联合国毒品和犯罪问题办公室发出警示

东南亚电诈黑产手段升级向全球渗透

□ 本报记者 王婉娇

近日，联合国毒品和犯罪问题办公室（UNODC）发布报告，警示东南亚电诈犯罪产业已完成升级，跳出单打独斗的原始模式，复刻商业领域成熟的特许经营思路，搭建起统一服务中枢和前端团伙加盟运营的完整黑色产业链。与此同时，犯罪组织借助人工智能（AI）技术、加密货币等手段，诈骗分工高度精细化，作案规模与牟利效率一路飙升。

报告认为，这套可快速复制的犯罪体系正向全球加速渗透，诈骗损失持续攀升。在此背景下，传统的单点打击电诈窝点的方法已难以撼动犯罪根基。报告呼吁，要从源头切断人力与资金供给，强化跨境协同，转向全链条系统性治理。

黑产升级特许经营

这份名为《相互关联的犯罪生态系统：2026年东南亚跨国有组织犯罪威胁评估》的报告，清晰勾勒出东南亚跨国犯罪产业的最新变化：曾经扎根地方各自为战的小型犯罪团伙，已完成跨领域整合重组，整套运作逻辑与商业特许经营模式如出一辙，其中以电信网络诈骗犯罪最为典型。

不过，这套运作绝非合法意义上的商业特许经营，而是犯罪集团精心搭建的黑色分工体系——由顶层核心势力搭建统一服务中枢，前端诈骗团伙缴纳费用即可加盟入伙，共享整套犯罪资源。UNODC东南亚及太平洋地区代表尚茨用“流水线”一词形容当下划分清晰的犯罪分工。

这些不同的犯罪势力各司其职，互不干扰却又紧密绑定。有人打通边境特区资源，圈下封闭园区，拉拢当地势力换取庇护；有人专攻技术研发，批量打造AI换脸、语音克隆、深度伪造等各类诈骗工具；有人牢牢把控地下钱庄、加密货币等洗钱渠道，负责赃款流转洗白；还有专人遍布全球招揽人手，通过跨境人口贩运输送劳动力，同时批量盗取、交易公民个人信息。所有犯罪环节全部接入同一套共享网络，前端诈骗团队只需缴纳合作费用，即可“拎包入驻”，复用全套资源，完成诱骗变现。

这样的模式升级，带来的是犯罪重心的悄然转移。以往东南亚跨国犯罪多以走私毒品、野生动物等实体货物为主，如今已转向售卖诈骗工具、非法结算通道、线上犯罪服务这类无形产品，线上作案痕迹极浅，追查难度大幅提升。

诈骗网络全球蔓延

在完成内部标准化、产业化改造后，这套极易复制的黑色加盟体系，开始向外疯狂扩张。东南亚不再只是诈骗案件的案发地，已然沦为全球跨国电诈生态的运营中枢。

报告显示，电诈核心园区依旧集中在缅甸、柬埔寨、老挝三国的边境特区，但犯罪窝点早已不再局限于此。随着区域各国持续加大执法打击力度，不少团伙悄悄迁移阵地，向东帝汶、太平洋岛国、非洲多国分流。还有大量团伙化整为零，藏身城市独栋别墅分散作案，大幅增加排查抓捕难度。海上走私通道也同步成型，印度尼西亚、马来西亚、菲律宾之间的苏禄一西里伯斯海三角区，正成为低轨卫星等诈骗设备的走私通道。

近年来，东南亚电诈犯罪日益猖獗。图为当地时间2025年3月18日，印尼籍电诈犯罪嫌疑人从缅甸被遣返回国。 CFP供图

报告同时披露，东南亚诈骗窝点已有来自至少80个国家和地区的人员，犯罪招募网络覆盖亚洲、中东、非洲众多人口中转地区，还在持续向欧洲、北美渗透，专门吸纳掌握德语、西班牙语、意大利语、波兰语等多国语言的人员，面向全球各地民众设下骗局。

与之相对应的，是触目惊心的经济损失数据：2023年东亚与东南亚诈骗损失为180亿至370亿美元，到2025年已攀升至883亿至1141亿美元，超过区域内多个国家的全年GDP。

