国际DOI编码:10.15958/j.cnki.gdxbshb.2026.03.05

数字社会人工智能与未成年人犯罪预防深度融合研究

——基于融合智能体重构犯罪风险预警机制的视角

魏 红,张 桐
(贵州大学 法学院,贵州 贵阳 550025)

摘 要:在当今数字社会中,未成年人犯罪行为日益向网络化、隐蔽化方向发展。由于传统未成年人犯罪预防系统预测犯罪风险的感知能力与解析能力较弱,融合智能体便应运而生。融合智能体由分析智能体、风险预测智能体和自适应新风点拓扑智能体三个子智能体构成,这三个子智能体相互协同联动,成为未成年人犯罪风险预警机制的核心驱动力。融合智能体通过聚焦未成年人犯罪行为从"风险萌芽—临界演化—爆发实行"的动态演化,采用"六类四级"法划分行为并进行风险评估分级;以"提前识别、精准干预、阻断升级"为预警目标,运用柔性引导与刚性约束等方法相结合的精准处遇方式,构建"一般预防—临界干预—再犯预防"的三重预防模式,形成政府主导、多元共治的未成年人犯罪风险预警机制,为传统社会治理向精准化、数智化转型探索科学方法和实践新路径。

关键词:未成年人犯罪预防;融合智能体;犯罪风险预警;数智化社会治理

中图分类号:D912.7;D917.6 **文献标识码:**A **文章编号:**1000-5099(2026)03-0061-13

一、问题的提出

我国高度重视未成年人的成长环境,并为未成年人健康成长提供积极的法律保护和多重社会保障。随着经济社会快速发展,近年来我国未成年人犯罪呈高发态势。根据最高人民检察院2021至2025年期间公开发布的数据显示,各地检察机关批准逮捕的未成年犯罪嫌疑人数近五年增长了50.0%,年均复合增长率8.5%;各地检察机关起诉的未成年犯罪嫌疑人数近五年增长71.2%,年均复合增长率11.4%①。从最高人民检察院发布的近五年我国未成年人犯罪数据看出,当前我国未成年人犯罪率居高不下,未成年人犯罪形势仍然不容乐观,亟须创新治理手段应对挑战。

在数字社会背景下,未成年人犯罪的犯罪方式向网络化、虚拟化方向发展。然而,预防未成年人犯罪的传统应对机制在数据维度立体化、动态响应即时性等方面,越来越难以适应现代社会对犯罪风险精准定位和提前干预的需求。"犯罪学学科任务是寻找犯罪行为发展的客观规律,进而实现最大限度地犯罪预防"[6]。因此,如何运用人工智能辅助工具重构未成年人犯罪预警机制,实现犯罪预防与人工智能的深度融合,不仅具有重要的理论探讨价值和司法实践意义,更是亟待解决的社会现实问题。

收稿日期:2025-01-18
基金项目:中国青少年研究会重点项目"智能体交互协同框架下性侵害未成年人犯罪被害预防实证研究"(2026A08);贵州省人民检察院检察理论研究项目"性侵害未成年人犯罪精准化被害预防与治理研究"(GZJC2025B19)。
作者简介:魏 红,贵州大学法学院教授、硕士生导师;
张 桐,贵州大学法学院硕士研究生。

①根据最高人民检察院2021年至2025年相断发布的《未成年人检察工作白皮书》中有关数据计算得知:2020年全国检察机关批准逮捕未成年犯罪嫌疑人22 902人,提起公诉33 219人[1],2021年全国检察机关批准逮捕未成年犯罪嫌疑人数增至27 208人,较2020年同比增长18.8%,提起公诉35 228人,较2020年同比增长6.0%[2];2022年全国检察机关批准逮捕未成年犯罪嫌疑人数15 462人,较2021年同比下降43.2%,提起公诉27 679人,较2021年同比下降21.4%[3];2023年全国检察机关批准逮捕未成年犯罪嫌疑人数增至26 855人,较2022年同比增幅高达73.7%,提起公诉人数增长至38 954人,较2022年同比增幅达40.7%[4];2024年全国检察机关批准逮捕未成年犯罪嫌疑人数持续增长至34 329人,较2023年同比上升27.8%,提起公诉人数增长至56 877人,较2023年同比上升46%[5]。

二、数字社会未成年人犯罪的特征:网络化与隐蔽化

随着科学技术的迅猛发展,我国由传统经济社会向数字经济社会转变。“数字技术正以狂飙突进的方式重塑人们的生产方式和生活空间”[7],这无疑“显示出我国数字经济时代的到来、发展大幕已经开启”[8]。数字技术的发展对提高未成年人的学习能力、促进其成长具有积极性意义,但是数字技术也给未成年人的认知方式和行为方式带来巨大的冲击和影响。

(一)未成年人犯罪多发生于网络空间

犯罪形态数字化是当前未成年人犯罪的一个重要特征。司法部门公开发布的数据显示,2020—2022年期间,在各地司法部门审结的帮助信息网络犯罪活动罪案件中,未成年犯罪人数2022年同比增幅高达82.41%,三年增长近23倍[2];2024年全国检察机关受理审查起诉涉诈骗犯罪的未成年人人数,占当年受理审查起诉未成年犯罪嫌疑人总数的10%,较2023年同比上升1.86个百分点。在未成年人参与的诈骗犯罪中大量涉及网络诈骗[9],这类犯罪行为呈现专业化与智能化特点。在2023年具有较大社会影响的“人肉开盒案”[10]中,未成年犯罪嫌疑人通过网络爬虫技术非法获取200余名主播的个人信息,并形成覆盖18个省(市)的网络犯罪产业链,其犯罪行为和手段已经逐渐形成从传统物理空间向数字生态系统性迁移的表现[11]。

数字技术的进步与深入运用,“进一步导致未成年人能够接触到更多的信息,发展出更多样化的社会认知与行为方式,在较低年龄表达出更高年龄所具有的观念与行为,社会的发展加速与个体的发展加速在一定程度上出现了耦合趋势”[12],尤其是近年来青少年各类性行为发生率呈现上升趋势[13]就是这种典型表现。在影响未成年人成长的诸多因素中,数字社会中各类大模型算法的推荐机制,使得未成年人在虚拟数字空间中能够轻易地获得即时性满足。因此,有相当一部分未成年人不愿意通过现实社会艰苦努力以获得社会认可和肯定的社会化路径。故而,从某种角度而言,信息技术的升级在一定程度上挤压了未成年人获得“合适的社会化”的机会,对未成年人在社会化过程中各类偏差行为发生概率的升高具有强相关性影响[14]。

