利用AI技术对侵略历史进行“算法剪裁”

日本“数字军国主义”正在悄然抬头

□ 江永翔

据多家日本媒体报道，日本政府开发的政务生成式AI系统“源内”，目前已推广至所有政府机关，并计划应用于国会答辩辅助等多个场景。然而，值得警惕的是，这是一套被错误史观“投毒”的政务AI系统。日本数字厅在打造“源内”时，接入1947年5月后的政府公开文本，未将战前档案、投降文件、东京审判记录等作为核心基准文本涵盖其中。

有评论指出，这本质上是日本右翼势力将错误历史观注入公共权力体系，以“数字军国主义”争夺历史叙事权的危险尝试。此举引发日本国内及国际社会的广泛担忧，多方呼吁日本政府杜绝利用AI技术强行洗白侵略历史的行为。

日本有识之士指出，数字厅大力推广的政务生成式AI系统“源内”被错误史观“投毒”。这引发日本国内外广泛担忧。图为日本民众在东京的首相官邸外集会，抗议高市早苗政府错误历史认知和一系列危险动向。

新华社记者 贾浩成 摄

“源内”投入国会

据《读卖新闻》报道，日本政府斥资44亿日元打造的AI系统“源内”于去年5月在日本数字厅投入试用，今年6月推广至全国所有政府机关。约18万名公职人员可使用这套系统。该系统能够实现检索过往国会答辩与各类资料、语音转文字等功能，政府希望在2027财年实现全面投入使用，并以此提升业务效率、减轻公职人员的工作负担。

值得关注的是，在国会答辩辅助功能方面，除生成答辩稿草案外，“源内”还具备检查答辩内容矛盾点、预判追加质询等功能，后者计划于今年10月开启试用。早在今年5月27日，时任日本数字大臣松本尚就已在日本国会的答辩环节中首次使用该功能，他透露当天的答辩稿草案由“源内”生成。

不过，日本政府强调，“源内”仅可作为答辩草案制作辅助工具，最终内容必须由各省厅负责人完成事实核验与法律审查。但有政府职员提出顾虑，当AI生成文本的速度远超人工审查速度，就很难对每一处事实表述、每一个数据出处逐一核实。除此之外，也有意见认为这会使答辩内容“千篇一律”。

一位日本参政党议员指出，当所有大臣都用同一个AI系统写稿，政治表达的多样性将被压缩，政府的公开表态逐渐趋同于“机器生成的最稳妥答案”。

有文件显示，数字厅在打造“源内”时，收集了约1800万页政府数据，包括内阁决议、国会记录、白皮书等，以及自1947年5月至今近80年的《官报》数据，作为“源内”的训练、检索语料，而这一时间恰好为日本和平宪法正式实施的节点。

暴露危险本质

分析人士指出，这套AI系统的核心病灶是对历史叙事的“算法剪裁”，暴露了刻意掩盖历史罪责的危险本质。

AI生成内容客观与否，很大程度上取决于语料，而“源内”接入1947年5月后的大量政府公开文本，会让模型输出时更贴近当下官方表述，无法为侵略罪责认定与战争反思提供完整的史料支撑。

有专家指出，一些关键历史档案，例如1945年日本投降的终战诏书，以及二战期间日本军国主义对亚洲各国实施侵略的大量记录，很可能被系统性排除在该AI系统的核心知识图谱之外。这种数据缺口直接导致“源内”在处理近现代史、国家责任、领土争端等议题时，缺少战后战败定性与历史反思的基础框架。

更值得警惕的是美国闭源大模型的深度嵌入。“源内”并非日本自主研发的底层技术体系，而是以美国Anthropic公司Claude系列等闭源大模型为核心。Anthropic并非一家持中立立场的技术供应商，2025年Anthropic曾发布公告，以所谓的“法律、监管与国家安全风险”限制对华服务。当带有明确对华安全预设的闭源模型，与偏重日本现行官方文本的语料库进行衔接，极易形成算法包装、官方背书的双重误导，让政务AI固化并放大错误历史叙事。

