Volume 40 | Issue 3 Article 2

3-20-2024

Research on artificial intelligence crime and China's countermeasures

Jianxin GAO
Beijing Municipal Public Security Bureau, Beijing 100006, China

Research on artificial intelligence crime and China’s countermeasures

Abstract

The rapid development of artificial intelligence technology has constantly given rise to new scenarios, models, and markets, changing the way information and knowledge are generated. Nevertheless, the security risks exposed by technology, such as algorithm bias, data leakage, false content generation, and improper use, are also prone to trigger various new types of crimes. There are still loopholes in legal regulation and technological prevention under the current situation, which poses severe challenges to crime crackdown. In order to effectively meet the new challenges of China’s artificial intelligence (AI) crime, we should supplement and improve the existing legal norms, improve the technical prevention ability, strengthen supervision and personnel training, expand the scope of international cooperation, and steadily improve the AI crime prevention ability.

Keywords

artificial intelligence (AI) crime ; response strategies ; legal norms ; monitoring and early warning ; cooperation and communication

Authors

Jianxin GAO, Jinping SUN, Yukun CAI, Chongpeng WANG, Yanyan YANG, and Kaiyue WANG

引用格式：高建新，孙锦平，蔡瑜坤，等．人工智能犯罪与我国对策研究．中国科学院院刊，2025, 40(3): 408-418, doi: 10.16418/j.issn.1000-3045.20241025004.
Gao J X, Sun J P, Cai Y K, et al. Research on artificial intelligence crime and China's countermeasures. Bulletin of Chinese Academy of Sciences, 2025, 40(3): 408-418, doi: 10.16418/j.issn.1000-3045.20241025004. (in Chinese)

人工智能犯罪与我国对策研究

高建新　孙锦平　蔡瑜坤*　王崇鹏　杨燕燕　王凯悦

北京市公安局　北京　100006

摘要　人工智能技术的高速发展，不断催生新场景、新模式和新市场，改变了信息和知识的生成方式，但技术暴露出的算法偏见、数据泄露、虚假内容生成、不当利用等安全风险也极易引发各类新型犯罪，现有形势下的法律规制与技术防范仍存在漏洞，为犯罪打击带来严峻挑战。为有效应对我国人工智能犯罪的新挑战，需补充完善现有法律规范，提升技术防范能力，加强监管与人才培养，扩大国际合作范围，稳步提升人工智能犯罪防范能力。

关键词　人工智能犯罪，应对策略，法律规范，监测预警，合作交流

DOI　10.16418/j.issn.1000-3045.20241025004

CSTR　32128.14.CASbulletin.20241025004

近年来，随着人工智能的突破性发展，以ChatGPT与Sora为代表的大模型正在重塑生产和生活方式，并催生了社会危害和犯罪风险的迭代升级。在公安实践中，自2015年起，已出现相关人员利用人工智能技术实施犯罪的案件，并且呈现逐年增长的趋势。人工智能介入各类犯罪后，使各类犯罪的法律风险及治理难度与日俱增。应高度重视人工智能可能带来的威胁风险，并理顺发展与安全、治罪与治理之间的关系[1]。

1 人工智能犯罪概述

随着人工智能技术发展，结合公共安全领域治理实践，本文认为人工智能犯罪主要分为3个类别。① 人类利用人工智能技术进行的犯罪行为。使用人工智能工具或借助人工智能技术实施诈骗、制作传播污秽物品等犯罪。② 针对人工智能系统进行的犯罪行

*通信作者

修改稿收到日期：2024年12月28日；预出版日期：2025年1月7日

为。人工智能研发过程中，研发者通过参数篡改、数据集污染等方式故意制造犯罪工具以实现自身犯罪目的。③ 人工智能独立实施的犯罪。未来可能出现的人工智能体，能够跳出原有设计功能，自主实施犯罪行为或者与使用者、其他人工智能共同犯罪。目前，第3种类型的犯罪虽尚未出现，按照技术发展的逻辑，不排除此类技术将带来相应的安全风险，建议科学界和法律界应联合开展前瞻研究，未雨绸缪，将威胁降至最低。

1.1 人工智能犯罪常见类型

按照现行刑法规定能够规制、规制不足和无法规制的犯罪进行分类，本文对近年国内外出现的人工智能相关犯罪进行统计分析，大多属于上述3种人工智能犯罪类型的第1种类型，即人类利用人工智能技术进行的犯罪行为，主要涉及以下4类案件。

（1）诈骗勒索类。利用人工智能进行诈骗勒索是人工智能犯罪最典型的类型之一。伴随着生成式人工智能模型的发展，深度伪造技术持续迭代，利用人工智能伪造的内容愈发逼真。犯罪分子借助这些模型或技术伪造出视频、图片及声音等虚假内容，骗取受害者的信任后实施诈骗，或是利用伪造的信息进行勒索等犯罪活动，给受害者造成巨大损失。例如，2023年11月，最高人民检察院发布检察机关依法惩治电信网络诈骗及其关联犯罪典型案例中[2]，诈骗团伙利用人工智能语音机器人实施电信网络诈骗的案例共骗取1 437人、3 586万余元。

