DOI: 10.54254/3050-2160/2025.25696

技术向善：人工智能辅助预防“隔空猥亵”未成年人行为研究

鲁霄飞

（苏州大学王健法学院，江苏省苏州市，215006；ludaniel1016@163.com）

摘　要：利用网络平台的“隔空猥亵”行为对未成年人的性利益和身心健康造成了严重威胁，违背了最有利于未成年人原则，人工智能辅助预防符合法律3.0时代的要求。同时，“隔空猥亵”未成年人行为可识别性强，人工智能技术能将“不应”变为“不能”并积极调节行为人道德。此外，三层监管责任、“无知之幕”、协商民主等理论的实际运用为可以为人工智能辅助预防“隔空猥亵”未成年人行为的可持续发展提供支撑，尽量避免人工智能技术的负面影响，促进人工智能技术向善。

关键词：“隔空猥亵”道德物化；人工智能；技术向善

引言

随着互联网与社会生活的深度融合，网络扩展了传统线下犯罪活动的发生空间与实施手段[1]，传统的强制猥亵罪和猥亵儿童罪的猥亵行为在线上扩展为“隔空猥亵”未成年人行为。2018年11月18日，最高人民检察院发布第十一批指导性案例，其中的“骆某猥亵儿童案”（检例第43号）首次从司法认定的角度指出，行为人以诱骗、强迫或者其他方法要求儿童拍摄、传送暴露身体的不雅照片、视频，行为人通过画面看到被害儿童裸体、敏感部位的，是对儿童人格尊严和心理健康的严重侵害，与实际接触儿童身体的猥亵行为具有相同的社会危害性，明确了对“隔空猥亵”儿童行为的处罚。2022年10月28日，《最高人民检察院关于人民检察院开展未成年人检查工作情况的报告》指出，2018年至2022年9月，起诉利用网络“隔空猥亵”未成年人犯罪1130人，利用网络对未成年人实施“隔空猥亵”和线上联系、线下侵害的犯罪占性侵未成年人犯罪的15.8%[2]。可见，“隔空猥亵”未成年人行为已经对未成年人的性利益和身心健康发展产生了不可忽视的威胁。

道德物化理论认为，技术不是中立的工具，人不仅在使用技术，技术也在不断影响和塑造人的行为，人与技术是一种互动甚至是互构关系[3]。我们的日常生活已经与技术紧密地交织在一起，社交媒体网络平台、跨平台通讯工具、短视频类应用程序等网络平台的用户可以通过终端设备向其他用户发送语音短信、视频、图片、表情、文字和链接等内容，为用户的沟通交流提供了线上渠道，也塑造着用户使用网络平台的习惯。技术的发展不是让我们的生活变得更加便利的简单的中立工具，在完成沟通功能的同时，也框定着我们可以做什么以及我们如何体验世界，并且以此方式积极地参与到我们的生活当中。网络平台为“隔空猥亵”未成年人行为提供了可能性与可行性，网络平台的发展客观上促成了“隔空猥亵”未成年人行为的产生。“隔空猥亵”未成年人行为的实施不单是行为人自己的决定与行为，更是行为人和网络平台的相互交织而做出决定并按照决定进行行动。网络平台为行为人的“隔空猥亵”未成年人行为提供了可能性，对行为人的决定产生了影响，并且为行为人按照决定进行行动提供了手段。网络平台通过对行为人的实践和体验塑形的促进，而促进着行为人的决定和行为的塑形。为了预防“隔空猥亵”未成年人行为，保护未成年人的性利益和身心健康发展，笔者认为应当发挥技术对人的行为的正向影响和塑造，利用人工智能技术对已经产生了“隔空猥亵”未成年人行为或具有产生“隔空猥亵”未成年人行为风险的网络平台进行干预，调节网络平台在行为人和未成年人沟通中的作用，促进对行为人和未成年人使用网络平台行为的塑形。

1. 人工智能辅助预防“隔空猥亵”未成年人行为的必要性

“隔空猥亵”未成年人行为已成为不容忽视的社会问题，严重威胁未成年人的性自主权和身心健康。从法律层面看，“隔空猥亵”违反刑事法律规范，侵害未成年人的法益；从社会层面出发，最有利于未成年人

原则要求我们必须对此类行为采取有效措施；从时代发展角度而言，法律3.0时代的到来，需要技术措施来补充法律规制。因此，运用人工智能辅助预防“隔空猥亵”未成年人行为具有现实紧迫性。

1.1.“隔空猥亵”未成年人行为的不法实质

随着网络图片与音视频传输等网络通讯技术与硬件的发展，猥亵行为从传统的直接身体接触形式呈现出向非接触式“隔空猥亵”形式的转变，行为人依托网络图片与音视频传输等网络通讯技术，既可远程操控未成年人自行实施暴露隐私部位、模拟性行为动作等具有性意义的行为，亦可迫使未成年人观看行为人或他人做出的色情行为，此类技术赋能的远程沟通手段，实质上消解了物理空间的阻隔效应，从而使得“隔空猥亵”能够达到与当面猥亵相当的对性自主权的侵害程度[4]。面对性侵未成年人犯罪，坚持“零容忍”和从严惩治态度，是司法实务的政策共识[5]。2023年5月24日，最高人民法院和最高人民检察院联合发布《关于办理强奸、猥亵未成年人刑事案件适用法律若干问题的解释》，其中第9条规定明确了行为人胁迫、诱骗未成年人以视频聊天或者发送视频、照片等方式，暴露身体隐私部位或者实施淫秽行为的，应当以强制猥亵罪或者猥亵儿童罪定罪处罚[6]。由此可见，“隔空猥亵”未成年人行为是对我国刑事法律规范的直接违反。