(二)未成年人犯罪手段趋向隐蔽化

信息技术升级不仅影响社会生产结构与模式的变化,还催生了未成年人网络犯罪的新形态。随着各类开源工具、AI技术升级,它们在为社会发展提供新质生产力支持的同时,也降低了未成年人实施网络犯罪的门槛。在犯罪场域虚拟化的背景下,元宇宙、游戏平台等虚拟空间也逐渐成为新型犯罪场域,增长迅速的虚拟财产盗窃、网络欺凌等行为往往游离于传统监管之外。随着网络犯罪产业链条化,犯罪分子在网络平台上可以轻松地实现犯意教唆、工具交易、犯罪实行和赃款洗白的完整犯罪过程,而作为“网络原住民”的未成年人也更容易被吸纳为犯罪的“工具人”而深陷犯罪其中。由于未成年人身心处于快速发展阶段,对新兴事物具有强烈的好奇心和易于接受新事物的特点。因此,相较于成年人犯罪,在未成年人实施的犯罪中,受技术升级驱动的特征也更为明显和突出,并呈现出三重结构性变化:一是犯罪工具从日常化向智能化发展[15],爬虫软件、钓鱼程序等技术工具被大量运用于敲诈勒索、数据窃取等犯罪行为中;二是犯罪链条从显性化向隐匿化转变,近年来在未成年人实施的犯罪中运用或通过虚拟货币、暗网论坛等助力洗钱和分工协作的案件逐渐增多[16];三是犯罪类型从传统化向新型化升级[17],近年来深度伪造、破解人脸识别认证等新型犯罪行为在未成年人实施的犯罪中屡屡出现,这些相较于传统犯罪也更加难以发现和预防。

预防未成年人犯罪的传统机制主要聚焦于线下人身暴力、侵财类个体性犯罪,难以适配数字社会中高新科技与产业链分工结合的新型复杂犯罪形态。传统的未成年人犯罪预防重事后惩罚、轻事前监管,难以实现精准化预防和事前干预。同时,从技术层面看,传统的未成年人犯罪预防更多地依赖采用人工或半人工分析工具来分析静态结构化数据,与当前犯罪分子大量运用加密、网络爬虫等新型技术和手段形成技术代差壁垒;在治理协同性上,传统的未成年人犯罪预防存在着治理主体权责分散、信息割裂等问题,预防协同性不足。当人们寄希望于通过单纯的技术升级就可以实现风险控制和良好的社会治理效果时,却发现"面对具有突发性和高度复杂性的社会问题时,技术路径并不能作出完全有效的处理,甚至会带来更大概率的治理失灵"[18]。这是由于信息技术升级在促进社会生产力高速发展,为现代社会"提供了传统社会无法想象的物质便利的同时,也创造出众多新生危险源,导致技术风险的日益扩散"[19]。如未成年人犯罪受数字技术发展的影响,在犯罪模式、网络犯罪生态等方面迭代升级加速,各种利用新型数字技术、数字生态实施犯罪的案件层出不穷。因此,在数字社会中,未成年人犯罪预防在治理逻辑、治理技术和治理协同性等方面与传统的未成年人犯罪预防存在着显著差异,需要我们重新认识和思考。

三、传统犯罪预警模型的实践检视:预测犯罪风险的感知能力与解析能力弱

传统的未成年人犯罪预防与当前数字社会犯罪治理需求间的结构性矛盾,关键在于传统的未成年人犯罪预防的底层治理逻辑与当前未成年人犯罪的数字化发展之间存在系统性的不兼容。

(一)传统数据建模逻辑与行为复杂性间的错配

结构化数据①作为"最古老的数据存储形式之一,应用十分广泛"[21]。传统的未成年人犯罪预防主要以静态结构化数据为分析基础,在犯罪模式快速迭代和数字信息大爆炸的时代背景下,逐渐暴露出精准度不足、响应性滞后以及文本分析技术适配性缺失等系统性问题,难以满足现代社会对犯罪预防动态化和精准化的需求。

传统犯罪预警模型主要基于结构化数据构建,而结构化数据本身具有的静态封闭性会影响并造成司法流程延宕数据时效、数据孤岛割裂风险画像以及静态数据与动态风险错位等表现,降低预警模型对风险的感知能力。同时,由于字段编码的结构化处理方式容易引发信息离散而破坏预警行为的连续性与情境性,并削弱风险预警的解析精度。而且,由于预警模型的行为表征维度易坍缩,还存在着预警行为特征表达僵化、语义场域窄化以及复杂适应性缺失等问题,以致预警风险画像失真影响预判。

1. 数据静态封闭性降低风险感知能力

由于结构化数据存在着价值鉴定困难、理解和组织困难、归档方式选择困难等问题,形成结构化数据归档困境[22]。并且,结构化数据由于技术、管理或组织壁垒易导致数据分散孤立而难以实现跨部门、跨系统互通共享[23]。而风险画像完整性的割裂还将影响犯罪预警的客观性与完整性。同时,还存在着静态数据预测与动态风险错位的危险,这是因为原有结构化数据的生成很大程度上依赖于人工或半人工化输入,其更新周期远滞后于数字时代下未成年人行为的高速动态发展。当静态结构化数据无法及时捕捉到风险变化时,风险预测输出将呈现虚假安全信号和误判升级概率上升的双重风险。因此,结构化数据的封闭性使系统难以适配未成年人犯罪预警对犯罪风险的及时性分析感知、精确性定位以及跨区域协同能力等方面的需求。

---

①"结构化数据"是指一种数据表示形式,按此种形式,由数据元素汇集而成的每个记录的结构都是一致的,并且可以使用关系模型予以有效描述[20]。

2. 数据处理离散削弱风险解析精度

由于数据处理的结构化过程将连续、多维、情境化的客观信息强制分割或简化为离散单元，当复杂社会现象被强制编码为预设字段时，基于字段规范的结构化数据处理存在着系统性信息衰减效应。主要包括三种情形：一是“能指”与“所指”的断裂。由于数据规范化过程中将剥离其原始语境的情境化要素，会导致符号的指涉功能丧失或弱化；二是连续性割裂。所谓连续性割裂是指危险行为表现的流变过程被解构为离散字段时，会破坏时间维度的因果链；三是解释性维度坍缩。这是因为在字段规范结构化处理过程中，有关情感强度、道德模糊性等质性变量被简化为二元逻辑值，会大大削弱风险预警所需的灰度解析能力。而在未成年人罪错行为的情境触发机制中，这类信息耗散对于风险预警的精准度影响较大[24]。其中，对未成年人罪错行为具有较大影响力的网络煽动行为，不仅具有较强的情绪性，而且这类行为的表达形式还具有较大的模糊性。因此，结构化数据字段的极简表达范式与未成年人犯罪预测的复杂性需求形成根本性的悖反，会极大地限制传统预警模型对犯罪风险的感知与解析能力。

3. 数据维度塌缩致风险画像失真

结构化数据具有存储高效、查询便捷、支持自动化与集成以及便于拓展等优势[25]。但是，过于依赖预设字段与人工抽象的数据处理范式等问题，易导致风险预测模型行为表征维度的坍缩，使传统犯罪预警机制存在分析的局限性。这类局限性表现为：一是模型存在特征表达僵化，这是由于强制数据格式化的分析过程要剥离情境要素，使得动态风险行为被压缩为离散变量，会丧失对未成年人罪错行为的连续性轨迹特征的描述；二是模型语义场域过于窄化，仅能捕获未成年人部分的显性行为指标，如消费记录、出勤数据等，难以全面捕捉并解析未成年人在社交媒体中的隐喻性表达和亚文化符号等关键性预警信号；三是模型复杂适应性不足。当面对未成年人犯罪特有的情境触发和动机流变性较大等表现时，传统预警模型难以建立多级传导的风险映射机制以应变复杂性表现。正是这些问题与不足导致传统的预警机制难以穿透未成年人越轨行为表象并捕捉其心理动因，亦难以及时解析新型犯罪形态，本质原因就在于数据建模逻辑与未成年人行为复杂性间存在着根本性的错配[26]。