（2）网络攻击类。攻击者可以使用人工智能技术自动收集被攻击者的身份、社交关系等社会工程学信息，批量生成带有个性化内容的钓鱼邮件，或自动生成恶意软件实施网络攻击。此外，还能使用人工智能技术对目标主机或网络实施自动化的漏洞扫描，展开自适应攻击。利用人工智能技术辅助网络攻击，降低了实施难度，提高了攻击的成功率。2023年5月，日本东京的智慧城市网络遭遇了人工智能驱动的勒索软件攻击，导致东京的地铁系统和交通信号灯系统瘫痪，造成严重的交通拥堵。据研究机构Gartner发布的《2024年第二季度新兴风险排行榜》，人工智能增强的网络攻击已成为全球企业、组织等数字化发展中的最大新兴风险①。

（3）侵犯公民个人信息类。人工智能模型在训练阶段，需要大量的样本作为训练数据，通常会在网络中自动爬取数据，并且在和用户相互的过程中收集用户相关信息，这些数据和信息包含公民个人信息，即人工智能模型获取这些信息的行为易涉嫌侵犯公民个人信息。在使用阶段，“AI换脸”等技术极易被人利用，造成侵犯公民个人信息的犯罪行为。例如，韩国出现多起利用深度伪造技术将女生头像和身体部位合成后进行淫秽影像传播的案件，被称为“新N号房”事件，据报道，涉案人数可能达22万人②。

（4）编造传播虚假信息类。用人工智能技术生成包含文字、图片或视频，甚至含有“AI换脸”“AI换音”等内容的虚假新闻信息，并在社交网络上传播，易对大众认知及社会秩序产生干扰。若传播的虚假信息中包含不实的灾情、警情等内容并严重扰乱社会秩序，可能涉嫌编造、故意传播虚假信息罪。例如，2024年以来，公安部公布10起打击整治网络谣言犯罪典型案例中[3]，有4起涉及违法人员使用人工智能技术生产的虚假内容进行造谣；四川、甘肃等多地公安厅公布了多起网民利用人工智能技术制造谣言的案例，引发舆论持续关注。

1.2 人工智能犯罪态势

（1）案件数量快速增长。随着人工智能技术的不

① 加大预算投入来对抗AI驱动的网络攻击已经势在必行. (2024-08-06) [2024-12-25]. https://www.163.com/dy/article/J8TLG1P00511ALHJ.html.

② 韩国出现“深度伪造”技术下的新“N号房”案件. (2024-08-30)[2024-12-25]. https://m.gmw.cn/2024-08/30/content_1303835019.htm.

断成熟和普及，越来越多的犯罪分子开始利用这一高科技手段进行犯罪活动。奇安信集团《2024人工智能安全报告》[4]指出，2023年，基于人工智能技术进行的深度伪造欺诈增长3 000%，基于人工智能技术生成的钓鱼邮件增长1 000%。在可预见的未来，利用人工智能进行犯罪的案件会越来越多，给社会治理带来严峻挑战。

（2）辐射范围不断扩大。人工智能犯罪从早期主要集中在个人资金安全、数据安全、网络安全等领域，逐步扩展到金融、医疗、交通、政治、军事、社会公共安全等多个方面，截至2024年9月，对法律法规、部门规章中公安机关执法办案相关案件开展2个维度的涉人工智能犯罪分析，① 主案别。复杂案件中起主导作用的案件类别称为主案别[5]，1 071个主案别中的犯罪行为可能涉及人工智能的案别有214个，占比19.98%。② 细分类案件性质。案件本身具有的本质属性，划分案件类型的依据是由实体法的内容决定，故称为细分类案件性质[6]，2 840个细分类案件性质中犯罪行为可能涉及人工智能的有596个，占比20.98%。

（3）犯罪手法加速升级。人工智能作为新一轮技术革命的代表，其模型算法不断加快迭代升级。犯罪分子持续优化算法、提高模型性能，使得犯罪手段更加智能化、隐蔽化。这种快速迭代的趋势，让公安机关在打击人工智能犯罪时面临更大挑战，需不断更新技术手段和侦查方法。

（4）技术门槛持续降低。随着人工智能技术发展和普及，犯罪变得更加多样且高效，可自动化实现复杂犯罪活动，即使不具备高技术背景的人也能参与其中[7]。例如，ChatGPT可以辅助实现网络攻击或实施扫描，编写恶意代码和诈骗脚本；生成式人工智能可实现单张照片、秒级音频生成高逼真的“AI换脸”以实施诈骗。

1.3 人工智能犯罪特点

相较于传统犯罪，人工智能犯罪具有以下4个特点。

（1）伪装程度高。与传统犯罪相比，利用人工智能技术生成内容实施的犯罪活动真实性更强，受害人难以识别出犯罪行为。深度伪造诈骗案例中，犯罪分子利用人工智能进行换脸、换音，冒充他人进行视频、电话诈骗，真实度越来越高，肉眼难以辨别真伪，普通人难以察觉。例如，2024年2月，中国香港发生一起涉及多人“AI换脸”的诈骗案，诈骗者利用深度伪造技术制作受害公司多位高管的视频并邀请受害职员参加视频会议，整个会议仅有1位受害职员是真人，该职员见到与现实容貌相同的虚拟高管，信以为真，根据指示前后转账2亿港币，后向总部查询才知被骗。

