。可以将 AI 伦理增设为教师继续教育的一项内容,涵盖隐私保护、数据安全、算法偏见识别等,帮助教师系统掌握 AI 伦理知识。在隐私保护方面,教师应当了解如何保护学生隐私,确保数据存储、传输、使用符合安全规范。

学生是 AI+校园欺凌治理的直接参与者、利益关联方。可以在中小学信息技术课中增设“AI 与隐私保护”模块,帮助学生掌握 AI 基本理论、运作机制,了解联邦学习、数据脱敏等知识,学会如何保护个人隐私。

2. 构建伦理审查机制

构建伦理审查机制有利于规范 AI 技术在校园欺凌治理中的应用,有利于防控隐私侵犯等风险、保障学生健康成长。面对伦理法规的“硬法缺失”,呼吁立法机构建立并完善伦理准则、规则和问责机制,以明确人工智能相关主体的权责边界,全面保障各方的法定权益。^①在制度设计上,校园 AI 伦理委员会、第三方评估与认证等至关重要,可以确保 AI 技术应用合法、合规,为有效保护学生隐私提供强有力的支撑。

校园 AI 伦理委员会一般由法律专家、教育工作者、家长代表等组成,应当赋予其较大的审查权。学校拟采用的 AI 技术方案应当通过校园 AI 伦理委员会的隐私影响评估(PIA)与公平性测试,确保 AI 技术应

^①张铤:《人工智能的伦理风险治理探析》,《中州学刊》2022 年第 1 期。

用不会损害学生隐私权等合法权益。动态监督也很重要，校园 AI 伦理委员可以定期、不定期评估 AI 系统的运行效果，灵活调整监控策略，根据实际需要及时优化 AI 技术应用。

为了保证 AI 防欺凌系统认证的公正与公平，为 AI 技术应用提供客观评价和监督，学校应当引入第三方评估与认证机制。在评估与认证标准体系上，技术认证是基础，学校采用的 AI 防欺凌系统应当符合校园 AI 伦理认证标准；持续审计不可或缺，有利于及时发现、纠正 AI 技术应用中的隐患，及时化解隐私侵犯等风险，切实保护学生的合法权益。

Analysis on Privacy Protection in AI + School Bullying Governance

ZHANG Han

(International College of Krirk University, Bangkok 10700, Thailand)

Abstract: Artificial intelligence (AI) is an important tool to improve the efficiency of school bullying governance, and school bullying governance is in urgent need of AI empowerment. However, AI also has defects and can give rise to risks. In the practice of AI + school bullying governance, the risks of privacy infringement derived from AI have emerged, spreading across links such as data collection, use, storage, and transmission. The root cause of privacy infringement risks lies in the fact that traditional school bullying governance cannot quickly adapt to the challenges brought by AI to educational management, which is specifically manifested in inadequate AI management, imperfect legal system for privacy protection, relatively low social acceptance of AI, etc. We should adopt a positive, inclusive and prudent attitude, vigorously promote the construction of AI ethics; strengthen privacy and security management, and build a multi-stakeholder collaborative risk prevention and control mechanism; improve legislation on AI-based anti-bullying, and perfect the accountability mechanism for privacy infringement. Through comprehensive prevention and control, we can minimize and resolve the risks of privacy infringement, and effectively protect the legitimate rights and interests of students.

Keywords: school bullying governance; AI+; privacy; risk; prevention and control

[责任编辑 治平]