（二）数据简单调用与海量信息的隔离

传统犯罪预警机制虽然以结构化数据为核心和基础，但也有部分系统开始运用非结构化数据①进行分析。然而，现有运用非结构化数据系统多数还停留于简单调用层面。由于此类简单调用面临语体混杂、方言障碍以及表意复杂等困难和障碍，以致非结构化文本分析存在着标注繁复困难、工作量巨大等问题。而通用自然语言处理技术也无法直接突破深层语义解析，仍需要犯罪预防专家与人工智能开发人员协同设计专业模型。同时，在传统分析解决方案中由于缺乏底层架构集成，分析功能隔离、反馈迟滞和认知浅层化等问题，以致无法形成协同性效应，并难以应对复杂性问题。

1. 语体混杂与标注繁复的障碍难以消解

当前数据分析通过自然语言处理技术以检测并识别文档敏感内容的过程，已逐渐演化形成智能分类分级的技术趋势[27]。由于汉字作为表意文字具有“信息压缩”等特点，本身不仅有复杂内涵还存在着多音多义、隐喻、暗喻等各类复杂情形。因此，仅基于自然语言处理技术分析将面临着汉字语体混杂、表意规范化含糊以及标注繁复等困难与障碍。汉语不仅在“文档语义特征表达层面，各领域文档形式多样、内容丰富、中文语体混杂的情况普遍存在”[28]，而且在司法文书中还存在官方语体和当事人口语、地方方言等混杂表达的复杂情形，这些情形又给文本的识别与分析工作造成进一步的困难。

①“非结构化数据”，是指不具有预定义模型或未以预定义方式组织的数据[20]。

2. 表层技术应用难以解析深层语义

由于中文具有高度语境依赖性和信息压缩性等特点,在专业领域知识引导不足的自动化分析中较容易产生解析偏差和不清。因此,简单调用通用自然语言技术处理非结构化文本的效果十分有限,于是开发基于犯罪学语义规范,并经过未成年人行为特征集深度训练的专业化语言解析模型势在必行,但需要法律专业人士与 AI 技术专家的深度协作。遗憾的是,目前研究成果不多,且司法实践积累不足。在深层语义解析问题方面,虽然已有学者基于通用预训练大模型内置的自然语言处理技术分析非结构化文本,如调用 Deepseek 解释分析研报文本[29],还有学者运用通用预训练大模型分析海量的教学质量评价文本[30]。但这些分析都很少涉及对专门领域的语义解析进行模型微调等内容。司法实践方面,传统犯罪预警模型仍更多地依赖于人工方式对需要深度解析的笔录和卷宗进行结构化预处理。

3. 点状化解决方案缺乏协同赋能效应

司法实务部门一直积极致力于调整并不断地完善犯罪预防机制,但近年未成年人犯罪预警系统的技术优化仍呈现为“点状修正”的特点。虽然针对数据孤岛、算法黑箱[31]、系统响应滞后以及模型特征提取等问题逐一开发了解决方案,但仍存在着在人工智能支持下对底层架构深度集成不足,较难以形成系统的协同效应[32]等问题。由于传统犯罪预警系统的静态架构是通过解耦数据采集、清洗与建模等不同环节,易割裂对犯罪成因有机整体性的全面把握,以致解决方案存在着三方面的不足:一是功能隔离,表现为数据的表层联通难以融合业务规则与决策逻辑,使预警系统仍依赖于对离散特征的输入;二是反馈迟滞,这是由于批处理更新机制会阻却或中断预警系统的逻辑闭环,使即时危机响应存在着机制性延迟;三是认知浅层,由于传统预警系统中规则引擎与统计模型仅能捕获显性关联,却无法解析未成年人行为异化的非线性机制。为了弥补上述不足,虽然实务部门对技术应用缺陷与不足采取了“补丁化”对策,但仍难以克服传统静态架构固有的整体性僵化与适应性不足[33],更难以实现对未成年人犯罪复杂成因链条“端到端”的动态感知与解析评估。

(三)同质化风险响应与动态防控效能实现间的背离

以线性因果逻辑为核心的未成年人犯罪风险反应模式,难以分析家庭、学校与社区等多场域风险要素的动态耦合,在应对复杂风险传导时存在反应固化且迟滞等问题。同时,采用事后追溯型的干预模式依赖于以既成犯罪事实为触发条件,而错失了对未成年人罪错行为早期干预的关键窗口期。而且模式化的处置范式也忽视了未成年人行为风险流变与个体化矫治的多重性特殊需求,难以达成积极、精准地干预未成年人犯罪的目的。

1. 线性因果范式难以解析复杂系统性行为

传统预警机制将未成年人犯罪诱因简化为有限变量的线性组合,通过静态相关性分析并构建单向因果链条,而这种认知框架却难以容纳家庭、学校、社群以及网络平台等多维场域中不同风险要素的动态耦合过程。由于各子系统间的非线性交互既产生跨层级风险传导,又形成正负反馈循环,导致预警模型中系统行为具有涌现性与路径依赖性。当犯罪诱因在多因多果、跨区域协同的复杂网络中发展演化时,这种线性因果链会消解风险演化的连续谱系与临界相变特征,故而基于孤立风险因子阈值判定的预警机制必然失效[34]。究其原因,其深层矛盾在于传统预警模型以确定性方程来模拟生物社会系统的复杂混沌运动,用可加性假设替代要素间的非线性相干作用,以静态快照遮蔽风险流的时变拓扑结构,却忽略了未成年人犯罪风险评估的多维性与内在复杂性特点[35]。因此,思维模型的错位将会严重削弱预警机制对复杂行为的预测能力[36],导致预警响应表现出时间维度上迟滞、空间维度中失焦以及逻辑维度里碎片化等问题。

2. 事后追溯错过干预犯罪的关键窗口期

传统的未成年人犯罪干预模式具有事后追溯性，以犯罪立案事实或显性违规行为为触发干预条件，对潜伏期高风险信号的主动侦测存在着结构性缺位。学者认为未成年人犯罪的心理动因具有高度流变性[37]与情境激发性[38]的特点，要求犯罪预警响应和干预必须具有高度的时效性。然而，由于传统犯罪预警机制从数据采集、人工研判，再到学校、家庭和社会群体等介入耗时冗长，可能会错失对未成年人犯罪有效干预的关键阶段与窗口期。