（2）智能化程度高。人工智能模型的智能化程度越来越高，由人工智能驱动的网络攻击犯罪可进行自动化漏洞扫描，自动化地与受害者对话实施诈骗，自动化生成个性化的钓鱼邮件，自动化生成虚假信息并传播等。犯罪分子只需将要求告知人工智能模型，人工智能即可智能地执行相应任务，大大降低犯罪实施的难度，减少犯罪成本。

（3）行为过程的不确定性风险更大。许多人工智能算法被称为“黑箱”，其决策过程复杂且不可解释；不法分子还有多种算法可供选择，甚至可以将多个算法结合使用；算法的演变过程快速、非线性，对于使用这些复杂算法的犯罪行为，分析其运行机制更加困难，防护成本也更高。因此，人工智能算法具有多样性、复杂性和“黑箱”等特性导致一些犯罪行为难以被及时发现或追踪，不确定性相比传统犯罪更高。例如，一个人工智能驱动的金融诈骗行为可能通过复杂的算法模式进行诈骗，传统的侦查手段很难理解其工作原理，甚至无法预测其下一步行动；人工智能生成的深度合成视频、虚假新闻、恶意代码等内容，很难追溯到源头。

（4）具备自主决策能力。在传统刑事案件中，犯

罪主体只能是法律意义上的人，包括自然人与法人等，对人工智能犯罪来说，除了人类利用人工智能技术或针对人工智能系统进行的犯罪行为，自动驾驶系统或具身智能机器人等由人工智能自身系统决策导致的独立犯罪行为也可能发生。虽然生产厂商对自动驾驶或机器人系统进行了大量安全测试与训练，但面对真实世界中更复杂的场景，智能决策机制可能会突破一些人类预先设定的基本原则，这将是人工智能犯罪完全不同于其他传统犯罪的独特之处。

1.4 国外人工智能犯罪治理的经验梳理

鉴于人工智能犯罪高发态势，世界各国在积极拥抱人工智能技术的同时，也在加紧探索人工智能犯罪治理方法，重点从完善规制立法、积极协同联动、加大技术研究、加强案件打击等方面开展治理工作。

（1）国外相关立法情况。国外非常重视人工智能立法工作，密集出台一系列人工智能安全法律法规，筑牢人工智能犯罪治理的法律基础和规范指导底座。① 美国的人工智能法律较为宽松，注重激励创新，以利于抢占科技制高点。美国在人工智能治理立法方面倾向于较为宽松和灵活的监管方式，强调创新和竞争力，整体上更侧重于行业自律和非强制性指导原则。同时，美国也在逐步加强监管力度，特别是在算法歧视和数据隐私方面进行规制。2018年至今，美国相继出台了《禁止恶意“深度伪造”法案》《“深度伪造”责任法案》《“深度伪造”报告法案》③，通过一系列立法对人工智能犯罪，特别是涉及深度伪造技术的犯罪，进行了防范。② 欧盟对人工智能的法律治理更为慎重，通过较严格的法规预防人工智能技术潜在的社会危害。欧盟侧重于通过全面立法来保护公民权利和建立高标准的监管环境，倾向于通过立法实现对人工智能的监管。2021年4月，欧盟发布《关于制定人工智能统一规则并修订某些欧盟立法的条例》，这是全球首个全面的人工智能法律框架[8]。2024年3月，欧盟正式发布全球首个人工智能治理综合性法律《人工智能法案》，该法案基于风险预防的理念，为人工智能构建了一套覆盖全过程的风险规制体系，是欧盟推进人工智能治理、抢占全球人工智能竞争高地的关键举措[9]。③ 其他国家或地区在人工智能治理立法方面各有侧重，反映出各自的政策导向、文化价值观和社会经济发展需求，形成了各具特色的立法方向。例如，2019年，新加坡推出亚洲首部《人工智能治理模型框架》④；2023年12月，加拿大发布的《生成式人工智能技术的基本原则：负责任、可信和隐私保护》明确了开发、提供或使用生成式人工智能的个人信息保护问题⑤。

（2）国际对话合作。全球各国积极推动人工智能安全领域的合作。① 政府和社会协同。美国政府将部分社会人工智能政策和举措编入法律并加以扩展，吸收社会先进方法纳入国家治理；德国通过购买社会人工智能监管服务来实现人工智能犯罪治理，积极成立包含广泛领域国际专家的人工智能国际专家顾问委员会。② 加强国际对话。人工智能技术的跨国界性使得构建国际治理框架尤为重要[10]。2024年9月，美国、英国和欧盟等签署了欧洲委员会制定的《人工智能、人权、民主和法治框架公约》⑥，该公约是全球首个具有法律约束力的人工智能国际公约，旨在确保人工

③ 人工智能时代，如何应对“换脸”危机？. (2024-12-18)[2024-12-25]. https://www.eeo.com.cn/2020/0831/406312.shtml.