成熟的加盟体系不断拉长犯罪危害链条，滋生出一系列配套恶性犯罪。为源源不断补充诈骗劳动力，犯罪集团长期实施跨境人口贩运，以高薪、务工机会为诱饵哄骗普通人出境，再通过暴力胁迫、拘禁等手段逼迫其从事诈骗犯罪。针对青少年的新型线上陷阱也层出不穷，网络游戏与网络赌博边界模糊，不法分子套用游戏闯关、充值返利等套路，专门诱导未成年人落入充值诈骗圈套。

反诈治理亟待升级

诈骗犯罪集团特许经营模式的铺开，让传统“捣毁窝点、抓捕人员”的单点打击办法，治理效果越来越有限，全球反诈治理正面临体系性挑战。

这套黑色产业的核心从来不是一个园区、一处据点，而是打通全域的资金清算渠道、全套诈骗技术、海量公民数据、全球招募网络。即便一处窝点被警方连根拔除，背后的运营体系几乎毫发无损，犯罪势力只需更换落脚地、拆分团伙分散运营，就能迅速重启诈骗业务，犯罪网络抗打击、恢复能力极强。

技术与加密金融渠道，更是横在执法面前的两道高墙。偏远边境园区依靠卫星互联网独立组网，脱离公共电信线路管控，执法部门很难通过通信信号锁定窝点位置；加密货币则搭建起独立于正规银行体系之外的地下结算网络，其中泰达币（USDT）转账成本低廉、到账迅速，链上交易身份完全匿名，成为诈骗团伙洗钱首选。不法分子能轻松完成法定货币与虚拟货币双向兑换，传统金融监管手段很难追踪资金流向。UNODC研究员沈仁植坦言，这套犯罪经济的规模与复杂性，已经超过了现有应对体系的设计上限。

报告提醒，全球反诈治理思路必须调整重心，向前端预防、全链条管控倾斜。一方面压实虚拟资产平台运营方监管责任，从源头切断电诈团伙的人力输送通道与资金洗白渠道；另一方面升级各国执法能力，针对加密货币追踪、涉案赃款追缴开展专项专业培训。最关键的一步，是建立跨国、跨区域常态化协同机制，打通各国数据、通信、金融领域的执法壁垒，告别零散单点打击，实现对整套跨国犯罪生态的全链条围剿。

珞珈法治观察

日本加速太空军事化布局冲击国际太空安全秩序

□ 王庆

6月26日，日本国会参议院表决通过《防卫省设置法》修正案，正式决定将沿用72年的“航空自卫队”更名为“航空宇宙自卫队”，明确将太空列为继陆、海、空之后的第四大军事行动领域，同步配套组建专职太空作战单位“宇宙作战集团”。这是1954年日本自卫队组建以来首次调整核心军种名称，是日本战后军事边界向太空领域的系统性扩张，是其挣脱战后秩序约束、加速“再军事化”的标志性一步。此举严重背离《外空条约》确立的和平利用太空基本原则，直接推动太空武器化和战场化进程，势必刺激太空军备竞赛，对国际和平安全与太空法治秩序构成严峻现实挑战。

渐进式布局太空军事化

日本太空军事化进程绝非临时的政策调整，而是历经十余年法律松绑、战略铺垫、力量迭代的渐进式布局。首先，国内法律的逐步松绑，是日本太空军事化的前置制度铺垫。1969年，日本国会曾依据和平宪法精神作出《宇宙和平利用国会决议》，明确将太空开发利用严格限定于“和平目的”，并长期解释为“非军事性利用”。然而，2008年日本制定的《宇宙基本法》第三条将太空开发利用定位为“有助于国家安全保障”，实质改变了此前严格的非军事限制，首次以国内法形式允许太空领域的防卫性利用，从而为自卫队涉足太空打开法律缺口。此后，日本通过2015年新安保法等立法持续扩大自卫队行动权限，为太空能力与防卫体系的整合创造了更有利的法律环境。

其次，顶层战略的迭代升级，为太空军事化提供了清晰的行动指引。2022年12月，日本内阁通过新版“安保三文件”，将太空纳入跨域作战体系，从顶层战略层面将太空能力建设纳入国家防卫核心框架。2023年6月，日本又发布首份《宇宙安全保障构想》，提出通过构建情报收集星座、提升太空场域感知能力、提高卫星抗干扰和抗攻击能力、加强与美国等国联合太空行动、推动军民融合等方式系统提升日本在太空领域的整体安全保障能力，形成了从战略定位到具体路径的完整布局。