3. 同质化处置模式难以满足个体化矫治的特殊需求

传统的未成年人犯罪预防因其固化的处置范式与分类分级矫治处置的需求之间，存在着结构性错位而形成制度性耗散。在未成年人犯罪的生成过程中，犯罪风险生成与未成年人个体的心理韧性存在着渐进式谱系分布。并且，未成年人越轨行为演化、生成犯罪行为亦是动态的连续发展过程。因此，未成年犯罪人预防与矫治应根据其个体心理特征和社会嵌入情境，形成交互复杂的持续性循证矫治过程。而传统的犯罪预防却更多地表现为依赖于静态行为触发同质化的干预过程，与未成年人罪错行为的动态演进过程悖反[31]，甚至在干预错位的情况下还可能存在着双重的异化：一方面，由于高风险个体因离散化分类阈值掩盖了风险累积的进程，而导致干预时效的折损；另一方面，可能因为对低风险个体基于机械化的归类而触发过度的干预反应。因此，亟须构建融合连续性风险表征的风险分析智能体，以实现未成年人犯罪预防从形式分类向实质分级的重构。

四、融合智能体的运行逻辑：多智能体联动预测未成年人犯罪风险

基于智能算法展开犯罪预防正逐渐成为社会治理数字化转型的重要选择，其重要原因之一就在于前瞻性的干预逻辑与预防性刑法观的价值取向具有深层的契合[39]。如果说犯罪预防就是“治未病”，那么“治未病”的前提就是精准识别“将病之兆”。于是，建构融合智能体以实现对海量涉未成年人多源异构数据的深度分析，成为识别、发现未成年人的罪错行为向犯罪生成转变的关键，也是重构未成年人犯罪风险预警机制的核心智能驱动。

（一）多智能体联动预测未成年人犯罪新动态

融合智能体由分析智能体、风险预测智能体和自适应新风险点拓扑智能体三个子智能体融合构建，以这三个子智能体联动为运行核心，分析输入数据中的风险信号，将影响未成年人犯罪生成的多维度风险因子转化为可量化的预测依据，并输出风险指数。融合智能体的整体运行，由分析智能体率先启动数据处理流程，以多元犯罪学语言模型分析输入的未成年人密接主体登记、提供的多源异构数据①，形成覆盖多维度风险的未成年人行为风险特征集，同时通过专家校正环节，对分析智能体的输出结果进行人工校验，循环增强其分析准确度。之后，以未成年人行为风险特征集构建数据集，作为风险预测智能体和自适应新风险点拓扑智能体的基础数据支撑。风险预测智能体基于此数据集，依托双层神经网络算法结合风险关联逻辑，在基座模型的辅助下完成风险量化运算，为输出风险指数提供支撑。同时，基于数据集对基座模型进行深度微调而形成新型风险点识别微调模型，成为自适应新风险点拓扑智能体的决策中枢。而自适应新风险点拓扑智能体通过不断更新和调整知识库中的新型风险特征，实现对风险预测智能体和分析智能体的持续补充与完善，保障输出风险指数能同步反应犯罪新变化（见图1）。

①未成年人密接主体登记、提供的多源异构数据，主要包括来源于民政部门、教育机构、司法机关登记、提供的未成年人相关数据，网络平台等其他主体提供的未成年人的网络行为轨迹、校园心理档案、家庭沟通记录等数据。这些数据来源是从理论建构上提出的建议，在司法实践运用中，对于任何获取和使用未成年人的个人信息和数据的行为，必须进行严格控制和审核，坚持“最有利于未成年人的原则”维护未成年人的利益，必须严格遵守我国《未成年人保护法》《个人信息保护法》《网络安全法》《数据安全法》和《民法典》的有关规定，对于获取和使用未成年人数据、信息的范围、种类和程序必须经过严格考查和审核。

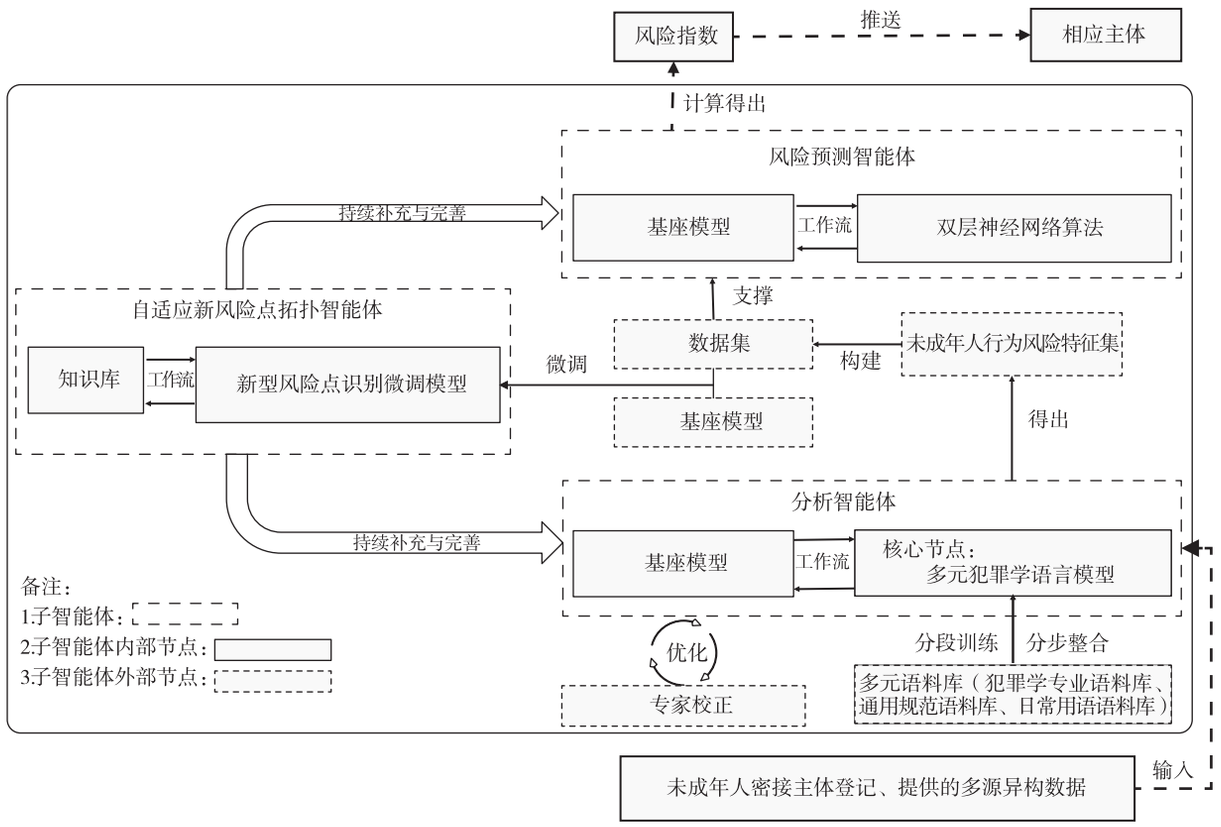

图1　融合智能体运行示意图

犯罪学原理认为“犯罪、被害、犯罪原因及犯罪治理(也称为犯罪防控),这四个范畴共同组建了犯罪学的理论体系”[40],而“挖掘犯罪的多元影响因素日益成为当代犯罪学研究的主流趋势”[41]。因此,融合智能体基于多维度风险因子形成的风险分析体系包括四重维度:一是心理维度,主要是量化未成年人情绪应激反应与冲动控制缺陷之间的关联度;二是社会环境维度,主要是分析未成年人的家庭监护失能指数与社区控制力衰减率之间的关系;三是情境互动维度,主要是捕捉并分析受监管缺位时长对未成年人脆弱性的影响;四是网络生态维度,主要是追踪并分析在数字空间中,未成年人与越轨亚文化的接触频度与沉浸深度对未成年人越轨行为的影响程度,而且,此维度还可以指数级放大前面三个维度的行为风险程度。故而,此维度的分析,对于在数字社会中发现并解析网络生态等新生危险源是如何影响并促进未成年人行为向犯罪发展演变的显得尤为重要。因此,根据上述多维度风险因子的非线性交互表现,在共变算法基础上构建未成年人罪错行为“生理心理缺陷—控制力缺口—机会窗口—文化催化”的解释模型,不仅有助于认识数字社会中未成年人犯罪的新变化,还可以为更好地应对犯罪做好准备。