④ 中国信通院孙小童等：解读新加坡《生成式人工智能治理模型框架》. (2024-07-04)[2024-12-25]. https://baijiahao.baidu.com/s?id=1803650916369583838&wfr=spider&for=pc.

⑤ 加拿大发布AIGC的基本原则：负责任、可信和隐私保护. (2023-12-11)[2024-12-25]. https://www.secrss.com/articles/61606.

⑥ 美英欧等签署全球首个具有法律约束力的人工智能国际公约. (2024-09-13)[2024-12-25]. https://baijiahao.baidu.com/s?id=1810086997343057511&wfr=spider&for=pc.

智能系统生命周期内的活动完全符合人权、民主和法治，同时有利于技术进步和创新。③ 国际警务合作。人工智能犯罪高复杂性和跨国性需要各国共同应对[11]。2023年6月，国际刑事警察组织与联合国区域间犯罪和司法研究所（UNICRI）共同发布了“人工智能警务创新工具指南”。在该指南指导下，中国与缅甸执法部门合作，成功打击了使用人工智能技术实施电信网络诈骗集团，中国公安机关将犯罪嫌疑人从缅甸押解回国，充分彰显国际警务合作在打击跨国人工智能犯罪的有效性。

（3）提前布局技术研究。世界各国通过成立专门机构、加大投入等方式，多措并举战略性布局人工智能技术研究。生成式人工智能的发展暴露出数据泄露、虚假内容生成等安全风险，这些风险需要通过前瞻性的技术研究和治理机制来应对[12]。随着人工智能技术的快速发展，各国纷纷加强对前沿领域的布局，以抢占战略制高点[13]，掌握未来产业发展的主动权，有效应对人工智能带来的安全风险。例如，美国国家科学基金会建立国家人工智能科学院（NAAI），成员包括国土安全部等政府部门和谷歌公司等科技巨头，组成美国政府重大基础前沿研究的“国家队”，全方位推动国家人工智能技术与产业的快速健康发展⑦；2024年2月，英国国家科研与创新署宣布将投资1亿英镑支持人工智能研究，重点是设立9个人工智能研究中心，以提供下一代创新和技术，使人工智能能够解决从医疗保健到节能电子等应用领域的复杂问题⑧。

2 我国人工智能犯罪治理现状及挑战

2.1 我国人工智能犯罪治理行动

（1）加强立法规制。目前，我国已经发布了一系列法律和政策，走出了一条探索人工智能犯罪治理的法治路径。2017年6月，《中华人民共和国网络安全法》⑨开始实施，规定数据处理和传输过程中，人工智能技术不能被用于危害国家安全或侵犯公民权利等事件。2021年11月，《中华人民共和国个人信息保护法》⑩中明确规定了个人信息处理合法性、透明性和数据最小化等基本原则，对依赖大数据、用户画像等人工智能技术的企业提出严格合规要求。2022年，《关于规范和加强人工智能司法应用的意见》⑪发布，是为了贯彻党的二十大精神和习近平法治思想、落实国家发展规划纲要的具体举措。2023年1月，《互联网信息服务深度合成技术管理规定》⑫实施，对深度合成技术严格监管，要求深度合成内容必须标识，防止其被用于制造虚假信息并误导公众。2023年8月，《生成式人工智能服务管理暂行办法》⑬开始实施，这是中国首个针对生成式人工智能的专门法规，要求提供生成式人工智能服务的公司确保技术安全和可靠，对提供、使用人工智能服务的企业和个人行为进行规

⑦ 美国国家基金会布局人工智能，宣布成立七个新国家人工智能研究所. (2023-05-05)[2024-12-18]. https://baijiahao.baidu.com/s?id=1765014475892691735&wfr=spider&for=pc.

⑧ 英国UKRI新设9个人工智能研究中心. (2024-06-14)[2024-12-18]. http://www.casisd.cn/zkcg/ydkb/kjqykb/2024/kjqykb2405/202406/t20240614_7189061.html.

⑨ 中华人民共和国网络安全法. (2016-11-07)[2024-12-18]. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

⑩ 中华人民共和国个人信息保护法. (2021-08-20)[2024-12-18]. https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm.

⑪《最高人民法院关于规范和加强人工智能司法应用的意见》全文(中英文版). (2022-12-09)[2024-12-18]. https://www.court.gov.cn/zixun/xiangqing/382461.html.

⑫ 国家互联网信息办公室 中华人民共和国工业和信息化部 中华人民共和国公安部 令. (2022-11-25)[2024-12-18]. https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm.

⑬ 生成式人工智能服务管理暂行办法. (2023-07-10)[2024-12-18]. https://www.gov.cn/gongbao/2023/issue_10666/202308/content_6900864.html.

