3月全球AI观察

智能体加速落地 技术滥用须警惕

新华社记者

人工智能（AI）浪潮正深刻重塑全球科技与产业格局。3月，以“开放之爪”为代表的AI智能体加速落地应用，使AI从对话交互迈向自主执行任务，并推动全球算力与词元需求迎来爆发式增长。

与此同时，多国密集出台战略规划、加大产业与算力布局。随着AI深度赋能千行百业，技术滥用、伦理失范与安全风险也日益凸显，如何在抢抓发展机遇的同时筑牢安全防线，成为全球AI产业发展面临的共同课题。

从对话到执行 智能体重塑AI应用格局

继聊天机器人之后，能“自主干活”的智能体正快速进入人们生活。今年以来，一款名为“开放之爪”（OpenClaw）的开源智能体在全球科技圈迅速走红。数据显示，它在开源平台GitHub上线仅两个多月便获得超过30万颗“星标”，显示出开发者社区对其高度关注。在中国，在设备上部署该工具被称为“养龙虾”，“养虾”热潮以超乎想象的速度火爆出圈。

支持智能体的AI模型、技术方案等也不断演进。美国开放人工智能研究中心（OpenAI）3月5日宣布推出最新升级版大模型GPT-5.4，称该模型是其首个能直接操作计算机的通用模型，不仅擅长编写代码，还能根据屏幕截图发出鼠标和键盘操作指令，使智能体能够操作计算机，并在不同应用程序之间执行复杂的工作流程。

英伟达公司创始人兼首席执行官黄仁勋3月16日在该公司年度GTC大会（GPU技术大会）上发布名为“NemoClaw”的软件栈，用于支持“开放之爪”智能体的开发和部署。

智能体等AI应用的兴起，使得“词元”（token）概念越来越多地进入公众视野。作为AI模型处理和生成信息的基本单位，词元逐渐成为衡量智能服务成本与价值的重要标尺。全球主要AI模型聚合平台“开放路由器”数据显示，用户通过该平台调用AI模型的词元总量从今年年初的一周约6万亿大幅上涨到截至3月22日一周的20.4万亿。

多国加码AI战略 算力与产业布局提速

为抓住AI发展机遇，多国政府和企业纷纷加大政策扶持与产业投资，在智能经济布局、国际合作与规则完善、算力基建扩容等方面持续发力。

中国在今年的政府工作报告中首次提出“打造智能经济新形态”，包括深化拓展“人工智能+”，促进新一代智能终端和智能体加快推广，推动重点行业领域人工智能商业化规模化应用，培育智能原生新业态新模式等。

韩国政府3月2日表示，韩国将与新加坡加强在AI领域的合作，计划在新加坡设立3亿美元的全球投资基金，吸引对初创企业和AI领域的投资。

欧盟理事会3月13日就一项简化部分AI监管规则的提案达成一致，主要涉及精简欧盟AI监管框架，减轻企业合规负担等内容。

德国政府3月17日公布的一项数据中心扩容战略规划显示，到2030年，德国通用数据中心的算力将在2025年基础上至少翻一番，其中专门用于AI的算力将至少增至2025年的4倍。

美国知名企业家埃隆·马斯克3月21日宣布，一个名为Terafab的大型芯片制造项目将落户得克萨斯州奥斯汀。根据他的构想，该项目将建设一座集设计、制造、封装和测试于一体的先进晶圆厂，目标支撑每年最高达1太瓦级算力需求，并服务自动驾驶出租车、人形机器人及太空数据中心等应用场景。

3月24日，OpenAI宣布将结束“天空”（Sora）视频生成业务。此举被外界视为OpenAI战略转型的重要举措。观察人士认为，关停Sora不仅是该公司在业务方面的一次“断舍离”，也是AI产业当前竞争转向的缩影。