“隔空猥亵”未成年人行为侵害的法益是未成年人的性自主权和性健全发展。

性自主权是一项人格权利，以对具有性意义的行为的自主决定与同意为其核心[7]，其内涵指向大体分为性支配权和性维护权[8]。性关系必须作为主体自主决定与同意的实践结果，否则就是对人格尊严的侵害。当个人对性关系存在有效的自主决定和同意时，可以认为该性关系是个人性自主权的实践结果。当欠缺主体对性关系的有效同意时，该行为就贬损了个人的主体地位，损害了人格尊严[9]。“隔空猥亵”未成年人行为在不存在有效同意的情况下，将未成年人降格为满足行为人性欲目的的工具，贬损了未成年人的人格尊严。

基于班杜拉的社会学习理论，性发展是一个社会学习的过程，个体通过观察、模仿、强化等多种方式与其他社会成员进行互动、协作和对话[10]，来学习与性相关的行为和态度。“隔空猥亵”未成年人行为发生后，未成年人受害者可能仍会在长时间内受到心理上的伤害，为其未来通过与其他社会成员互动来促进自身性健全发展产生了不可忽视的阻碍。在“隋某某利用网络猥亵儿童，强奸，敲诈勒索制作、贩卖、传播淫秽物品牟利案”（检例第200号）中，隋某某以传播14岁的被害人的裸照、裸体视频相威胁，多次强迫被害人与其发生性关系并索要钱财，还将被害人的私密视频通过朋友圈售卖，导致视频在被害人所在学校多名学生间传播，不仅对被害人当下的身心健康产生实在损害，也会对被害人未来与性行为相关的行为和态度的发展产生长期不良影响，损害被害人性健全发展。

1.2. 最有利于未成年人原则的实践要求

约翰·埃克拉尔围绕儿童的基本利益（对身体、情感和治理关怀的需求）、发展利益（良好的成长环境和发展机会）以及自治利益（选择自己生活方式的自由）三个层次的内涵，层层递进式对儿童最大利益作出解释[11]。自从1990年加入联合国《儿童权利公约》，我国政府部门结合国际通行的儿童利益最大化原则，确立了具有本土特色的最有利于未成年人原则[12]。2023年5月24日，最高人民法院、最高人民检察院、公安部、司法部联合印发《关于办理性侵害未成年人刑事案件的意见》第2条规定：“坚持最有利于未成年人原则，充分考虑未成年人身心发育尚未成熟、易受伤害等特点，切实保障未成年人的合法权益。”最有利于未成年人原则在中国特色少年司法的语境中，具体表现为对未成年人特殊、优先保护以及有利于其身心健康发展[13]。

对未成年人的特殊保护就是基于未成年人主体的特殊性，法律赋予其有别于成年人的特别权利，并为了其充分行使权利创造有利条件[13]。社交媒体网络平台、跨平台通讯工具、短视频类应用程序等网络平台在软件许可及服务协议中通常会专门明确特殊保护未成年人的条款，要求用户不得发布侵害未成年人合法权益或者损害未成年人身心健康的内容，并制定专门的未成年人隐私保护条款。此外，在未成年人行驶合法权利时给予适当限制，18周岁以下的未成年人在使用网络平台前需要取得家长或法定监护人的书面同意。

未成年人的优先保护是指未成年人权利与成年人权利并非出于同等分量级别，二者发生冲突或存在紧张关系时，若无法实现平衡则应把保障未成年人权利置于更优先的地位[14]。具体到社交媒体网络平台、跨平台通讯工具、短视频类应用程序等网络平台，用户具有发送语音短信、视频、图片、表情和文字等服务的权利，但是当用户使用网络平台所提供的服务的权利与未成年人权利相冲突时，对未成年权利的保护优先于用户使用网络平台所提供的服务的权利，用户不得利用网络平台所提供的服务进行“隔空猥亵”未成年人行为，不得侵犯未成年人的权利。

我国未成年人互联网普及率较高，同时未成年人对于性问题的认知和理解程度相对较低，更容易在网络平台遭受性关系相关的侵害后留下身心健康创伤。根据共青团中央发布的《第6次中国未成年人互联网使用情况调查报告》显示，2023年我国未成年网民规模为1.96亿，未成年人互联网普及率达97.3%，39.5%的未成年网民会在网络平台发布笔记或短视频，58.6%的未成年网民经常在网上聊天[15]。据调查显示，我国48.4%的青春期未成年人遭遇过“不想谈时，有人在网络上与其试图谈论性话题”，20%的被调查者在青春期遭遇过