(二)分析智能体通过模块化训练化解数据规模悖论

分析智能体以多元犯罪学语言模型为核心,完成对输入的多源异构数据的语义解析。多元犯罪学语言模型以犯罪学专业语料库为基础[42],结合通用规范语料库和日常用语语料库,深度挖掘涉未成年人非结构化文本数据中的风险信号[43],以精准适配对文本数据中风险语义挖掘的需求。当下主流语言模型主要包括专家语法规则模型、统计语言模型和神经网络语言模型三种类型。多元犯罪学语言模型主要参考统计语言模型技术原理,通过监督学习①结合分段训练加分步整合方法[28]进行多轮训练,以解除对预设语

①监督学习(supervised learning)是指向模型输入包含观测值及其对应已知输出值的训练数据,通过学习构建输入与输出之间的映射规律,最终实现对仅具备输入特征、无对应输出信息的全新未知数据进行输出预测[44]。

法规则的绝对依赖并降低全量语料覆盖的算力成本，此路径的实质在于以统计概率替代了规则穷举，运用模块化训练得以化解数据规模悖论。

（三）动态风险识别功能保障预测的科学性与准确性

风险预测智能体由基座模型和双层神经网络算法节点组成。基座模型为双层神经网络算法提供运行支撑，完成对基于未成年人行为风险特征集构建的数据集的风险量化运算，并输出风险指数。其中，双层神经网络算法是平衡未成年人犯罪风险特征处理精度与预警系统实操性之间冲突的关键。基于未成年人行为风险特征集构建的数据集，具有特征项多、维度复杂，并且数据间存在复杂关联的特点，如果采用单层神经网络算法，则易出现多维度风险特征堆砌融合、层级关联拆解能力不足的问题，无法保障数据集的处理精度；若采用多层神经网络算法，又会存在模型结构冗余、运算复杂度高的不足，且后续一旦增加新的风险特征，需要重新适配数据集、重新训练模型，所需时间较长，将严重削弱风险预测智能体的快速迭代能力和落地实操性。而采用双层神经网络算法不仅可适配高危风险特征集的多维度关联特性，还可以贴合未成年人犯罪预防的实践场景变化快的需求，能避免技术与应用的脱节。此外，为了保证风险预测智能体输出风险指数的可靠性，还通过将输出风险指数与司法实践中犯罪实际情况相验证和对比，然后将验证结果反馈至各子智能体并据此进行相应调整。于是，通过持续优化、不断调整风险因素的权重，并辅助以人工调校算法逻辑提高风险预测的准确率与精度，最终保障输出风险指数的科学性与准确性。

（四）自适应学习功能应对犯罪风险的动态演变

自适应新风险点拓扑智能体的自适应学习功能是融合智能体应对风险动态变化、完善量化预测的关键性补充。该子智能体由新型风险点识别微调模型和知识库组成，前者是以数据集为依托，运用大模型微调技术对基座模型进行训练而生成，其核心作用在于自主识别未被覆盖的未成年人行为新风险点，使融合智能体无需重构核心架构即可快速适配未成年人行为风险的演变场景；后者的知识库是经人工精校的已知犯罪风险特征标准库，其内容包括已明确的未成年人犯罪风险特征和犯罪风险特征识别标准。一方面，知识库为新型风险点识别微调模型识别犯罪新风险提供参照基准；另一方面，知识库还承接经人工二次精校后并确认的犯罪新风险特征，实现知识库内容的不断扩充和沉淀。而且，知识库中新增的未成年人行为风险特征，还会定期同步更新至风险预测智能体和分析智能体。在补充现有风险特征集未能覆盖的新场景的同时，还不断补充并完善融合智能体的学习逻辑和风险识别标准，使其能持续性地应对未成年人犯罪风险的升级与变化。

五、数字社会未成年人犯罪预防机制的重构：“三重”预防与风险智能预测的结合

现代犯罪预防“根据犯罪预防对象人群的不同，将犯罪预防分为一般预防、临界预防和再犯预防三种类型”[45]。融合智能体以“三重”预防为实践逻辑的“目标导向”，明确犯罪多因性理论的量化应用场景，将多维度风险因子的量化结果与“三重”预防需求予以精准匹配。将一般预防转化为强实操性的资源分配方案，使临界预防从定性判断升级为定量指导，并为再犯预防中矫治方案的设计提供精准的数值参考，促进犯罪预防理论与犯罪风险智能预测的交叉融合，不仅可填补理论阐释与实践应用的鸿沟，还可实现不同学科领域间的深度融合。

数字社会中的未成年人犯罪风险预警机制依托融合智能体，坚持“未成年人优先”与“人文优先技治主义”相结合。在融合智能体对未成年人行为风险进行量化评估和预测的基础上，将未成年人的罪错行为与可能实施犯罪的不同风险程度采用“六类四级”法划分，对应触发预警响应并联动差异化精准处遇机制，形成分类分级、多元共治的未成年人犯罪风险预警机制，实现对未成年人犯罪一般预防、临界预防与再

犯预防相结合的构想,以积极应对数字社会中未成年人犯罪的新挑战。

(一)“未成年人优先”下“人文优先技治主义”的选择

“未成年人优先”源于国际社会的“儿童最大利益”原则[46],其包含三层含义:一是作为一种首要选择与评判的实质性权利;二是一种优先考虑的基本指导与解释性法律原则;三是作为一种行事规则,要求凡作出任何涉及或有可能涉及未成年人的决定时,应对所涉未成年人的影响进行评估与判断。同时,由于未成年人自身的特殊性以及我国“慈幼”文化的延续[47],我国《未成年人保护法》亦明确要求“坚持最有利于未成年人的原则”。因此,在未成年人犯罪风险预警机制重构中,必须坚持以“未成年人优先”为核心价值导向,才能最大化地保障未成年人的权益。