赋能千行百业 伦理与安全风险不容忽视

3月初在西班牙巴塞罗那举行的2026年世界移动通信大会上，机器人手机等多款AI创新产品成为焦点，AI推动产业变革的趋势也愈加明显，移动通信正从单纯提供连接的网络基础设施，向融合云计算、数据平台与各行业数字化应用的更广泛数字生态加速演进。

不过，在AI持续赋能各行各业同时，新技术带来的问题也逐渐显露。美国《科学》杂志近期发表的一项研究显示，当人类用户就人际困境等问题向AI模型寻求建议时，AI常表现得过度迎合或谄媚；对于一些有害甚至违法的提问，AI常常肯定用户的立场。

马斯克旗下企业xAI公司开发的聊天机器人“格罗克”此前因被滥用于生成基于真实人物的虚假色情内容，被多国监管机构调查。近期，“格罗克”又因发布冒犯内容和歪曲事实引发舆论关注和批评。

美国和以色列对伊朗发动军事打击以来，大量运用AI技术实施目标识别与打击，造成严重平民伤亡和民用设施损毁。AI技术在军事领域滥用的伦理与安全风险急剧凸显，推动形成具有广泛共识的全球AI治理体系已刻不容缓。

中国国防部新闻发言人蒋斌3月11日答记者问时表示，中国愿与世界各国一道，推进以联合国为核心的人工智能多边治理进程，加强风险预防和管控，确保人工智能始终朝着有利于人类文明进步的方向发展。

（新华社北京电）

3月18日在安徽省合肥市拍摄的“腾讯小龙虾一站式服务日”活动现场。
新华社记者 周 牧 摄

3月13日，在北京市一家商场内，对“龙虾”智能体感兴趣的市民在美团组织的安装“龙虾”活动上进行咨询。 新华社发

AI智能体“龙虾”为何引发广泛警惕

新华社记者 冯玉婧

今年年初以来，一款俗称“龙虾”的人工智能（AI）智能体工具“开放之爪”（OpenClaw）凭借其自主执行复杂任务、可扩展技能包等强大能力，在开源社区迅速崛起。但爆火之后，“开放之爪”接连被曝出存在多重安全隐患。

目前，多国监管机构和科技企业已陆续发布针对“开放之爪”的使用指南和规范。4月1日，中国国家知识产权局发布风险提示说，“开放之爪”等智能体工具被曝光默认安全配置脆弱，易引发严重安全风险。与此同时，使用此类智能体撰写专利申请文件，也可能诱发多重风险。

安全漏洞频发

“开放之爪”由奥地利软件工程师彼得·施泰因贝格尔开发，是一款开源AI智能体软件。该智能体采用层级化架构，将社交即时通讯软件与自动化智能体深度耦合，同时借助插件系统扩展各种工具能力。这种分层架构虽赋予了“开放之爪”灵活性与可扩展性，但也带来了多维度的安全风险。

1月下旬，开源平台GitHub上发布的一项安全审计报告显示，“开放之爪”存在512项安全漏洞，其中有8项被归类为“严重”，涵盖了身份验证、机密管理等领域。

2月下旬，国际网络安全机构“绿洲安全”研究人员发布报告说，“开放之爪”核心系统中存在一个名为“ClawJacked”的重大安全漏洞，攻击者可能通过恶意网页接管该智能体，从而获取设备权限和访问系统数据。“开放之爪”团队将漏洞定级为“高度危险”，并在24小时内发布了修复版本。

3月30日，中国360数字安全集团在官方微信公众号上发文说，在“开放之爪”平台中发现一处高危漏洞，影响范围覆盖全球50多个国家和地区。

谨慎使用（漫画） 新华社发 冯德光 作

广泛的攻击风险

美国微软公司安全团队发布的风险报告显示，使用“开放之爪”可能面临两类攻击风险：恶意技能插件和间接提示词注入。

“开放之爪”的执行能力依赖于社区平台提供的技能插件。绿盟科技公司近期发布的安全报告指出，如果缺乏严格的代码审计和签名校验，攻击者可通过发布包含恶意提示词和代码的恶意技能插件实现“代码投毒”。用户可能只因一次点击就加载了此类插件，攻击者可在受害者系统中获得持久驻留能力。而攻击者上传自定义技能插件的门槛非常低，只需要注册一个非实名的GitHub账号即可。