“不想看时，有人在网络给其发送他（她）私密部位的照片、视频或黄色图片、视频”[16]。网络平台如此高概率的性诱惑容易使得未成年人对于性问题的认识和理解产生认知偏差，对其身心健康产生恶劣影响。

1.3. 法律3.0时代技术措施的规制补充

在人工智能时代，智能监控、人脸识别、用户画像分析、“码治理”等各类算法和“数治”技术越来越多地被私人机构和公共部门运用，对社会治理和个人生活产生了重大而深远的影响。运用人工智能辅助预防“隔空猥亵”未成年人行为是从技术发展的视角观察违法行为后得出的弥补法律规范的滞后性和保障未成年人权利的实践方法，是目光不断在法律、规制和技术之间往返流转而得出的启示。

罗杰·布朗斯沃德在《法律3.0：规则、规制和技术》一书中首次提出了“法律3.0”的概念，即技术措施被用作规制问题的解决方案，并认为其发挥着支持、补充或替代法律规则的作用，从而与传统的法律规则、思维模式乃至法治理念均处于并存、协作或冲突的复杂关系[17]。他认为，“法律兴趣领域”共经历了法律1.0、法律2.0、法律3.0三个阶段。法律1.0阶段奉行融贯主义的思维模式，是关于将法律的一般原则（及其更具体的规则）适用于具体的事实情况[16]。工业革命后新技术的出现和应用颠覆了法律1.0阶段的部分原则和规则，随着农业、小规模社会中的制度在19世纪末期处理纠纷时的不适合性逐渐凸显[18]，法律2.0阶段到来。法律2.0阶段奉行规制工具主义的思维模式，对话的内容不仅是法律内部一致性和一般法律原则的适用，也包括关于规则是否适合用于新兴技术的目的[17]。在人工智能时代，无处不在的算法和蓬勃涌现的新技术不仅是法律2.0视角下规制的对象，也在塑造着人类自身，法律3.0阶段强调融贯主义、规制工具主义、技术主义思维模式的共存与对话，在法律原则之下利用技术措施补充法律规制来实现规制目标。

人工智能发展可以分类为三种主要类型：限制领域人工智能、通用人工智能和超人工智能[19]，现在我们已经进入限制领域人工智能类型阶段，并向通用人工智能类型阶段发展。自2022年底以来，ChatGPT、文心一言、Kimi Chat、DeepSeek R1等生成式人工智能产品陆续发布，标志着限制领域人工智能的蓬勃发展。2024年底，Large World Model、Genie 2世界模型产品发布，标志着人类在迈向通用人工智能的道路又进了一大步。随着人工智能技术的发展与人工智能产品的应用，人工智能正在越来越深入地影响着人类的行为。例如，社交媒体网络平台使用人工智能技术基于设备信息、IP地址、位置信息、个人主页信息以及用户行为信息等分析用户偏好，推送个性化内容，吸引用户的注意力，塑造用户的浏览习惯。那么，社交媒体网络平台、跨平台通讯工具、短视频类应用程序同样可以基于涉及“隔空猥亵”未成年人行为的图片、视频、语音等数据运用随即梯度下降等算法训练人工智能，使人工智能具备识别“隔空猥亵”未成年人行为的能力，进而塑造潜在的行为人和未成年人的行为方式，将“隔空猥亵”未成年人行为从“不应”变为“不能”。

2. 人工智能辅助预防“隔空猥亵”未成年人行为的可行性

“隔空猥亵”行为自身具有显著特征，这为人工智能识别提供了可能。同时，人工智能技术可以同时处理大量数据，能够将法律层面的“不应”转化为技术层面的“不能”，切实阻止此类行为的发生。此外，人工智能还能从符号性层面调节行为人的道德，影响其行为决策。因此，人工智能在预防“隔空猥亵”未成年人行为方面具备坚实的可行性基础。

2.1. “隔空猥亵”未成年人行为可识别性较强

“隔空猥亵”未成年人行为具有可以分为传播型猥亵、拍摄型猥亵、语音型猥亵、直播型猥亵4种类型，司法实践中的“隔空猥亵”未成年人行为往往属于上述某种类型或多种类型并存。

传播型猥亵行为是行为人利用网络平台向未成年人发送带有性意义的图片、视频、音频等内容的行为。在刘某盛猥亵儿童案中，刘某盛为寻求刺激，通过QQ搜索添加未成年被害人为好友，2022年1月4日至18日期间，多次发送调戏语言及多个淫秽视频给被害人。最终，结合其其他犯罪行为，法院判决刘某盛的行为构成猥亵儿童罪，判处刘某盛有期徒刑一年八个月[20]。

拍摄型猥亵行为是行为人利用网络平台诱骗、胁迫未成年人拍摄暴露身体隐私部位的图片或视频。在蒋某猥亵儿童案中，2015年5月至2016年11月，蒋某虚构身份，谎称自己是童星演艺公司的工作人员，代表公司招募童星，在QQ聊天软件上结识女童，以检查发育情况和身材比例为由，要求被害人在线拍摄和发送裸照。最终，结合其其他犯罪行为，法院判处蒋某犯猥亵儿童罪，判处蒋某有期徒刑十一年[21]。