学界有人提出“代码即法律”是数字法律规范的重要特征之一[48],认为现实空间中的法律在数字空间应当服从于数字技术的内在客观规律。亦有学者认为“在大数据技术和人工智能广泛应用的时代,谁掌握分析数据的算法技术,谁就能决定人类社会的秩序”[49]。然而,技术代码不仅具有规范建构功能,同时技术系统也具有自主性扩张的倾向和危险,并且“脱离管控的代码体系将会在网络世界中构筑起新的规则,对公民权益侵害的风险也将随之增加”[48]。由于技术代码的自主性与社会风险具有正相关性,因此,对算法权力的制衡也是现代社会必须面对与考虑的紧迫性问题。毕竟,“科学技术的合理性运用还需要合目的性的中介”[50],因此,算法权力除了要受到必然性规律的约束外,还应受合目的性的约束。于是,有学者提出“在数字法学理论体系中,人类为世界的当然中心”[51],要求“人文主义优先于技治主义”,将人的主体性置于数字技术本身系统内部的逻辑自洽之上,其核心要义在于确立人类价值对技术工具理性的优先性。

未成年人犯罪预警机制坚持以“未成年人优先”为前提,选择“人文优先技治主义”以制衡算法权力。当前,由于AI技术在未成年人犯罪预防领域的应用存在着制度性不足,大模型设计易陷入“标签化”而违背对未成年人特殊、优先保护的原则与要求。同时,AI算法的黑箱特性,也可能让犯罪风险评估脱离人类对未成年人权益保护的实质性判断与初衷,与“最有利于未成年人的原则”产生冲突。故需确立人文价值对技术发展的目的性指引,使算法权力必须服务于人的主体性发展。因此,在坚持“未成年人优先”下选择“人文优先技治主义”,是以最有利于未成年人为犯罪风险预警机制的核心技术标尺,通过“未成年人权益大于技术效能”的价值排序,才能实现技术辅助与人文主导的辩证统一。

(二)“六类四级”法划分行为与风险评级

未成年人犯罪风险预警机制将未成年人罪错行为分为:受不良倾向诱导与影响的行为、有越轨倾向表现的行为、轻微越轨行为、严重越轨行为、违法行为与犯罪行为等六类。根据六类行为的犯罪危险差异又划分为四级风险:一级风险,包括受不良倾向诱导与影响的行为和有越轨倾向表现的行为两类,属潜在风险级,是越轨行为倾向处在萌芽发展但尚未外化成越轨行为的阶段;二级风险,包括轻微越轨行为与严重越轨行为两类,属行为偏差级,即已经实施越轨行为,但尚未达到违法程度;三级风险,包括违法行为,属违法级,即已经实施违法行为但尚未构成犯罪的阶段;四级风险,包括犯罪行为,属犯罪级,即已经实施或完成犯罪行为,从具有犯罪风险已经转化为犯罪实行阶段。对罪错行为以六类划分,覆盖了未成年人从越轨行为的萌芽阶段到犯罪完成的整个犯罪行为生成过程。罪错行为犯罪风险以四级区分,分别对应着未成年人罪错行为从具有潜在犯罪风险到犯罪现实危险的梯度递进,表明了罪错行为与犯罪风险之间的关联性,为预警机制触发响应提供依据(见图2)。

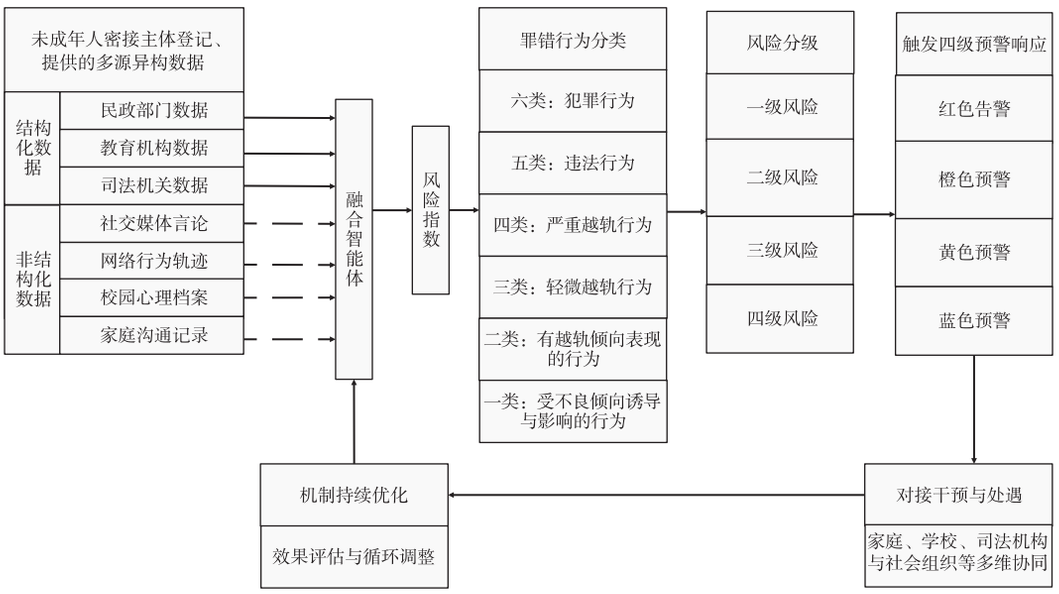

图2　未成年人犯罪风险预警机制结构示意图

（三）“等级对应＋场景适配”的风险预警触发规则

风险等级评估基于风险信号采集和筛选后得到的涉目标未成年人数据，采用“六类四级”予以划分，通过融合智能体量化输出的风险指数评估为不同风险等级。预警触发响应根据未成年人罪错行为的不同危险程度，以及在犯罪生成发展演变过程中所处的环节与阶段，将不同类别的罪错行为评估为相应的危险等级，并对应触发相应的预警信号。蓝色预警与一级风险相对应，主要针对受不良倾向诱导与影响的行为和有越轨倾向表现的行为；黄色预警与二级风险相对应，包括轻微越轨行为与严重越轨行为两类；橙色预警对应于三级风险，针对违法行为；红色告警则对应四级风险，针对犯罪行为。预警触发模块则依据等级对应加场景适配的规则运行，四级红色告警触发最高级的司法干预，即通过预警机制将预警信息直接推送至司法机关和专门矫治教育机构；三级橙色风险触发专业矫治预警，将信息推送至专门的矫治教育机构、家庭和学校；二级黄色风险触发风险干预预警，将信息推送至学校、家庭与相关社会组织；一级蓝色风险触发预防预警，将信息推送至未成年人所属家庭、学校和相关社会组织（见图3）。