据美国派拓网络公司2月发布的数据，研究人员已在相关平台上发现超过800个针对“开放之爪”的恶意技能插件。

提示词注入是一种针对大语言模型的攻击技术，分为直接注入（攻击者直接输入恶意指令）和间接注入（通过网页、文档等外部数据源实现攻击）两种方式。

美国“众志”网络安全服务公司近期在官网发文说，提示词注入的首要威胁是敏感数据泄露，考虑到“开放之爪”对敏感文件与系统的高访问权限，这一风险尤为严重。间接注入则会进一步放大风险，因为攻击者无需直接与“开放之爪”交互，只需污染其读取的数据，恶意指令即可悄悄进入软件决策流程。

多国机构及企业发布使用规范

对于“开放之爪”是否适合在企业中部署应用，“众志”公司的文章指出，若员工在企业设备上部署“开放之爪”或将其接入企业系统，且配置不当、缺乏安全保护，它就可能成为系统“后门”，执行攻击者的指令。

业内人士建议，个人或企业用户不要在常规办公与涉密设备上运行“开放之爪”，如需部署须采取权限治理、沙箱机制、持续监控与全周期安全防护等严格管控措施。

据媒体报道，出于风险管控的考虑，美国元宇宙平台公司、韩国多音通讯公司等多国科技企业已禁止员工在办公设备上使用“开放之爪”。与此同时，多国监管机构也发布了关于使用“开放之爪”的安全指南。

荷兰数据保护局2月发布公报，建议用户和组织不要在存有敏感或机密数据（如访问码、财务行政资料、员工数据、私人文档或身份证明文件）的系统上使用“开放之爪”及类似AI智能体；建议谨慎对待外部插件，实施严格的访问控制，在存在泄露风险时及时更新登录信息。该监管机构还呼吁将“开放之爪”等AI智能体纳入欧盟《人工智能法》的管辖范围。

3月22日，中国国家互联网应急中心等发布了“开放之爪”安全使用实践指南。此前，工业和信息化部网络安全威胁和漏洞信息共享平台组织相关机构研提了“六要六不要”建议，以防范“开放之爪”开源智能体安全风险。

（新华社北京电）

从“试水”走向“深耕”，AI技术深度赋能银行业

新华社记者 吴 雨

在AI赋能全社会的浪潮中，银行业是落地较快、应用较广的领域之一。从信贷审批到风控决策，从智能客服到产品营销……人工智能正深度融入各类业务场景，助力降本增效，成为推动金融创新发展的重要引擎。

银行业AI应用多点开花

近期，农业银行推出一款自主研发、绿色金融专属的AI智能体，引起业内关注。

“我们希望帮助客户经理自动加工绿色项目数据、交叉验证多维信息和智能生成尽调报告，让办贷流程更便捷、更高效、更安全。”农业银行董事长谷澍近日的公开表态，道出银行布局AI的务实初衷。

近两年，从国有大行到股份制银行，再到区域性城商行，越来越多金融机构亮出AI布局规划，AI赋能金融已从早期试水探索，转向深耕细作，逐渐成为衡量银行科技实力与发展潜力的关键赛场。

工商银行将“数字工行”战略升级为“数智工行”，在30多个业务领域落地500多个AI应用场景；建设银行AI应用覆盖近400个场景，借助AI赋能授信审批全流程；交通银行在跨境贸易、财富管理、零售数字化经营等领域深化AI应用，2025年智算规模同比增长超过50%……