语音型猥亵行为是行为人利用网络平台发送音频、语音通话等方式，对未成年人表达带有性意义的话语，或者要求未成年人发送音频、语音通话做出与性行为相关的语言表述。在某学校教师猥亵未成年人学生案中，2012年至2022年，某学校教师A某利用其对特长生的优先升学推荐权，为满足性刺激，多次向学生甲提出“语音做爱”，甲再三拒绝，但由于害怕A报复，遂接受。之后在A的要求下，甲又通过语音通话自己对自己实施性行为并配合发送喘息声、淫秽话语[22]。

直播型猥亵行为是行为人诱骗、胁迫未成年人在与其进行视频通话的过程中进行暴露身体隐私部位或做出与性行为相关的动作的行为。在乔某某以视频裸聊方式猥亵儿童案中，2014年3月至8月，乔某某通过QQ添加未成年人为其好友，并冒充生理老师以视频教学为名，先后诱骗多名未成年人与其视频裸聊。最终，法院判决乔某某犯猥亵儿童罪，判处乔某有期徒刑四年[23]。

通过以上案件，结合学理分析，“隔空猥亵”未成年人行为的行为特征较为明显，从信息的传播方式可以依次分为：说脏话、讲色情笑话等延时文字信息；喘息声、模拟性行为的性冒犯等延时语音信息；暴露性器官等敏感部位的照片、视频等延时图片、视频信息；包括抚摸、性交等性活动的延时图片信息；暴露性器官等敏感部位和具有性意味的行为的视频直播信息[24]。除含有性意义的语言不具有明确的识别标准外，模拟性行为的喘息声、暴露性器官的图片或视频画面均具有较强的可识别性，人工智能可以通过声纹、图像分析等技术较为准确的辨别上述行为，从技术层面遏制“隔空猥亵”未成年人行为。

2.2. 人工智能的技术功能将“不应”变为“不能”

在人工物的物质性层面，通过使用人工智能技术，可以将“隔空猥亵”未成年人行为从法律层面的“不应”变为技术层面的“不能”，从而有效预防此类犯罪。布鲁诺·拉图尔在描述人工物对人类行动的影响时，引入了“脚本”的概念，如同电影或剧场演出的脚本，人工物规定使用者在使用它们时该如何行动[25]。譬如，减速带的脚本是“当你靠近我时，请减速”，一个一次性纸杯的脚本是“用后请扔掉我”。当脚本工作时，物作为物质的物，而不是非物质的符号在调节行动。减速带致使司机减速的原因不是其在人与世界的关系中表示的意思，而是由于汽车高速经过减速带时会产生强烈颠簸。使用后丢弃一次性纸杯并不是因为产品说明书告诉我们应该如此做，而是因为一次性纸杯清洗后会漏水，无法再次使用。

人工智能辅助预防，可以让远程操控未成年人自行实施暴露隐私部位、模拟性行为动作等具有性意义的行为，或迫使受害人观看行为人、他人做出的色情行为从“不应”变为“不能”。

根据《网络安全法》第24条的规定，为用户提供信息发布、即时通讯等服务的网络平台应当在与用户签订协议时要求用户提供真实身份信息。网络平台根据用户提供的身份信息通过“人脸识别”等手段进行身份识别后，可以判断哪些账号是未成年人注册的账号。对于有未成年人用户参与的线上通讯，可以利用人工智能实时扫描聊天内容中涉及的语音、图片、视频信息，通过如卷积神经网络等深度学习模型检测裸露、性暗示语音、图像或视频，即使内容经过模糊、裁剪处理，仍能通过肢体动作、场景特征进行判断。对于检测到的涉及裸露、性暗示语音、图像或视频，进行屏蔽处理，并在屏幕页面内显著位置提示上述内容发送者其发送的内容涉嫌违反法律法规和平台规范。此外，人工智能可以为未成年人扮演“虚拟监护人”角色，当检测到未成年用户被诱导发送个人信息或涉及性裸露、性暗示语音、图片、视频信息时，自动弹出风险提示，并且限制此类敏感信息的传输功能。

在视频通话的场景中，人工智能可以通过视觉动态检测，分析行为人和未成年人的人体姿态实时监测视频中的不当触摸、暴露隐私部位等裸露、性暗示动作，并识别视频通话画面中出现的成人用品、性暗示道具等物品，结合私密空间布置等环境特征评估风险。此外，通过语音分析视频通话中的对话内容，识别是否有“脱衣服”、“私下见面”等敏感词汇，并且通过声纹分析技术分析判断对话中的语气，是否出现语调急促、猥亵性措辞、喘息声、成人内容播放声等异常声音。如果出现以上异常画面或声音，人工智能应当即时进行干预。当检测到裸露画面或敏感动作时，立即对违规区域打码，并向双方发送“检测到不当内容，已屏蔽”的警示信息。对持续多次出现异常画面和语音对话的视频通话，人工智能应当自动切断通话并记录证据，同时向监护人和网络平台举报。