范，明确了数字水印、安全评估、技术检查等监管手段，同时禁止人工智能被用于虚假信息的传播、诈骗等违法活动。

（2）探索共治模式。① 持续推进行业治理。国内头部科技企业推动人工智能自律自治，百度、华为等企业参与了《人工智能产业担当宣言》⑭的发布，强调企业在人工智能治理中的责任，确保人工智能系统的安全、可靠、可控，提高算法的透明性和可解释性；全国网络安全标准化技术委员会发布了《人工智能安全治理框架》1.0版⑮，提供基础性、框架性技术指南，促进人工智能的健康发展和规范应用。② 积极开展国际合作。2023年，习近平主席提出《全球人工智能治理倡议》，提议各国秉持共商共建共享理念，协力共同促进人工智能治理[14]。第78届联合国大会上，中国提出了加强人工智能能力建设国际合作的决议，该决议获140多个国家联署，强调人工智能发展的3项原则，鼓励国际合作和互助，共同提高全球人工智能发展能力。我国每年举办世界人工智能大会，促进全球科学家、企业家的交流合作，共同探讨人工智能的发展与治理，连续3年组织召开“全球公共安全合作论坛（连云港）”⑯，汇聚各国政府、执法部门及学者共商公共安全治理策略，针对人工智能治理达成共识，共同倡议加强应对人工智能潜在风险等领域的合作。

（3）加大技术攻关。人工智能技术作为新兴前沿技术的代表，涉及多学科交叉，其发展融合了基础研究和系统工程研究，但在多个领域的应用并未真正发挥作用，尤其在公共安全领域，人工智能技术赋能警务滞后于人工智能犯罪。国家高站位推动谋划，部署推进“科技兴警三年行动计划（2023—2025年）”[15]，将人工智能技术列为构建公安战略科技力量体系的重要方面。通过人工智能相关的人才梯队建设、科研项目保障及促进科技成果转化等多种方式，加大人工智能相关技术攻坚。公安机关立足国家战略和公共安全需求，研究人工智能技术赋能实战的关键核心技术，加强与中国科学院等科研机构深入合作，探索研发公共安全领域治理工具，以科技手段持续提升公安机关预警、防范、打击、处置能力，充分利用人工智能技术提高人工智能犯罪侦查的效率和精准度，全面提升公安实战科技含量。

（4）加强犯罪打击。近年来，公安机关通过“净网”“夏季行动”等一系列专项行动，打击处理了一批利用人工智能技术开展的网络谣言、电信诈骗、制作传播淫秽色情音视图文等案件，对人工智能相关犯罪形成了有效震慑[3]。尤其是在打击“AI换脸”系列犯罪的过程中，公安机关联合相关全国重点实验室等单位开展重点研究，适时组织人脸识别与活体检测技术的安全测评，测评范围覆盖即时通信软件、网络平台、游戏平台、金融软件等需要进行人脸识别登录验证的系统，及时发现人脸识别验证系统存在的风险隐患，开展专项整治，升级安全保护措施和人脸识别算法，不给不法分子可乘之机[16][17]。2023年8月，公安部新闻发布会上公布，依托“净网”专项行动，侦破“AI换脸”案件79起，抓获犯罪嫌疑人515人，有效遏制了人工智能相关犯罪势头。

⑭ 行业首个《人工智能产业担当宣言》发布. (2021-08-04)[2024-12-18]. https://baijiahao.baidu.com/s?id=1707150456180133392&wfr=spider&for=pc.

⑮《人工智能安全治理框架》1.0版发布. (2024-09-09)[2024-12-18]. https://www.cac.gov.cn/2024-09/09/c_1727567886199789.htm.

⑯ 全球公共安全合作论坛. [2024-12-18]. https://www.lianyungangforum.org.

⑰ 公安机关打击治理电信网络诈骗犯罪成效显著 2023年1至11月破案39.1万起 8月以来发案数连续下降. (2024-01-05)[2024-12-18]. https://www.mps.gov.cn/n2255079/n4876594/n5104076/n5104077/c9367701/content.html.

2.2 我国人工智能犯罪治理面临的挑战

我国在人工智能犯罪治理方面，形成了多主体、多领域协同共治的良好局面，取得显著成效。但随着人工智能发展进入快车道，技术的演进衍生出复杂多变的新型风险，防范化解面临挑战。

（1）人工智能安全法律法规体系有待完善。我国在人工智能立法上进行了有益探索，通过多层级、地域化、领域化立法初步构建了人工智能法律治理框架，但仍存在立法层级低、立法规定落后、体系衔接不畅等问题。2021年起，我国密集出台了多部相关政策法规，对生成式人工智能相关技术应用进行了系统规范，但责任认定与归结等问题不够明确，操作性欠缺，尚未形成统一的生成式人工智能法律框架[17,18]。当前，我国人工智能安全相关规制条款分散在不同层级的法律法规中，缺乏统一的监管法律，难以形成治理合力。同时，现有法规缺乏具体操作标准，导致在实际应用中难以执行和监管。例如，对使用人工智能技术制作传播网络谣言，仍需使用现行民法中侵犯名誉权、诽谤罪等及刑法中诈骗罪、帮助信息网络犯罪活动罪等法条进行起诉。

（2）人工智能监管措施尚未完善。人工智能技术存在着较大的不确定性和不可控性，应该接受严格的安全评估，需要监管来确保这些安全措施得以实施。随着人工智能技术的快速发展，尤其是生成式人工智能的滥用现象日益严重，对现有的监管方式带来了重大挑战。以合成视频或合成语音为例，现有大模型生成工具提高了视频和语音生成内容的合理性和逼真度，降低了虚假信息的生成成本，加大了公共安全领域风险治理难度。目前，大模型测评管控及人工智能犯罪风险评估等领域尚未形成成熟的监管措施，随着技术的进一步发展，人工智能技术模拟真实物理世界的能力将进一步增强，监管措施存在明显的滞后性，难以适用新的安全防范要求，需加强技术应对。