AI应用场景不断扩容的背后，是银行业在底层技术、自研模型、数据治理等多个维度的持续发力。

打造“工银智涌”技术体系、构建大模型弹性算力池、集成十多款业界主流模型……工行在金融转型变革的前沿持续发力，全力提升数智化发展水平。

“数智化不是选择题，而是必答题，是我们抢占先机、掌握主动的战略选择。”工商银行副行长赵桂德说。

“银行业正全面拥抱AI，助推AI从辅助工具升级为银行核心生产力。”北京师范大学经济与工商管理学院副院长江婕表示，借助高质量金融数据与先进算法模型，AI将从单一技术支撑逐步转变为懂业务、懂风险的智能力量。

科技赋能助力银行降本增效

技术的价值，最终要靠业务增长与效率提升来印证。从多家银行2025年业绩报告公布的一组数据，可以看出AI带来的切实改变。

在工行，金融市场智能询价助手助力交易笔数同比提升50%，个人客户经理营销助手带动重点产品成交额增加千亿元；在建行，借助AI赋能授信审批全流程，受理量两位数增长、平均处理时间下降30%以上；在邮储银行，在智能推荐模型的助力下已实现个人授信超60亿元，财富类产品销售超14亿元。

“人工智能的应用带来业务效率提升，贷款审批速度和账户开户速度大大提高，人工操作错误减少，有助于提升服务效能和业务规模。”上海金融与发展实验室副主任董希淼表示，在人工智能的加持下，银行风险管理也在升级，逐步从“规则防御”向“智能预判”转变。

2025年8月，国务院印发的《关于深入实施“人工智能+”行动的意见》对外发布，部署在金融等领域推动新一代智能终端、智能体等广泛应用。

在金融业数智化转型按下加速键的同时，AI应用中的潜在风险不容忽视。

业内专家提醒，金融业务不可过度依赖数据模型，仍要坚持见人见事实地调查，对事实进行多维度交叉验证，确保真主体、真经营、真用途。要警惕“数据孤岛”带来的模型偏差，以及人工智能“幻觉”等新型技术风险。

工商银行行长刘珺在近期举行的业绩发布会上表示，全面且系统的验证是推广AI相关技术的前提条件，必须把安全可靠的应用推向客户、服务于实体经济。

面向未来深化AI布局

当前，越来越多银行将AI能力建设提升至更高战略位置，聚焦全场景应用、深度业务赋能、安全可控运行等方面，谋划更为长远的布局，以科技硬实力持续引领金融创新。

“科技自身能力不强，服务业务就会成为一句空话。”交通银行副行长钱斌表示，今后将在需求分析、产品研发、系统测试、生产运维、网络安全等全链条深化AI应用，推动科技服务向智慧化、精益化升级，以科技硬实力精准赋能业务发展。

立足长远，邮储银行透露，已启动“十五五”IT规划编制，明确以人工智能为核心驱动，打造营销、客服、财务分析、综合办公等多类智能体，深化数字生态建设，将技术能力转化为可感知、有温度的金融服务。

“力争用两年的时间，让AI渗透到我行每一个重要决策和业务环节。”中信银行副行长谢志斌给出明确时间表，表示力争到“十五五”末实现90%以上核心业务流程AI重塑，完成从AI辅助到AI原生的跨越。

2025年12月，金融监管总局发布《银行业保险业数字金融高质量发展实施方案》，明确提出加快发展“人工智能+金融”。金融监管总局表示，将指导金融机构积极稳妥推进数字化转型，赋能金融服务提质增效，同时统筹好发展与安全。

江婕认为，人工智能加速银行业效率革命与商业模式重构，正系统性重塑银行的成本结构、决策效率与服务能力。持续加大AI投入、深化场景落地的银行，将显著提升成本管控与科学决策水平，构筑更强劲的核心竞争力。 （新华社北京4月10日电）

中共运城市委主管主办 社址：运城市河东东街3969号 邮政编码：044000 电话：总编室 2233336 发行部 2233450 广告部 2233411 广告许可证：1427014000001 全年报价：528元 零售每份2元 运城市运报印务有限公司印刷 地址：运城市老东街116号