2.3. 人工智能可以调节行为人的道德

在人工物的符号性层面，人工智能可以促成人的道德决定和道德行动。人作为道德层面的责任主体，具有意向性和实现其意向的自由两个层面能力[3]。由于“隔空猥亵”未成年人行为损害了未成年人的合法权利，如果所有人从事该行为，人类的性伦理将会崩溃，因此该行为违反了康德的普遍法则原则。此外，“隔空猥亵”未成年人行为仅仅将未成年人受害者当作行为人满足性欲或性刺激的工具，贬低了未成年人的人格尊严，因此该行为也违反了康德的人性原则。所以，除了使用人工智能物质性的技术手段调节“隔空猥亵”未成年人行为的行为人实现其意向的自由外，也应当利用人工智能的符号性层面的作用调节行为人的意向性。

人工物对人的道德决定和道德行动的影响往往不可忽视。公共场所中的监控摄像头，在不知道其运行状况的情况下，其存在本身所代表的符号就对行为人的行为起到了监督作用，促使行为人在做出决定时会更加考虑道德和法律规范，不敢轻易做出不道德或违法的行为。美国旧金山大桥加装护栏后自杀人数降低也是人工物调节人的决定的案例，自1937年建成以来，共有超过1500人在未安装护栏的旧金山大桥跳桥自杀，2018年安装护栏后，根据追踪调研，当年515个想在旧金山大桥跳桥自杀的人只有25个人又选择了其他方式自杀，

其余的490人，在看到护栏后都放弃了自杀的决定并且没有再实施自杀行为。由此可见，人存在于其中的物质环境的符号性特征在很大程度上调节了人的决定和依据该决定产生的行为。

当行为人在网络平台的行为具有“隔空猥亵”未成年人的倾向性时，人工智能应当即时向行为人警示其行为的违法性与后果，影响其道德决定与行动，促使其放弃实施“隔空猥亵”未成年人行为。在监测层面，当监测到相关行为时，人工智能可以向行为人发送精准的警示信息，明确指出其行为可能涉及的违法性以及可能面临的法律后果，让行为人清楚认识到自己行为的错误性和严重性，从而影响其道德决定，促使其放弃进一步的不当行为。此外，人工智能可以提示行为人其涉嫌“隔空猥亵”未成年人行为的行为会被记录，包括行为发生的时间、内容、涉及的账号等信息，以便平台或相关监管部门在需要时进行追溯和调查，这也在一定程度上对行为人起到了威慑作用，使其放弃实施“隔空猥亵”未成年人行为。对于人工智能识别的出的有多次涉嫌“隔空猥亵”未成年人行为倾向的行为人，网络平台应当限制或禁止其使用部分或全部功能、账号封禁直至注销、回收账号的处理，并公告处理结果。

3. 人工智能辅助预防“隔空猥亵”未成年人行为的发展可能性

从理论支撑角度来看，三层监管责任理论、“无知之幕”理论和哈贝马斯协商民主理论为人工智能的发展提供了有力的指引。三层监管责任理论确保人工智能使用向善，“无知之幕”理论促进人工智能机器学习向善，哈贝马斯协商民主理论推动“技术民主”。在这些理论的指导下，人工智能辅助预防“隔空猥亵”未成年人行为在未来有望不断完善和发展，为保护未成年人权利发挥更大的作用。

3.1. 三层监管责任理论促进人工智能使用向善

利用人工智能辅助预防“隔空猥亵”未成年人行为是运用技术措施规制人的行为，需要防止技术运行和发展过程中不合目的的变异。雅克·埃吕尔在《技术社会》的前言中指出，我们需要警惕那些“致力于寻找不断改进的手段，以达到未经仔细审查的目的”的技术[26]。为了促进人工智能等技术向善，罗杰·布朗斯沃德提出了“三层监管责任理论”，在第一层，监管者对任何类型的人类社会共同体均负有一种“管护”责任；在第二层，监管者有责任尊重某一特定人类共同体的基本价值；在第三层，监管机构有责任在不同的合法利益之间追求一种可接受的平衡[17]。

第一层是对公共品的监管责任，即保护人类生存的基本条件、人类能动性和自我发展的一般性条件以及道德能动性发展和实践的必要条件。人类需要足够的自我感和自尊感，以及对同伴的充分信任和信心，还要有足够的可预测性来作出规划，从而以互动和有目的的方式行动，而不仅仅是防御性的。在使用人工智能辅助预防“隔空猥亵”未成年人行为的过程中，一方面，人工智能识别到的关于未成年人的具有性意味的语音、图片、视频等仅能用作预防“隔空猥亵”未成年人行为保护未成年人合法权利的目的，在数据收集、处理和储存过程中，对未成年人及相关用户的个人信息应进行严格保密。另一方面，应当保护网络平台用户的言论自由，仅能使用人工智能技术用于识别和分析涉及“隔空猥亵”未成年人的的图片、语音和视频信息，不应出于其他目的对用户的信息传输内容进行监视。