（3）在人工智能犯罪治理领域国际合作面临诸多挑战。在人工智能犯罪治理中，国际合作面临的挑战和冲突主要源于地缘政治因素、文化和伦理差异、技术发展与安全问题以及国际合作机制不足等方面。此外，各国尚无人工智能技术使用的统一标准规范，尚未达成统一的关于人工智能治理的国际条约，不利于人工智能技术快速发展。例如，美国主导的“理念一致国家同盟”[19]与联合国注重全球共识和可持续发展目标的实现，倡导建立一个敏捷、网络化的全球治理机制⑱存在冲突，这种分歧体现在监管模式、技术发展和安全问题的优先级设定等诸多方面；欧盟倾向于以人为本的监管模式，而美国则倾向于自主自治，欧盟和美国在人工智能监管方面的不同态度也导致了跨大西洋监管合作的困难；中国主张数据主权原则，美国采取利益攸关方准则，两国跨境数据流动政策的核心关注、政策基调和战略诉求依然存在根本差异。

3 我国应对人工智能犯罪挑战的对策建议

随着人工智能技术高速发展，我国安全监管刚刚起步，面对治理人工智能犯罪的挑战，可从顶层设计、技术治理、加强监管、人才培养、宣传教育与合作交流等方面综合发力，构建全方位人工智能犯罪治理体系。

3.1 强化人工智能安全顶层设计，完善相关法律规制

① 建立高级别领导指挥体系，在省部级政府机关成立人工智能安全领导小组和专家组，为领导决策提供专业支撑。② 制定专门性法律，借鉴美国和欧盟的应对策略，在早期发展阶段更加关注技术创新，在技

⑱ 联大首份人工智能决议出炉：技术监管的美国角色与全球未来. (2024-04-12)[2024-12-18]. https://www.163.com/dy/article/IVJT2QM80521RRCK.html.

术成熟期和大规模应用时实行更严格的监管，以确保人工智能技术尽量少带来负面社会影响，推动法律法规适应人工智能技术的发展和应用场景变化，采取动态的法律更新机制，定期对法律条款进行更新，以适应新兴技术。③ 对于人工智能的开发者、提供商、使用者等相关方[20]，建立相应的责任追究机制，完善处罚与追责体系；对于现行法律规定规制不足或无法规制的涉人工智能犯罪，采取完善相关司法解释、调整相关犯罪的构成要件及设立新罪名的法律应对策略[21]。④ 创新人工智能治理方式，健全人工智能相关配套政策与技术规范，系统制定和完善与人工智能犯罪治理相关的国家和行业标准，特别是尽快制订人工智能生成、合成内容标识的技术规范及配套的强制性标准，提高生成模型透明度，解决法庭科学中对证据结论可解释的要求。

3.2 加强人工智能安全技术创新，提升技术应对能力

积极发挥人工智能技术的正向赋能作用，依托公安部与科学技术部联合部署推进的“科技兴警三年行动计划（2023—2025 年）”，加强技术攻关与应对。鉴于其“双刃剑”效应，应从 2 个方面加强技术保障。① 利用人工智能技术提升犯罪预防和侦查效率，建立人工智能犯罪发现和处置系统，开展常态化监测预警。② 提升人工智能系统与应用自身安全，构建准确、稳健、安全、隐私、公平和可解释的人工智能算法，加强对软硬件设施的安全防护，定期对人工智能系统进行全面的安全检查，及时发现和解决潜在的安全威胁，以确保人工智能系统满足安全要求。

3.3 加强人工智能系统安全管理，建立分级监管机制

① 参照网络安全等级保护体系，创设动态的人工智能分级分类监管机制，根据生成式人工智能服务的风险高低进行分类分级监管；并根据生成式人工智能服务适用的不同领域进行行业部门监管。分级分类监管和行业部门监管 2 种监管政策相辅相成，共同促进人工智能的体系化监管进一步加强。② 建立一套全面覆盖技术、伦理和社会安全的人工智能风险评估体系，构建人工智能风险防控和应急响应机制。对高风险人工智能技术，可以借鉴欧盟《人工智能法》中的“沙盒监管”模式，“沙盒监管”可以减少法律对产业发展的负外部性，即在特定范围内允许人工智能系统进行有条件的应用，进行小规模测试，同时在监管机构的指导下调整技术细节。③ 加强人工智能安全相关信息的收集和分析工作，按照危害程度、影响范围等因素对人工智能安全事件进行分级，及时向相关企事业单位通报预警信息，制定相应的应急预案并定期组织演练，以应对人工智能发展中的各类安全风险[22]。

3.4 加强人工智能安全人才培养，提升犯罪打击能力

打击人工智能犯罪需要培养一批“善侦查、专技术、懂法律”的交叉复合型高素质人才，应当优先在公共安全相关高校设置专门的人工智能安全相关专业，持续跟进人工智能及其安全前沿趋势，加强实际案例解析与培训，研究设计科研实践平台与攻防靶场，举办相关竞赛活动，搭建交流学习平台，建立并培养人工智能犯罪打击专门力量，定期开展人工智能犯罪专题研究，多角度应对潜在风险。