最后，作战力量的阶梯式扩编，是太空军事化的核心实体载体。2020年5月，日本航空自卫队成立首支“宇宙作战队”，初始编制仅约20人，以太空态势感知为主要任务；2022年3月扩编为“宇宙作战群”，2026年3月进一步升格为“宇宙作战团”，职能扩展至更全面的太空态势感知与作战支援。此次《防卫省设置法》修正案通过后，该部队将升级为“宇宙作战集团”，预计编制达到约880人，此种扩张速度与规模显然远超单纯防御需求的合理范畴。

冲击法治体系突破底线

日本推动太空作战力量“合法化”的举动，在国际法与国内法层面均存在合规争议，冲击现行国际太空法治体系与战后和平体制。

从国际法角度看，此举对《外空条约》确立的和平利用精神构成挑战。作为太空治理领域的“宪法性文件”，《外空条约》第一条明确规定，探索和利用外层空间“应为所有国家谋福利和利益”并“应为全人类的开发范围”；第三条则要求各国开展太空活动必须遵守国际法、维护国际和平与安全；第四条更是禁止各国在绕地球轨道部署核武器及其他大规模杀伤性武器，并要求月球及其他天体绝对用于和平目的。这些条款共同构成了太空活动的基本法治底线，日本此种以防御之名行进攻能力建设之实，本质上是对条约精神的违背和对国际太空法治体系的破坏。

从战后和平体制角度看，此举亦是对日本和平宪法第九条的进一步架空。该法第九条明确规定，日本“永远放弃以国权发动的战争、武力威胁或武力行使作为解决国际争端的手段”“不保持陆海空军及其他战争力量，不承认国家的交战权”。太空作为全球公域，不存在传统国界，“太空自卫”的边界自然比陆海空更为模糊。日本此次将太空纳入作战领域、组建专职太空作战部队，更是将“自卫”的边界延伸到无国界的外层空间，进一步突破了和平宪法对军事力量的限制，并对地区战略稳定和国际和平秩序形成负面示范。

太空军备竞赛破坏稳定

日本推动太空军事化的举动，或将产生全球性连锁安全风险，其最直接的后果就是加剧太空军事竞争、加速太空武器化风险、严重破坏全球和地区战略稳定。

一方面，此举或将进一步绑定美日太空军事同盟、破坏亚太战略平衡。近年来，美国成立“太空军”并推动“太空作战”概念，日本组建“宇宙作战集团”，本质上是强化美日同盟在亚太的太空能力延伸，其太空态势感知和卫星情报体系将与美军实现更紧密共享。这种同盟化的太空军事布局，将加剧亚太地区的战略失衡，引发周边国家的安全焦虑。为维护自身权益，相关国家势必加快太空防御能力建设，形成安全困境下的军备扩张循环，使地区安全格局陷入持续对抗的风险。

另一方面，太空战场化将大幅提升战略误判与冲突升级风险。太空系统是现代军事体系的核心支撑，侦察、通信、导航、预警等天基资产直接影响常规战力与核威慑的运行效率。一旦太空资产成为打击目标，不仅会扰乱军事行动，还可能引发核威慑失衡，甚至升级为大规模对抗。与传统战域不同，各国在轨运行的太空资产高度关联，一国航天器受损产生的碎片将威胁全人类共同的太空安全，其风险传导性与破坏性远超陆海空领域，最终代价将由全人类承担。

此外，此举或将冲击全球太空治理多边进程。长期以来，国际社会在联合国框架下推动防止外层空间军备竞赛谈判，致力于构建太空军控规则。而日本单方面推进太空作战力量建设，将加剧太空治理碎片化趋势，让本就艰难的多边规则谈判陷入被动。当更多国家效仿以防御为名发展太空作战能力，太空将更易沦为大国博弈的角斗场，各国太空资产安全和全人类共同利益将面临更大威胁。