第二层是对尊重特定共同体的基本价值，正如每一个人类能动主体都有能力发展自己的独特身份一样，将这一点扩大到人类能动主体所在的共同体也是如此。当用户同意网络平台的服务协议并根据网络平台的要求注册账号，就意味着用户有在不违反法律法规的基础上使用网络平台提供的文字、语音、视频、图片表情等功能的权利。在网络平台的用户这个共同体中，自由地使用平台提供的服务不受平台的不合法不合理干预是用户注册使用平台的目的和基本价值。网络平台应当通过强化训练提升人工智能识别、分析涉及“隔空猥亵”未成年人行为相关的语音、图片、视频的能力，避免识别错误，将与性意味无关的内容误认为与性意味相关，进而损害用户的正常使用权利。

第三层是寻求可接受的利益平衡的监管责任。人工智能辅助预防“隔空猥亵”未成年人行为涉及言论自由与未成年人性自主权、身心健康之间的平衡，也涉及“隔空猥亵”未成年人行为和成年人之间互相发送带有性意味的信息的区别。出于最有利于未成年人原则的要求，利用人工智能技术限制用户向未成年人发送带有性意味的语音、图片、视频，或限制未成年人发送带有性意味的语音、图片、视频，但不应利用人工智能技术识别、分析、限制用户或未成年人发送的文字信息。一方面，人工智能作为技术措施补充法律规范不应当超出法律规范的范围，人工智能辅助预防“隔空猥亵”未成年人行为不应超出强制猥亵罪和猥亵儿童罪的规范范围，不应介入并不会对未成年人性自主权和身心健康产生显著不良影响的文字信息领域。另一方面，文字含义具有多样性、句子具有歧义性，文字信息的性意味并没有语音、图片、视频等方式的明确标准。此外，由于成年人具有较强的自主判断能力与独立行为的能力，人工智能不应介入成年人之间互相发送带有性意味的信息的场景中。

3.2.“无知之幕”理论促进人工智能机器学习向善

约翰·罗尔斯构思了公正原则形成的思想实验：假设一群人聚集在一起来选择各种原则，同时假设有一道“无知之幕”挡在了每个人身前，导致参与选择的人不知道任何关于自己是谁的信息，不知道自己的职业信息、经济状况、生理特征、受教育程度、政治观点、种族归属或宗教信仰。当不知道任何关于自己是谁的信息的时候，参与选择的人实际上就会从一种原初的平等状态而进行选择。从这种假想的契约中，会产生两种公正原则，第一个原则为所有公民提供平等的基本自由，第二个原则关心社会和经济的平等，只允许那些有利于社会最不利者的社会和经济的不平等[27]。由此可见，对正义原则的确立需要预先利用“无知之幕”遮挡对自身或某个群体特定利益的认识。

“无知之幕”理论同样适用于辅助预防“隔空猥亵”未成年人行为的人工智能的机器学习阶段。在“无知之幕”后，人工智能的设计者需假设自己可能属于网络平台的成年人用户、网络平台的未成年人用户、网络平台的潜在用户等任何群体，避免人工智能算法过分偏向特定的利益集团，打破可接受的利益平衡，过分扩大或限缩辅助预防“隔空猥亵”未成年人的目的。在数据采集、模型训练阶段必须“去身份化”，模糊性别、民族、地域等对“隔空猥亵”行为不产生直接影响的信息，促使算法关注普世性特征，减少歧视风险。

机器学习主要包括无监督学习、监督学习、强化学习这3个领域[28]。无监督学习，机器被直接给予一堆数据，目的是理解数据，找到模式、规律、有用的方式来提炼、表示或可视化数据；监督学习，系统被给予一堆已分类或标记好的例子进行学习，然后用习得的模型对从未见过或尚不清楚基本事实的心例子进行预测；强化学习，系统被置于一个有奖惩的环境中，就像“胡萝卜”和“大棒”并存，目的是找出最小化惩罚和最大化奖励的最优方法。根据“无知之幕”理论，笔者认为在训练辅助预防“隔空猥亵”未成年人行为的人工智能时，应当使用无监督学习的方法。在建立语料库时先过滤掉广告等无效信息，然后人工智能从海量未标记的聊天记录、图片传输、社交网络关系等原始数据中自主学习，识别性诱导、伪装年龄的内容和频繁接触多个未成年人账号的成年人账号等“中心节点”，并以此为依据做出判断与行动。

3.3. 哈贝马斯协商民主理论促进“技术民主”

协商民主理论包括“民主”和“协商”两个方面。民主是指公共决策由一切利害相关者或其代表平等参与和投票来做出；协商是指一切参与者依照理性和无偏私性的原则，经过自由说理后来决策[29]。协商民主致力于“达成正当决策”。利用人工智能辅助预防“隔空猥亵”未成年人行为时，也应当警惕其先进性同时伴随的风险。从中美两国的人工智能实践应用来看，它们存在决策机制透明度、决策因素歧视等问题[30]。为了有效遏制上述问题，应当将哈贝马斯的协商民主理论用于人工智能技术实践的过程当中。

哈贝马斯的协商民主理论强调通过主体间平等对话形成共识的交往理性[31]，这一理论为人工智能“技术民主”提供了重要范式。在人工智能技术深度介入社会运行的背景下，协商民主理论的核心主张——程序正义、共识导向与主体间性——可有效应对算法歧视、数字黑箱和技术垄断等问题。