3.5 加强人工智能安全宣传教育，提升公众认知能力

① 加大人工智能安全教育与宣传力度，借鉴全民反诈宣传模式，依托知识科普和短视频平台等，提高公众安全意识，使公众能有效识别人工智能犯罪形式、保护个人信息、防范智能攻击，提升辨别能力。② 聚焦数据安全、网络谣言等热点问题，通过线上线下相结合的方式宣传人工智能安全法律法规、政策文件、国家标准等内容，及时通报公安机关打击人工智能犯罪、加强人工智能安全监管的工作成效，提升公众对人工智能安全观和相关法律法规的认识。

3.6 加强人工智能犯罪国际合作，提升跨境处置能力

在全球化背景下，人工智能犯罪常具跨国性，各国需要共同应对。应大力推动国际执法合作，发挥高层互访的引领作用，有效利用有关会晤机制平台，定期与外国执法部门和有关国际组织就人工智能安全、打击跨国犯罪、追逃追赃等议题深入磋商，共享研究成果和治理经验，共同研究制定国际规则和标准，建立人工智能犯罪专属数据库，实时共享犯罪信息，有效凝聚共识、管控分歧，对跨境人工智能犯罪形成有力震慑。

参考文献

1 戚永福, 翁音韵, 曹瑞璇. 生成式人工智能介入网络犯罪的治理难点及应对. 人民检察, 2024, (6): 66-70.
Qi Y F, Weng Y Y, Cao R X. The difficulties and countermeasures of using generative artificial intelligence to intervene in cybercrime governance. People's Procuratorial Semimonthly, 2024, (6): 66-70. (in Chinese)

2 郑雪. AI诈骗新套路不断,"眼见不一定为实"如何防范. 中国经济周刊, 2024, (11): 103-104.
Zheng X. New tricks of AI fraud keep emerging, how to prevent "seeing is not necessarily believing". China Economic Weekly, 2024, (11): 103-104. (in Chinese)

3 杜洋. 守住人工智能技术应用法治底线. 法治日报, 2024-07-04(06).
Du Y. Defend the bottom line of the rule of law in the application of AI technology. Legal Daily, 2024-07-04(06). (in Chinese)

4 刘津宁. AI"复活",边界何在？. 中国妇女报, 2024-04-22(008).
Liu J N. Where are the boundaries of using AI for people's virtual rebirth?. China Women's News, 2024-04-22(008). (in Chinese)

5 卢刚. 司法体制改革背景下法院分案制度之构建——基于传统分案和随机分案的样态检视// 法院改革与民商事审判问题研究——全国法院第29届学术讨论会获奖论文集(上). 北京: 最高人民法院, 2018.
Lu G. Construction of court case division system under the background of judicial system reform: A study of traditional and random case divisions// Research on Court Reform and Civil and Commercial Trial Issues: Selected Papers from the 29th National Academic Symposium on Courts (Part 1). Beijing: The Supreme People's Court of the People's Republic of China, 2018. (in Chinese)

6 陈世华. 确定案件性质的依据和作用. 犯罪研究, 1990, (5): 24.
Chen S H. Basis and function for determining the nature of cases. Crime Research, 1990, (5): 24. (in Chinese)

7 王彩玉, 梁立增. 新加坡人工智能犯罪生态治理的实践与探索. 现代世界警察, 2024, (1): 53-61.
Wang C Y, Liang L Z. Practice and exploration of AI crime ecological governance in Singapore. Modern World Police, 2024, 1: 53-61. (in Chinese)

8 杨延超. 大模型时代, 人工智能犯罪如何治理？. 群言, 2023, (7): 32-34.
Yang Y C. How to govern artificial intelligence crimes in the era of big models?. Popular Tribune, 2023, (7): 32-34. (in Chinese)

9 罗昕. 聊天机器人的网络传播生态风险及其治理——以ChatGPT为例. 青年记者, 2023, (7): 91-94.
Luo X. The ecological risks and governance of network communication of chatbots: A case study of ChatGPT. Youth Journalist, 2023, (7): 91-94. (in Chinese)

10 薛澜, 赵静. 人工智能国际治理: 基于技术特性与议题属性的分析. 国际经济评论, 2024, (3): 52-69.
Xue L, Zhao J. International governance of AI: Analysis based on technical characteristics and issue attributes. International Economic Review, 2024, (3): 52-69. (in Chinese)

11 聂江波. AI犯罪前瞻及人工智能侦查系统构建. 公安研究, 2024, (5): 86-91.
Nie J B. AI crime foresight and construction of artificial intelligence investigation system. Public Security Research, 2024, (5): 86-91. (in Chinese)

12 程乐. 生成式人工智能治理的态势、挑战与展望. 人民论坛, 2024, (2): 76-81.