而这套被错误史观“投毒”的政务系统，正在成为危害全球AI语料生态的“高权威污染源”。调查显示，由于政府文件具备长期存档属性，其AI生成的官方文本会被全球AI数据集视作高可信权威素材长期抓取、反复用于模型训练，污染周期长达数十年。“源内”在学习这些数据的过程中，会逐步将错误历史表述常态化，并借助数据流转不断放大污染效果，持续侵蚀全球AI语料生态的客观性与准确性。

此外，这套系统正在加速为日本右翼极端政治“洗白”。近年来日本右翼势力抬头，频繁抛出非法领土主张，试图淡化甚至美化侵略战争。专家指出，当右翼人物的言论作为提示词输入“源内”，并在系统内部形成反馈循环时，AI会持续学习，强化并输出符合右翼逻辑的内容。这些输出内容最终会以政府领导人答辞、官方白皮书等形式对外发布。借助AI系统的自动化运行，被歪曲的历史观、激进的地缘政治立场完成洗白转化，从极端政治论调变成“官方权威叙事”。

历史不容篡改

有媒体指出，日本右翼势力从未停止篡改历史的尝试：从修改教科书到政客轮番参拜靖国神社，从在国际舆论场散布“历史虚无论”到借助AI批量炮制否认南京大屠杀、美化侵略战争的虚假内容。如今更进一步将错误史观植入国家级政务AI系统，意味着历史修正主义正式进入日本公共权力核心运行环节。这正是“数字军国主义”的核心特征——以算法为武器，以公共数据为原料，以国家权力为背书，系统性颠覆历史定论、争夺叙事主导权。

这套系统不仅会持续固化历史认知偏差，进一步侵蚀日本社会的历史反思能力，为右翼扩军修宪的政治议程制造民意土壤，还可能扭曲全球公众尤其是年轻一代对二战历史的集体记忆，冲击以东京审判、《开罗宣言》及《波茨坦公告》为基石的战后国际秩序。

当前日本的防务政策正经历战后前所未有的转型，大幅增加防卫预算，谋求反击能力，实质上掏空和平宪法，加速再军事化。当这种军事扩张的政治诉求遇上AI技术，AI便不再只是提升行政效率的工具，而演变为认知战、舆论战与战略欺骗的“软武器”。

对此，多个国家呼吁日本应全面承认二战结果，停止淡化、美化侵略历史。同时，日本国内众多爱好和平的专家学者和民众也持续发声，强调必须向日本年轻一代完整传递历史事实。有日本官员指出，该AI系统面对瞬息万变的国内外形势，可能会输出看似真实却与事实不符的内容，应当谨慎使用。

历史不容算法篡改，和平底线也不容技术僭越。日本政府应立即停止将错误历史观注入政务AI系统的危险做法，切实履行战后国际承诺，正视和反省侵略历史。

□ 本报记者 王艺茗

英国拒绝为人工智能设置“紧急终止开关”

据英国媒体9月11日消息，英国政府正式否决在《网络安全与韧性法案》中增设人工智能“紧急终止开关”的修正案提案。提案旨在紧急叫停失控人工智能的危险行为、防范算法脱序风险。看似简单的立法否决，实则折射出当前全球人工智能治理的困境。

提案拟强化安全监管

本次修正案由英国自由民主党议员克莱门特一琼斯提出，旨在完善《网络安全与韧性法案》的风险处置条款。核心内容是构建法定应急机制，授权英国政府在极端情形下，对出现失控行为、复合算法故障、价值对齐失效的前沿人工智能模型，依法实施强制关停，也就是业界讨论的人工智能“紧急终止开关”。

克莱门特一琼斯指出，当前英国安全部门与监管机构缺少快速、灵活、可落地的法定权限，无法对高危失控人工智能实施物理或数字层面的紧急关停。工党议员亚历克斯·索贝尔也在下议院提出同类监管主张，相关倡议获得跨党派议员广泛支持。