哈贝马斯提出“理想言说情境”的预设条件，要求所有参与者享有平等表达权且不受强制干预[32]。即每个利害相关者，都应当以反思的姿态，与其他利害相关者一起，进入无偏私的和理性的对话过程。在辅助预防“隔空猥亵”未成年人行为人工智能系统开发与训练中，必须建立多方参与的协商平台，要求人工智能系统设计者、网络平台用户、未成年人、法律顾问、公众代表甚至包括曾被指控“隔空猥亵”未成年人的犯罪嫌疑人或罪犯共同制定技术标准。这种平等参与权不应仅限定于人工智能系统生成后的检验阶段，而应扩展至人工智能系统建立的全过程，从目标设定、模型搭建、数据投喂与应用，既要包括关于如何利用人工智能系统达到辅助预防“隔空猥亵”未成年人行为这个最终目的的协商，也要包括在利用人工智能系统达成最终目的的过程中如何避免人工智能带来的负面影响。

在协商的过程中，不同意见是不可避免的，可能达不成共识。为了避免无休止的协商，尽快得出决策以解决“隔空猥亵”未成年人的紧迫问题，在充分协商后，应当把对共识的要求转化为对多数决的要求。通过投票的方式得出可以被质疑的阶段性共识。协商民主理论的应用在人工智能的自动驾驶领域已有先例，为了回应“电车难题”，德国在经历多场讨论后于2017年发布《自动和网联车辆交通伦理准则》，提出在损害无法避免时，应优先保护人的生命利益，在对人生命利益的侵害无法避免时，不得根据个人特征预先设置选择的权重，并于2021年将这些准则正式写入《自动驾驶法案》[33]。这种阶段性共识机制既能即使回应现实需求，又有利于确保伦理准则随社会发展持续迭代。

此外，传统人工智能监管依赖政府单边规制，易产生监管滞后与技术俘获等问题。根据协商民主理论的要求，应当构建监管部门、人工智能技术设计者和使用者、公众三元监督网络，监管部门提供法律政策框架，人工智能设计者和使用者公开算法影响评估报告，公民通过“数字监督委员会”的参与算法审计。相比于单边规制，三元监督网络拓宽了观察问题的视角，集思广益。

4. 结论

“隔空猥亵”未成年人行为借助网络隐蔽性强、危害大，严重侵犯未成年人的性自主权和身心健康，违背最有利于未成年人原则，也凸显出传统法律规制的滞后性。人工智能辅助预防具有重要意义和实践价值。从必要性上看，它是应对“隔空猥亵”不法行为、践行未成年人保护原则、顺应法律3.0时代发展的必然选择；在可行性方面，“隔空猥亵”行为的特征使人工智能能够有效识别，其技术功能和对行为人道德的调节作用，为预防工作提供了有力保障；从发展可能性而言，相关理论为人工智能的应用和发展指明了方向，有助于实现技术向善。然而，目前人工智能辅助预防“隔空猥亵”未成年人行为仍面临诸多挑战。在技术层面，尽管人工智能在识别相关行为上有一定能力，但仍需不断优化算法，提高识别的精准度和效率，避免误判和漏判。同时，随着“隔空猥亵”手段的不断翻新，人工智能需要具备更强的适应性和学习能力，及时识别新型犯罪行为模式。在伦理和法律层面，如何在保护未成年人权利的实践中，平衡好与用户言论自由等权利的关系，也是不可忽视的问题。

未来，应加强技术研发投入，鼓励科研机构和企业合作，推动人工智能技术在预防“隔空猥亵”未成年人行为领域的创新发展。通过开发更先进的声纹、图像分析技术，以及结合大数据、云计算等技术，构建更加智能、高效的监测预警系统。另一方面，要完善相关法律法规和伦理准则。立法部门应根据技术发展和实践需求，及时制定和修订相关法律，明确人工智能在预防“隔空猥亵”未成年人行为中的权利和义务、责任界定等。同时，建立健全伦理审查机制，确保人工智能的应用符合伦理道德要求。此外，还需加强社会宣传和教育，提高公众对“隔空猥亵”危害的认识，增强未成年人的自我保护意识和能力，形成全社会共同参与、共同防范的良好氛围，从而更好地利用人工智能保护未成年人免受“隔空猥亵”的侵害，为未成年人创造一个安全、健康的网络环境。

参考文献

[1] 刘艳红. 网络犯罪的刑法解释空间向度研究 [J]. 中国法学, 2019(6): 202-223.

[2] 张军. 最高人民检察院关于人民检察院开展未成年人检察工作情况的报告 [N]. 检察日报, 2022-10-30(2).

[3] 维贝克 P P. 将技术道德化: 理解与设计物的道德 [M]. 闫宏秀, 杨庆峰, 译. 上海: 上海交通大学出版社, 2016: 111.

[4] 李川. 网络隔空猥亵犯罪的规范原理与认定标准 [J]. 法学论坛, 2024, 39(1): 51-62.