Cheng L. The situation, challenges, and prospects of generative artificial intelligence governance. People's Tribune, 2024, (2): 76-81. (in Chinese)

13 王恺乐, 陈云伟, 熊永兰. 人工智能监管政策与措施的国际比较及启示. 世界科技研究与发展, 2024, 46(4): 456-468.
Wang K L, Chen Y W, Xiong Y L. International comparison and implications of regulatory policies and measures for artificial intelligence. World Sci-Tech R&D, 2024, 46(4): 456-468. (in Chinese)

14 李猛. 人类命运共同体视角下人工智能风险全球治理的国际法规制路径探究. 宁夏社会科学, 2024, (2): 112-124.
Li M. Exploration of international legal regulatory paths for artificial intelligence risks global governance from the perspective of a community with a shared future for mankind. Ningxia Social Sciences, 2024, (2): 112-124. (in Chinese)

15 邢玉秋, 罗子杰. 公安院校无人机警务实战化教学研究. 中国军转民, 2024, (11): 23-25.
Xing Y Q, Luo Z J. Research on the practical teaching of unmanned aerial vehicle policing in public security universities. Defense Industry Conversion in China, 2024, (11): 23-25. (in Chinese)

16 张梓琪. C市电信网络诈骗协同治理研究. 长春: 吉林大学, 2023.
Zhang Z Q. Study on Cooperative Governance of Telecom Network Fraud in C City. Changchun: Jilin University, 2023. (in Chinese)

17 郑秋伟, 李前进, 程晓东. 人工智能驱推思想政治教育变革: 逻辑、趋向与策略. 教育理论与实践, 2024, 44(15): 31-36.
Zheng Q W, Li Q J, Cheng X D. AI driving the transformation of ideological and political education: Logic, trends, and strategies. Theory and Practice of Education, 2024, 44(15): 31-36. (in Chinese)

18 中国信息通信研究院, 京东探索研究院. AIGC赋能百业, 助力产业升级迭代. 大数据时代, 2023, (8): 6-29.
China Academy of Information and Communications Technology, JD Explore Academy. AIGC empowers all trades and fosters industrial upgrades. Big Data Time, 2023, (8): 6-29. (in Chinese)

19 丁迪. 大国竞争、战略稳定与治理合作——美国人工智能多元安全议程的特征、逻辑与战略影响. 国际观察, 2024, (4): 1-30.
Ding D. Big power competition, strategic stability and governance cooperation—Characteristics, logic and strategic impact of the US artificial intelligence multi security agenda. International Observation, 2024, (4): 1-30. (in Chinese)

20 于波, 应雨晴, 程得琳. AI应如何"合理"运用他人作品. 中国新闻出版广电报, 2024-03-21(A5).
Yu B, Ying Y Q, Cheng D L. How AI should properly use others' works. China Press Publication Radio Film and Television Journal, 2024-03-21(A5). (in Chinese)

21 何镭. 涉人工智能犯罪刑事责任问题研究. 蚌埠: 安徽财经大学, 2021,
He L. Research on Criminal Responsibility of Crimes Involving Artificial Intelligence. Bengbu: Anhui University of Finance and Economics, 2021. (in Chinese)

22 杨建军, 张凌寒, 周辉, 等. 人工智能法: 必要性与可行性. 北京航空航天大学学报(社会科学版), 2024, 37(3): 162-174.
Yang J J, Zhang L H, Zhou H, et al. Artificial intelligence law: Necessity and feasibility. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2024, 37(3): 162-174. (in Chinese)

Research on artificial intelligence crime and China's countermeasures

GAO Jianxin　SUN Jinping　CAI Yukun*　WANG Chongpeng　YANG Yanyan　WANG Kaiyue

（Beijing Municipal Public Security Bureau, Beijing 100006, China）

Abstract　The rapid development of artificial intelligence technology has constantly given rise to new scenarios, models, and markets, changing the way information and knowledge are generated. Nevertheless, the security risks exposed by technology, such as algorithm bias, data leakage, false content generation, and improper use, are also prone to trigger various new types of crimes. There are still loopholes in legal regulation and technological prevention under the current situation, which poses severe challenges to crime crackdown. In order to effectively meet the new challenges of China's artificial intelligence (AI) crime, we should supplement and improve the existing legal norms, improve the technical prevention ability, strengthen supervision and personnel training, expand the scope of international cooperation, and steadily improve the AI crime prevention ability.

Keywords　artificial intelligence (AI) crime, response strategies, legal norms, monitoring and early warning, cooperation and communication

高建新　北京市公安局副局长。主要研究领域：人工智能安全、网络安全等。E-mail: wcpwct@126.com

GAO Jianxin　Deputy Director of Beijing Municipal Public Security Bureau. Main research areas cover artificial intelligence security, network security, etc. E-mail: wcpwct@126.com

蔡瑜坤　北京市公安局网络安全保卫总队副总队长。主要研究领域：人工智能安全、网络安全等。
E-mail: caiyukun24@mails.ucas.ac.cn

CAI Yukun　Deputy Commander of the Network Security Guard Corps of Beijing Municipal Public Security Bureau. Main research areas include artificial intelligence security, network security, etc. E-mail: caiyukun24@mails.ucas.ac.cn

■责任编辑：文彦杰　梁小星

*Corresponding author