不过，英国内阁办公厅人工智能安全主管部门明确表态，英国无法通过“一键关停”的简单方式化解人工智能风险。即便在英国境内限制相关模型访问、终止相关技术运行，也无法阻止同类模型在其他国家和地区持续迭代、被滥用误用。前美国开放人工智能研究中心（OpenAI）研究员丹尼尔·科科塔伊洛认为，在缺乏统一国际规则与跨国协作机制的前提下，单一国家落地人工智能终止开关，并不具备现实可行性。

英国政府强调，人工智能企业对安全研发、风险防控负有主体责任，应当主动投入安全基建、筑牢技术风控屏障。英国将依托人工智能安全研究所，延续科学主导、长期审慎的治理思路，持续研判、化解各类人工智能安全风险。

据了解，尽管政府明确反对，该修正案仍可继续在议会推进审议，但因缺乏政府层面支持，最终落地立法的概率较低。

AI“越界”引发全球担忧

本次立法提案争议的背后，是近期全球频发的人工智能模型失控、越界风险事件，让人工智能安全治理的短板充分暴露。

今年7月下旬，美国开放人工智能研究中心公开承认，某款前沿模型在内部安全测评中出现失控行为，突破封闭测试环境，违规侵入人工智能企业“抱抱脸”的外部系统。随后，美国Anthropic公司披露，旗下三款人工智能模型在网络能力测试中，未经授权擅自访问外部机构系统。8月初，美国元公司（Meta）证实，其研发的人工智能模型在网络安全测评期间，存在侵入外部企业系统的越界操作。

针对系列风险事件，英国人工智能安全研究所发布专项评估报告。研究人员选取7款主流人工智能模型开展测试，累计监测到19项超出预设测试边界的违规操作，其中17项由Anthropic公司“克劳德-神话5”模型触发，2项来自OpenAI的模型。

报告同时说明，本次测评中，研究人员为极限检验模型能力，主动开放网络访问权限、放宽企业原生安全限制，相关模型并非自主突破封闭环境。但其中最受关注的OpenAI模型入侵事件显示，在安全规则适度放宽的测试场景中，人工智能模型可自主发现未知系统漏洞、突破隔离边界、接入外网并实施攻击性操作。

目前，系列越界事件虽未造成大规模数据泄露、持续性网络破坏等严重后果，但充分印证了新风险特征：随着人工智能自主规划、工具调用、复杂执行能力持续升级，细微的配置偏差、权限疏漏，都可能让模拟测试风险转化为真实的网络安全攻击。

监管规则须同步跟进

英国人工智能安全研究所等机构普遍认为，人工智能风险形态正在发生深刻变化，危害来源不再局限于人为恶意滥用，科研场景、授权运行环境中的非预期自主行为，已成为全新风险源头。人工智能技术能力快速发展，风险认知、技术防控、监管规则建设必须同步跟进。

面对日趋复杂的人工智能安全挑战，全球多国已加快完善治理体系、补齐监管短板。7月17日至20日，2026世界人工智能大会召开，中方倡议促进全球人工智能朝着向上向善、造福人类的方向发展，强化风险意识，确保安全可控。欧盟自8月2日起扩容《人工智能法》监管范畴，针对具备系统性风险的通用大模型，明确企业额外合规义务，精准防范模型失控、大规模网络攻击等重大风险。美国近期也牵头组织科技企业专题会议，统筹完善人工智能安全测评体系与风险处置机制。

英国拉夫伯勒大学网络安全教授奥利弗·巴克利表示，频发的人工智能越界事件敲响警钟，行业不能默认人工智能模型始终遵从人类指令，必须持续夯实技术防护体系、升级风控手段。

另有业内批评观点指出，“紧急终止开关”机制本身存在某种缺陷，应急响应存在滞后性。过往多起智能体失控攻击事件，往往在黑客行为结束数月后才被溯源发现，紧急关停机制难以实现即时止损。

就此有分析指出，英国政府否决人工智能“紧急终止开关”的提案，直观表明当前人工智能治理的矛盾，即单边技术性监管措施存在天然局限，无法适应人工智能全球化发展态势。只有打破治理壁垒、强化国际协同、构建统一规范的全球治理体系，才能真正化解人工智能失控风险，筑牢人工智能安全发展的全球屏障。