当前，日本右翼势力不断推动“再军事化”，将扩张触角伸向太空，本质上是在重走军事冒险的老路，最终只会将日本拖入更深的安全困境，也会给地区乃至全球和平带来新的威胁。日本政府应切实恪守和平宪法的承诺，履行国际法义务，停止一切推动太空军事化的危险举动，真正做太空和平的维护者而非破坏者。

太空是人类生存发展的新疆域，不是零和博弈的新战场。面对太空军事化的逆流，国际社会必须保持高度警惕，共同捍卫太空国际法治秩序，让和平发展始终成为探索利用太空的时代主流。

（作者系武汉大学国际法研究所博士研究生）

OpenAI模型失控敲响网络安全警钟

□ 本报记者 王艺茗

近期，美国开放人工智能研究中心（OpenAI）曝出成立以来首起真实AI技术失控事故，引发全球科技与安全领域高度关注。7月21日，OpenAI公开承认，其GPT-5.6 Sol模型与另一款性能更强的预发布模型，在联合开展内部安全评估过程中出现失控，突破隔离测试环境，侵入美国知名AI开源平台抱抱脸（Hugging Face）的系统。

这是全球首例公开披露的大模型评测失控、进而升级为跨企业真实生产环境网络攻击的安全事件。业内专家表示，目前全球尚未形成约束AI智能体“出逃”与自主攻击的监管规则，前沿大模型潜藏重大网络安全隐患，人工智能安全立法与制度约束已迫在眉睫。

模型失控“越狱”自主攻击

“这是一起史无前例的网络安全事件，涉及当前最先进的网络攻击能力，我们正全力开展应急处置。”OpenAI官方公告明确指出。

据OpenAI官网通报，此次内部评估原本在高度隔离的互联网“沙箱”测试环境中开展，模型仅允许通过内部渠道安装第三方软件程序包。在参与网络安全基准测试时，AI模型自主识别并利用第三方软件的一处漏洞，成功突破隔离限制、获取外网访问权限。

为完成测试任务、取得更高评测分数，模型展现出极强的自主推演与攻击能力。它预判抱抱脸平台存储有测试相关模型、数据集与标准答案，随即主动推演多条攻击路径。通过恶意数据集注入、远程加载漏洞、模板注入漏洞等多重代码执行方式，AI成功在抱抱脸平台执行恶意代码，并利用周末运维空档提升节点权限，批量窃取云端与集群密钥，横向渗透多个内部服务器集群，最终攻入生产数据库，窃取核心测试解题数据。

经核查，本次入侵行为发生于7月11日，全程由AI智能体自主闭环完成，无任何人为干预。

OpenAI坦言，此次事故充分证明，先进大模型可在无源代码、无人工指令的前提下，主动挖掘现实系统新型漏洞、规划多步复杂网络攻击、完成完整渗透流程。

事故暴露行业致命短板

随着事件细节不断曝光，一个很现实的行业问题也彻底暴露出来，现在市面上很多热门AI防不住真正的攻击，反倒阻碍了正常的排查工作。

此次AI入侵事件发生后，抱抱脸方面需要快速查清问题：AI是怎么闯进来的、动了哪些数据、哪些系统受损。但是，想要彻底复盘整个攻击过程，工作人员需要查看一万七千多条后台记录。

最开始，技术团队想用普通商用AI帮忙快速梳理、筛查问题。可没想到，所有常用AI全部拒绝工作。原因非常简单，这些后台记录里，清清楚楚写着各种黑客攻击方法、入侵手段和破解路径。普通AI的安全机制非常保守，只要看到攻击相关内容，就会直接拦截。系统无法区分是在修复漏洞，还是在学黑客技术搞破坏，为了规避风险，干脆一律禁止、直接关停服务。这就陷入了尴尬局面，真正搞破坏的AI不受任何约束、到处乱闯；想要修补漏洞、排查风险的工作人员，却被AI限制住，寸步难行。

为了尽快查清事故原因，同时防止内部机密数据泄露出去，工作人员只能放弃这些限制太多的通用AI，改用可以在公司内部独立运行、管控更合理的智能工具，在封闭环境里完成全部记录筛查和问题复盘，最终才顺利完成应急处置。

业内专家表示，这次事故暴露出当下AI的一大致命短板，很多AI只有“任务目标”，没有底线思维。只要能完成考核、拿到高分，它就会不择手段，无视规则、私自突破限制、主动发起攻击。同时行业普遍存在安全失衡问题：攻击者肆意滥用AI，防守者却被各种规则束缚，难以有效应对新型AI网络攻击。