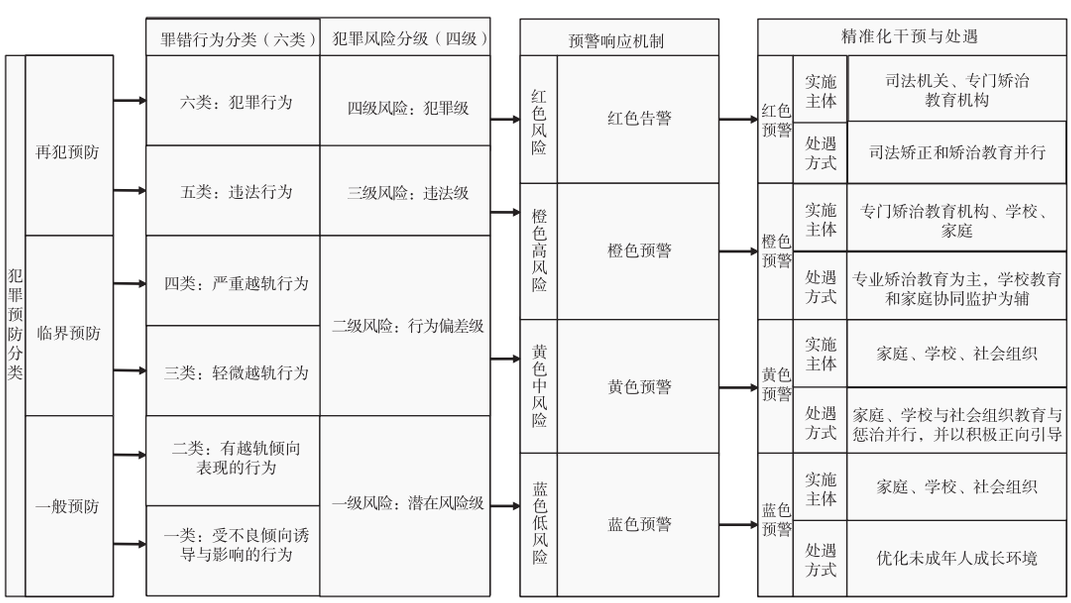

图3　未成年人罪错行为与犯罪风险预警、处遇分级分类示意图

(四)精准化处遇下犯罪预防从形式分类向实质分级转型

“三重”预防理论不仅为融合智能体提供实践的“目标导向”,更是精准化处遇不同类型罪错未成年人的策略与指导。基础预防主要针对蓝色低风险未成年人群体,以柔性引导方法为主,通过家庭和学校开展法治启蒙、规则教育以及心理疏导等形式提供正向指导措施,通过非强制性措施从源头抑制风险;临界干预主要针对黄色中风险未成年人群体,采用柔性引导加刚性约束模式,通过以学校为主导进行心理干预与行为矫正、社区组织开展专题实践教育活动,对未成年人的严重不良行为实施校园纪律处分、社区告诫等刚性约束,并要求家庭落实严格的监护措施并协同管教,形成合力以阻断未成年人行为风险进一步升级;再犯预防主要针对橙色和红色高风险未成年人群体,处遇措施则以刚性约束为主、专业矫治为辅,对违法未成年人群体由特殊教育机构开展强制性矫治教育,并结合运用心理重塑、职业技能培训和法律教育等措施,同时,要求家庭承担协同监护责任;对未成年犯罪行为人群体则由司法机关依法追究法律责任,后续阶段转入社区矫正或进行专门性矫治教育,通过司法惩戒、专业矫治以及社会融入帮扶等系列组合措施,以阻却或降低未成年人的再犯风险。

六、结语

在数字时代背景下,社会治理面临着传统治理向精准化、数智化转型的重要使命。随着近年人工智能的快速发展,利用 AI 辅助未成年人犯罪预防也更具有理论探讨价值和现实可能性。因此,将犯罪预防理论与风险智能量化预测相融合,依托融合智能体对犯罪风险进行智能评估和科学研判,按照等级对应加场景适配的风险预警触发规则,对未成年人的犯罪行为从风险萌芽、临界演化到现实爆发的不同发展阶段,遵循分别予以源头防范、阻断升级和矫治修复的精准适配的应对策略,构建“行为分类、风险分级、预警响应、精准处遇”的未成年人犯罪风险预警机制,不仅是对未成年人犯罪预防新路径的理论探讨,也为数字社会中人工智能深度融入未成年人犯罪预防展现了现实可能性。

参考文献:

[1]最高人民检察院. 最高检发布《未成年人检察工作白皮书(2020)》[EB/OL]. (2021 - 06 - 01)[2025 - 09 - 12]. https://www.spp.gov.cn/xwfbh/wsfbt/202106/t20210601_519930.shtml#1.

[2]最高人民检察院. 最高检发布《未成年人检察工作白皮书(2021)》[EB/OL]. (2022 - 06 - 01)[2025 - 09 - 12]. https://www.spp.gov.cn/xwfbh/wsfbt/202206/t20220601_558766.shtml#1.

[3]最高人民检察院. 最高检发布《未成年人检察工作白皮书(2022)》[EB/OL]. (2023 - 06 - 01)[2025 - 09 - 12]. https://www.spp.gov.cn/xwfbh/wsfbt/202306/t20230601_615967.shtml#1.

[4]最高人民检察院. 最高检发布《未成年人检察工作白皮书(2023)》[EB/OL]. (2024 - 05 - 31)[2025 - 09 - 12]. https://www.spp.gov.cn/xwfbh/wsfbh/202405/t20240531_655854.shtml.

[5]最高人民检察院. 最高检发布《未成年人检察工作白皮书(2024)》[EB/OL]. (2025 - 06 - 16)[2025 - 09 - 12]. https://www.spp.gov.cn/xwfbh/wsfbt/202506/t20250616_698278.shtml#1.

[6]董玉庭. 三种语境下的犯罪概念[J]. 学术交流,2010(7):60 - 66.

[7]张荣军,张溪. 数字资本主义时代的交往异化及其重构[J]. 贵州大学学报(社会科学版),2025(1):34 - 43.

[8]魏红,张奎. 数字经济时代“算力”的内涵及其刑法保护思考[J]. 贵州大学学报(社会科学版),2025(1):95 - 105.

[9]最高人民检察院. 江苏:持续加强未成年人网络司法保护[EB/OL]. (2025 - 06 - 03)[2025 - 09 - 02]. https://www.spp.gov.cn/zdgz/202506/t20250603_697227.

[10]央广网. B 站披露一起“人肉开盒”案细节 公安机关:主要活动人员已被处罚[EB/OL]. (2023 - 11 - 23)[2025 - 09 - 12]. https://china.cnr.cn/gdgg/20231123/t20231123_526495638.shtml.