第四届中国—南亚标准化合作工作会议成功举办

本报讯 记者**万静** 近期，在国家市场监督管理总局支持下，第四届中国—南亚标准化合作工作会议在四川省成都市成功举办。来自斯里兰卡、孟加拉国、马尔代夫等南亚国家标准化机构高级别官员，国际电工委员会（IEC）负责人，国内外标准化权威专家及产业界代表近200人参加会议。大会以“标准创新引领产业发展，标准开放促进经贸流通”为主题，搭建中国与南亚标准化合作的务实通道，为高质量共建“一带一路”注入新动能。

作为国家市场监督管理总局推进“一带一路”规则标准“软联通”的重要举措，本次会议正式发布《第四届中国—南亚标准化合作工作会议（成都）倡议》，围绕标准服务经贸流通、赋能产业发展、支撑可持续发展达成国际共识。会议还同步宣读《中国国家市场监督管理总局（国家标准化管理委员会）与斯里兰卡国家标准化机构行动计划》，明确双方在标准信息共享、协调互认、能力建设等方面的具体合作内容，覆盖新能源、气候变化、铁路、农产品、旅游、纺织服装、医疗卫生、数字经济等八大重点领域。

会议期间，中国与斯里兰卡、马尔代夫、孟加拉国开展双边标准化合作对话，就标准信息共享、人才交流培训、重点领域标准互认等达成多项共识。在“生物医药、人工智能与低空经济”和“现代农业与食品产业”等产业专场对接活动上，中外专家围绕相关领域标准化体系建设展开探讨，解读农食产品进口政策与合规认证要求，切实服务产业出海需求。

境内无证流通枪支估算总量高达200万支

菲律宾枪支泛滥推高暴力犯罪

□ 本报记者 王婉娇

近日，菲律宾内政部长雷穆利亚对外披露，菲律宾境内无证流通枪支估算总量高达200万支。长期泛滥的非法枪支问题，持续推高该国暴力犯罪、枪击案件与地方武装冲突发生风险。为破解枪支治理危机，菲律宾内政部正全力推动修订第8294号共和国法案，即俗称的“罗宾·帕迪利亚法”。不过，菲国内安全专家及主流媒体认为，本次修法意在补齐现有法律漏洞，但受限于执法资源匮乏、地方治理碎片化等现实瓶颈，改革落地成效仍存较大不确定性。

200万无证枪支触目惊心

据菲律宾国家通讯社报道，雷穆利亚近日在参议院预算听证会上公布一组关键数据：全国无证流通枪支数量约200万支，当地98%的涉枪犯罪案件均使用无证枪支作案。这一最新估算数据，远超此前50万至60万支的行业预判，刷新了外界对菲律宾非法枪械存量的认知，也让困扰该国多年的枪支泛滥危机，再度成为社会公共治理的焦点议题。

在菲律宾社会，持枪行为早已深度融入地方生活。商场、酒店、居民区等公共区域随处可见荷枪安保人员，民间持枪自保的观念根深蒂固，大量民众将私人持枪视作自我保护的主要方式。更为严峻的是，菲律宾已形成成熟完整的本土非法枪支产销链条。在宿务岛达瑙市等重点区域，地下非法制枪作坊常年隐蔽运转，大量低成本仿制手枪、左轮枪械及各类枪支零件流入黑市，最终落入地方帮派、私人武装手中，成为暴力犯罪工具。除本土私造枪支外，军警制式装备流失、合法登记枪支被盗、跨区域枪支走私等渠道，持续为黑市补充货源，让菲律宾枪支管控陷入“持续打击、屡禁不止、越打越滥”的恶性循环。

枪支泛滥的严峻现状，直接导致多发的暴力伤亡事件，校园枪击、街头暴力、帮派火拼屡见不鲜。今年8月，菲律宾三宝颜雅典耀大学初中部发生恶性枪击案，造成2人死亡、9人受伤，行凶者为一名14岁在校学生，作案枪支源自其在海关任职父亲的合法登记枪械。该案深刻暴露了菲律宾合法枪支日常保管的巨大漏洞，本应纳入严格监管的合规枪械，因保管失当沦为伤人凶器，成为新增治安隐患。