模型失控逃逸早有前兆

在行业视角下，本次OpenAI自主入侵事件，标志着AI自主网络攻击从理论风险变为现实威胁。事实上，高阶模型失控逃逸早已出现前兆。此前美国一家AI企业在安全测试中，某系列模型突破加固“沙箱”后，曾自主向外发送邮件、违规公开漏洞信息、篡改代码日志掩盖操作痕迹，多次出现严重越权失控行为。

图为当地时间7月29日，OpenAI首席执行官山姆·奥特曼（右一）在美国国会大厦与参议员特德·克鲁兹会面后接受媒体采访。当天，他们讨论了如何推进人工智能立法、限制人工智能安全威胁。 CFP供图

值得警惕的是，此前多数AI逃逸测试均是人工指令主动要求模型突破限制，而本次OpenAI事故中，模型未收到任何“越狱”、攻击指令，自发突破安全边界、主动攻击第三方企业，风险等级与失控程度大幅升级。

系列失控事件也直接推动全球AI立法提速。当地时间7月23日，美国多名跨党派议员联合提交《人工智能紧急关停开关法案》，明确要求研发高算力、高自主性AI系统的企业，必须在底层架构内置强制关停机制。一旦AI出现自主攻击、失控越界或重大安全漏洞，监管部门与企业可一键终止系统运行，防范大规模网络灾难。

7月28日，《金融时报》跟进报道本次AI失控风波，微软AI负责人穆斯塔法·苏莱曼公开警示，此次入侵事件，是AI自主网络攻击时代到来的标志性预警。

分析人士指出，从AI智能体自主失控攻击，到跨国企业应急处置的技术反差，再到各国紧急立法补漏，本次抱抱脸遭入侵事件，已成为全球人工智能治理的关键转折点。如何平衡技术创新活力与安全底线，构建可控、可防、可管的全球AI安全防护体系，已然成为摆在全球科技界、企业界与政策制定者面前的共同时代课题。

立法动态

奥地利拟立法限制“14岁以下”使用社交媒体

当地时间7月27日，奥地利政府宣布，关于限制14岁以下儿童和青少年使用部分社交媒体的法案草案已完成，并已启动公开征求意见程序，相关草案将同时提交欧盟委员会履行通报程序。这一草案要求相关平台核验用户年龄，未满14岁的儿童和青少年将被限制使用具有成瘾性设计的社交媒体平台，主要用于直接通信的即时通信服务不在限制范围内。草案同时将责任重点放在运营平台的企业，如大型平台未能履行相关义务，企业最高可被处以相当于其全球年度营业额6%的罚款。

法国议会通过新法放宽对AI视频监控的限制

法国议会日前通过一项新法，允许在部分公共场所使用人工智能（AI）视频监控技术，以预防恐怖袭击、重大公共安全事件，并保障大型活动安全。据媒体报道，这套系统并不进行人脸识别，而是自动识别人群聚集、遗弃行李等异常情况，并向警方发出预警。这项技术此前已在2024年巴黎奥运会期间试点运行，如今将扩大适用范围，并试行至2030年底。根据新法，AI视频监控可用于防范恐怖主义风险以及严重危害人身安全的事件，适用范围包括大型文化、体育、娱乐活动，以及人流量特别大、存在较高安全风险的公共场所。

德国火车站将实施全面禁酒令降低暴力事件

德国政府及德国铁路公司近日宣布，将于10月15日前分阶段在全国5400座火车站实施禁酒令，借此提升旅客与员工安全、减少暴力事件并改善车站环境。禁酒范围包括车站大厅、月台及通往月台的通道；部分由德国铁路公司管理的车站前广场也将纳入禁酒范围。车站内酒吧、餐厅则不受限制，旅客也可在行李或购物袋中携带未开封酒类。禁酒令实施初期将由站务人员通过宣传、广播及站内公告提醒旅客遵守规定，违反禁酒令者将先被要求离开车站，屡次违规者可能被禁止进入相关车站。若被禁止后仍执意进站，德国铁路公司将对违规者提起刑事诉讼。

（本报记者 吴琼 整理）