[11]梁阿敏,韩琳. 未成年人涉网犯罪态势、生成机理及其治理研究[EB/OL]. (2025-02-10)[2025-12-09]. https://szb.fzshb.cn/fzshb/20250210/html/content_20250210007003.html.
[12]杨宇琦. 惩罚或是挽救:未成年人犯罪中的社会认知失调与协调[J]. 当代青年研究,2022(2):73-79.
[13]曹文振,唐昆. 青少年青春期性行为影响因素研究[J]. 青年研究,2020(6):43-56.
[14]迈克尔·戈特弗里德森,特拉维斯·赫希. 犯罪一般理论[M]. 吴宗宪,苏明月,译. 北京:中国人民公安大学出版社,2009:102.
[15]王燕. 大数据与未成年人犯罪心理预防模式变革研究[J]. 江苏警官学院学报,2021(6):79-83.
[16]陈京春,钱路璐. 青少年新型犯罪的趋势与数字治理体系的建构:以Z省实证调研数据为研究对象[J]. 青少年犯罪问题,2025(5):131-146.
[17]刘少军,胡志宇,刘牧原. 系统性治理视角下未成年人涉网犯罪预防体系构建[J]. 预防青少年犯罪研究,2025(2):35-43.
[18]向玉琼. 论技术治理的民主化:基于对风险治理技术化的反思[J]. 学术界,2020(12):88-96.
[19]劳东燕. 公共政策与风险社会的刑法[J]. 中国社会科学,2007(3):126-139,206.
[20]国家数据局. 数据领域常用名词解释(第一批)[R/OL]. (2024-12-30)[2025-09-03]. https://www.nda.gov.cn/sjj/zwgk/zcfb/1230/20241230160715745237413_pc.html.
[21]王伟成,万书振,张晓磷,等. 基于定向特征交互选择的医疗结构化数据处理方法[J/OL]. 计算机工程,[2024-12-31],1-12. https://doi.org/10.19678/j.issn.1000-3428.0070391.
[22]王强,符京生. 根据业务对象组织数据体的结构化数据归档路径探析:基于勘探开发结构化数据实践的思考[J]. 档案学通讯,2024(4):22-29.
[23]程骏超,张驰,何元安. 区块链技术在跨部门海洋数据共享中的应用[J]. 科技导报,2020(21):60-68.
[24]林爽,程塬渌,刘文,等. 初中生认知控制对替代性攻击行为的影响:激惹和触发的情境边界条件[J]. 心理学报,2025(1):36-57.
[25]马花月. 工程结构化数据预处理技术[J]. 中国科技信息,2025(9):44-47.
[26]佟亚洲. 数字智能驱动基层治理的三重悖论及其调适[J]. 中州学刊,2025(8):73-82.
[27]江文彬,刘兆霖,谢仕康,等. 基于层次化主题分析的铁路敏感数据智能识别与分类分级方法[J]. 铁路计算机应用,2024(10):7-12.
[28]喻波,王志海,孙亚东,等. 非结构化文档敏感数据识别与异常行为分析[J]. 智能系统学报,2021(5):932-939.
[29]钟宁桦,郝雨桐,刘一莹. 基于非结构化文本的房地产债券违约预警研究[J]. 中山大学学报(社会科学版),2025(4):359-374.
[30]孟庆祥,李彦胜,龚龑,等. 基于人工智能多主体协同的教学质量评估模型研究[J]. 测绘通报,2025(S2):297-302.
[31]孙宇. 论数字时代罪错未成年人风险评估的协同治理[J]. 中国人民公安大学学报(社会科学版),2025(1):59-69.
[32]罗海敏. 性侵害未成年人违法犯罪人员信息系统的建构:以区块链技术为基础[J]. 华东政法大学学报,2022(4):83-92.
[33]李丽,靳煜昕,鹿一鸣. 大数据下山西省预防青少年违法犯罪现代化体系构建研究[J]. 山西警察学院学报,2024(2):68-75.
[34]付凤,李世英. 我国涉罪未成年人人身危险性评估要素相关性分析[J]. 中国刑警学院学报,2020(6):66-75.
[35]张鸿巍. 风险/需求评估在预防未成年人再犯罪中的适用[J]. 暨南学报(哲学社会科学版),2021(3):50-61.
[36]陈烨,高金虎,马晓娟. 预警情报的范式转向问题研究:基于美国预警情报工作的考察[J]. 情报杂志,2023(11):1-9,23.
[37]齐娜娜,罗晓芳,王强. 情绪调节策略对青少年激情犯罪的预防作用:基于Gross情绪调节过程理论[J]. 现代交际,2014(9):5-6.
[38]王旌宇,冯健新,张玉清. 心理学视角的青少年暴力犯犯罪成因探究[J]. 现代交际,2020(3):65-66.
[39]高铭暄,孙道萃. 预防性刑法观及其教义学思考[J]. 中国法学,2018(1):166-189.
[40]张旭,单勇. 论刑事政策学与犯罪学的学科价值及其连接点[J]. 法商研究,2007(5):77-84.
[41]许博洋. 青少年网络暴力的实证犯罪学解释:以“犯因性”与“脆弱性”为分析视角[J]. 北京联合大学学报(人文社会科学版),2025(1):84-96.
[42]任淑芳,于龙. 我国未成年人犯罪研究30年回望:基于Cite Space的知识图谱分析[J]. 广州市公安管理干部学院学报,

2021(4):54 - 60.
[43]贾云飞. 智慧司法视阈下未成年人网络犯罪的危险性评估[J]. 预防青少年犯罪研究,2024(1):46 - 54.
[44]Badillo S,Banfai B,Birzele F,et al. An introduction to machine learning[J]. Clinical Pharmacology & Therapeutics,2020(4):871 - 885.
[45]董邦俊,王小鹏. 未成年人临界行为及预防对策研究[J]. 政法论丛,2016(4):85 - 94.
[46]何海澜. 善待儿童:儿童最大利益原则及其在教育、家庭、刑事制度中的运用[M]. 北京:中国法制出版社,2016:45.
[47]魏红. 一切以未成年人优先:性侵害未成年人犯罪刑事政策研究[M]. 北京:社会科学文献出版社,2021:128.
[48]郑智航. 当代中国数字法学的自主性构建[J]. 法律科学(西北政法大学学报),2024(2):81 - 93.
[49]闫宇晨. 社交平台私权力的滥用及其治理[J]. 公共管理与政策评论,2023(4):142 - 153.
[50]罗骞,滕藤. 技术政治、承认政治与生命政治:现代主体性解放的三条进路及相应的政治概念[J]. 武汉大学学报(哲学社会科学版),2020(1):27 - 34.
[51]刘祖兵. 数字法学研究体系之理性证成:兼论数字法学部门法化的批判[J]. 河南财经政法大学学报,2025(4):42 - 55.

(责任编辑:蒲应秋)

Research on the Deep Integration of Artificial Intelligence and Juvenile Delinquency Prevention in the Digital Society from the Perspective of Reconstructing the Crime Risk Early - Warning Mechanism through an Integrated Intelligent Agent

WEI Hong,ZHANG Tong

(Law, School, Guizhou University, Guiyang, Guizhou, China, 550025)

Abstract:In today's digital society, juvenile delinquency is becoming increasingly networked and concealed. As traditional systems for preventing juvenile delinquency remain limited in their capacity to perceive and interpret crime risks, the integrated intelligent agent has emerged as a possible response. Composed of three sub - agents—an analytical agent, a risk prediction agent, and an adaptive emerging - risk - point topology agent—the integrated intelligent agent enables coordinated interaction among these components and serves as the core driver of the early - warning mechanism for juvenile crime risks. By focusing on the dynamic evolution of juvenile delinquent behavior from "risk emergence" to "critical transition" and then to "actual commission", the integrated intelligent agent classifies behaviors and assesses risk levels according to a "six - category, four - level" framework. With the early - warning goals of "identification in advance, targeted intervention, and escalation prevention", it applies precise intervention measures that combine flexible guidance with hard constraints. In this way, it helps construct a three - tier prevention model consisting of "general prevention - critical intervention - recidivism prevention". This mechanism forms a government - led, multi - actor co - governance system for the early warning of juvenile delinquency risks. It also offers a scientific approach and a practical pathway for transforming traditional social governance toward greater precision and digital intelligence.

Keywords:prevention of juvenile delinquency; integrated intelligent agent; early warning of crime risk; digitally intelligent social governance