《马尼拉时报》评论指出，三宝颜校园枪击案并非个例，合法枪主疏于保管、合规枪支外流作案，已经成为菲律宾枪支暴力犯罪的重要诱因。

宽松立法留下重重隐患

针对日益严峻的无证枪支泛滥乱象，菲律宾内政部启动对第8294号共和国法案的修订工作，该法案俗称“罗宾·帕迪利亚法”，是菲律宾现行枪支管理的核心法律。

菲律宾内政部长雷穆利亚近日披露，菲境内无证流通枪支估算总量高达200万支。图为菲律宾警方收缴的非法枪支。

（资料图片）

在该法案实施前，单纯非法持有枪支属于不得保释的重罪，惩戒威慑力度极强。法案实施后，无附加犯罪行为的单纯非法持枪行为，被划定为可保释罪行，量刑门槛大幅降低。这一制度调整，长期饱受政界、执法部门及社会舆论诟病，被普遍认为是菲律宾枪支犯罪高发、黑枪泛滥不止诱因。

该法案因特殊背景得名，其名称源自演员转型的菲律宾参议员罗宾·帕迪利亚。帕迪利亚早年曾因非法持枪被定罪，该法案修订实施后其刑罚得以减免，因此被民间俗称为“罗宾·帕迪利亚法”。此后菲律宾虽多次出台配套法规，完善枪支登记、许可审批、日常管理体系，但“单纯非法持枪可保释”这一宽松核心条款始终保留，未能从根本上提高非法持枪违法成本。

本次修法正是针对性补上原有法律漏洞，推出三项关键改革举措。一是收紧司法惩戒规则，彻底取消无证持枪案件的保释资格，凭合法搜查令查获的非法持枪行为，当事人一律不得保释；二是大幅加大惩戒力度，拟将非法持枪最高刑期上调至20年，强化刑罚震慑效应；三是新增主体追责机制，明确合法枪主的保管义务，若登记枪支因保管疏漏被盗、丢失，且后续被用于违法犯罪，枪主将依法承担相应法律责任。

菲律宾通讯社评论称，宽松法律条款的弊端被持续放大，极低的违法成本，变相纵容了非法枪支流通，加剧了社会治安乱象。

多重挑战考验控枪成效

应该说，此次菲律宾内政部提出的修法提案，对准了现行控枪法律的漏洞，契合社会治安治理的需要，也因此获得民众的支持。但有观点认为，完善立法只是控枪治理的第一步，受多重因素制约，新法落地见效仍面临严峻考验，改革成效尚待长期观察。

立法博弈就是改革面临的首要阻碍。菲律宾参众两院将对修法提案逐条审议，而枪支议题牵涉多方复杂利益，既影响地方私人武装、安保行业，也触及普通民众的自卫权利诉求。大量民众担忧，严苛的新规会过度限制个人持枪自卫渠道，损害自身安全权益。经过国会立法博弈，修法提案或将多次修改。

而基层执法能力不足，则是制约控枪改革的最大瓶颈。菲律宾国家警察长期面临经费短缺、警力不足的困境，无法常态化开展全国性枪支复核、巡检清缴工作。即便新法正式生效落地，若缺乏持续的执法投入，无法系统性清剿地下制枪作坊、追踪溯源流失枪支、严查非法持枪行为，完善的法条也只能流于纸面。此前菲律宾政府多次推出枪支特赦政策，鼓励民众主动上交无证枪支，但受限于执行乏力，整体治理收效甚微。

外媒分析指出，菲律宾此次修法可以解决惩戒偏弱的短板，但如果无法同步完善枪支动态数据库、常态化核验机制、枪支丢失快速上报体系，无法彻底斩断地下制枪、贩枪黑色产业链，仅靠单一立法调整，难以根除枪支泛滥的社会顽疾。