[5] 钱日彤. 网络猥亵儿童入罪的扩张解释与教义学检视 [J]. 河南警察学院学报, 2023, 32(6): 92-101.

[6] 陈兴良. 刑法教义学中的缩小解释与扩大解释 [J]. 法学论坛, 2024, 39(2): 5-17.

[7] 郭卫华. 论性自主权的界定及其私法保护 [J]. 法商研究, 2005, 22(1): 60-66.

[8] 王凤民. 性自主权的法理思考与现实考量 [J]. 福州大学学报(哲学社会科学版), 2008, 22(4): 61-64.

[9] 李高伦. “性身心保护” 罪质观下网络隔空猥亵儿童行为的刑法处罚 [J]. 预防青少年犯罪研究, 2023(1): 4-14.

[10] 吴刚, 黄健. 社会性学习理论渊源及发展的研究综述 [J]. 远程教育杂志, 2018, 36(5): 69-80.

[11] 吴啟铮. 少年司法中的协作型儿童利益保护机制：以儿童最大利益原则为基础 [J]. 法治论坛, 2019(2): 223-240.

[12] 张启飞, 胡馨予. 我国福利型少年司法处遇制度的检视与形塑 [J]. 河南财经政法大学学报, 2022, 37(6): 102-111.

[13] 马雷. 最有利于未成年人原则在网络隔空猥亵儿童案件中的适用 [J]. 青少年犯罪问题, 2023(5): 111-121.

[14] 苑宁宁. 最有利于未成年人原则内涵的规范性阐释 [J]. 环球法律评论, 2023, 45(1): 141-155.

[15] 宋莉. 报告：未成年人网络普及率达97.3%建议引导 “小网民” 用好互联网 [EB/OL]. (2024-11-24) [2025-02-13]. https://baijiahao.baidu.com/s?id=1816594015027879399& wfr=spider& for=pc

[16] 朱光星. 网络隔空猥亵儿童的定罪研究：以保护儿童为分析视角 [J]. 中国政法大学学报, 2022(4): 206-217.

[17] 布朗斯沃德 R. 法律3.0：规则，规制和技术 [M]. 毛海栋, 译. 北京: 北京大学出版社, 2023: 29, 45-53, 103-112.

[18] VINEY G, GUÉGAN-LÉCUYER M. The development of traffic liability in France [M]// MARTIN-CASALS M. The development of liability in relation to technological change. Cambridge: Cambridge University Press, 2010: 50.

[19] 巴菲尔德 W, 帕加洛 U. 法律与人工智能高级导论 [M]. 苏苗罕, 译. 上海: 上海人民出版社, 2022: 221.

[20] 广东省高级人民法院. 刘某盛猥亵儿童案：2023-02-1-185-003 [EB/OL]. (2023-02-01) [2025-02-13]. https://rmfyalk.court.gov.cn/view/content

[21] 最高人民法院. 蒋某猥亵儿童案：2023-02-1-185-005 [EB/OL]. (2023-02-01) [2025-02-13]. https://rmfyalk.court.gov.cn/view/content

[22] 潘孝舜, 沈澄, 沈勐儿. 特殊职责人员 “隔空猥亵” 未成年人入罪问题研究 [J].中国检察官, 2023, 34(4): 27-31.

[23] 利用互联网侵害未成年人权益的典型案例 [N/OL]. 人民法院报, 2018-06-04 [2024-05-01]. https://www.jcrb.com/xueshupd/jx/201806/t20180604_4885226.html

[24] 刘伟. “隔空猥亵” 目的性扩张解释的限缩规制 [J]. 青少年犯罪问题, 2024(6): 110-122.

[25] LATOUR B. Where are the missing masses? The sociology of a few mundane artifacts [M]// BIJKER W E, LAW J. Shaping technology/building society: studies in sociotechnical change. Cambridge, MA: MIT Press, 1992: 225-259.

[26] ELLUL J. The technological society [M]. New York: Vintage Books, 1964: 6.

[27] 罗尔斯 J. 正义论 [M]. 何怀宏, 何包钢, 廖申白, 译. 北京: 中国社会科学出版社, 2009.

[28] 克里斯汀 B. 人机对齐：如何让人工智能学习人类价值观 [M]. 唐璐, 译. 长沙: 湖南科学技术出版社, 2023: 9.

[29] 翟小波. 民主论脉络内的协商民主论 [C]// 高全喜. 从古典思想到现代政制：关于哲学，政治与法律的讲稿. 北京: 法律出版社, 2008: 623-651.

[30] 张玉洁. 规 “智” ：人工智能的法律挑战与回应 [M]. 北京: 社会科学文献出版社, 2022: 7.

[31] 哈贝马斯 J. 在事实与规范之间：关于法律和民主法治国的商谈理论 [M]. 童世骏, 译. 北京: 生活·读书·新知三联书店, 2003.

[32] 哈贝马斯 J. 交往行为理论：第一卷 [M]. 曹卫东, 译. 上海: 上海人民出版社, 2018.

[33] 郑志峰. 自动驾驶汽车“电车难题”的伦理困境与立法因应 [J]. 法学, 2024(11): 107-123.