当“蒸馏”受阻，中国大模型企业“出海”路在何方

▶ 本报记者 孙庆阳

近日，美国国家安全局、网络安全与基础设施安全局、联邦调查局联合发布公告，指责中国人工智能企业对美开展“工业规模蒸馏”，获取美前沿模型能力，并向美企提出相关防御措施建议。而中国商务部用8个字回应：于事无凭、于法无据。

据了解，所谓的“蒸馏”如同师傅教徒弟，即大模型这位“老师”把“知识”浓缩后传授给“学生”小模型，这也是人工智能领域各个模型间互相学习的通行做法，本质是中性技术手段，包括美企在内的全球模型企业都在用。

从芯片到模型，再到访问权限，规则冲突升级。中国大模型企业想要顺利“出海”，该如何通过合规这道关？

一场模型的规则之争

商务部国际贸易经济合作研究院助理研究员刘立菲介绍说，美国的抓手经历了3次升级：2020年针对华为等中国实体实施管制；2022年10月开始大范围限制对华出售高端芯片与半导体设备；自2023年起，又限制境外生产的芯片和有关国家利用第三国转移或获取受限物项。

“芯片只是训练模型的‘工具’。”在刘立菲看来，大模型的真正家底是海量模型参数，即“模型权重”。此外，还有训练方法、数据推理能力等看不见的无形资产。

“2026年，美国相关部门将矛头直接对准了模型本身。”刘立菲说，4月份，美国众议院外交事务委员会通过《阻止美国人工智能模型窃取法案》，主张用出口管制、实体清单甚至经济制裁对付“模型盗窃”；6月份，美国政府对商用大模型下达“按国籍封锁”的紧急指令，导致全球顶级大模型一度停服下架；9月份，美国国家安全局、网络安全与基础设施安全局、联邦调查局又给6家中国企业扣上“工业规模蒸馏”的帽子。

刘立菲认为，美国出口管制的抓手从芯片延伸到模型，再延伸到访问权限，已从控制训练用硬件，逐步转向管制训练形成的模型能力，从有形的芯片转向无形的访问权。

“模型‘蒸馏’本质上涉及产品使用方式，更多属于知识产权和合同方面的问题，一旦纳入出口管制可能会造成国际法层面的数量限制、国家间贸易对抗等问题，将私人主体的纠纷转化为国家间的贸易壁垒。这一步一旦跨出，正常的科研交流、技术评测都可能被怀疑。”刘立菲说。

缺乏多数国家认可的约束性条约

对于美国有关部门试图以美国法律监管中国企业的行为，上海交通大学法学院教授沈伟表示，当前全球人工智能治理相当混乱，所谓“国际规则”大多是软法及白皮书、政府倡议之类的文件，缺乏一份被多数国家认可且具备约束力的条约。“在国际规则缺位的情况下，某些国家的国内法会向其它国家‘扩张’，哪个国家的算法技术、人工智能技术领先，其规则的影响力就越强。”

沈伟举例说，目前各国都有外资审查制度，而美国还有芯片法案、出口管制等环环相扣，这些规定对其他国家的外溢效果非常明显。

对外经济贸易大学法学院教授熊光清介绍说，当前欧盟也在推进与美国类似的动作，如欧盟《人工智能法（2024）》已于2024年8月出台。该法案的适用范围较为广泛，即便相关企业是欧盟范围以外人工智能技术的供应商，但在欧盟范围内使用其人工智能技术的产品或提供服务，也需要适用法案的相关规定。再配上市场准入、高额罚款等，欧盟标准事实上已成为全球相关企业必须遵守的“准入门槛”。

“很多中国企业的大模型产品只要有欧盟用户，就需要符合欧盟制定的规则，欧盟标准事实上已经成为全球性标准。”熊光清坦言，虽然“输出地”标准在法理上存在争议，然而真正值得警惕的是欧盟《人工智能法（2024）》第二条的输出地标准。该规定的核心问题在于其效果连接点被过度稀释，最终使作为有限域外管辖依据的效果原则转化为潜在的全球管辖依据。

国内企业如何应对

对于国外设置的相关门槛，中国大模型企业又该如何接招？对此，专家给出两套办法。

第一，以柔克刚。刘立菲认为，应对分两个层面：一是规则层面，在国际上积极发声，联合更多伙伴争取治理空间，达成人工智能作为公共基础设施的国际共识，同时在外贸方面推动人工智能相关产品与服务出海，拓展国际市场；二是根本层面，要依靠国内企业在政策空间内加快技术突破，摆脱技术限制，真正形成竞争力。

“美国自身也在不断试探管制边界，从管制算力，到管制模型权重，再到管制远程访问，管制手段不断升级，但实际执行中面临很多技术规避手段，完全管住非常困难。”刘立菲说。

第二，在国际规则上“抢话语权”。熊光清认为，中国未来在为人工智能综合立法时，应避免复制无限扩张的域外管辖模式。“真正具有国际法正当性的人工智能规则，应当是建立在实质联系、比例约束、相互承认的基础之上的规则。”

熊光清表示，当前中美在大模型、底层技术上的差距仍然不小，美国的严格管制从侧面反映了这种差距；而自我封闭带来的负面影响可能比美国管制的影响更大。换言之，“出海”既是考题，也是练内功的机会。



近日，打通智能分拣、无人接驳、机器人上门配送的全链路智慧校园物流示范场景正式落地哈尔滨工业大学（深圳）。通过与室内外一体化配送机器人协同作业，可实现物流全校区域覆盖。

图为在哈尔滨工业大学（深圳）智能快递服务中心，分拣机械臂分拣快件。

新华社记者 毛思倩/摄

智能体正在重塑算力基础设施

本报讯（记者 张伟）近日，国际数据公司（IDC）与浪潮信息联合发布的业内首份聚焦智能体基础设施的研究报告《2026年中国人工智能计算力发展评估报告》（以下简称《报告》）指出，智能体（Agent）正在重塑算力基础设施，以乘数效应推动Token（词元）消耗激增，带动中国智能算力规模斜率大幅度上调，算力跃升为与水利、电网、铁路等同等级的国家战略基础设施。

IDC预测，全球活跃智能体数量将从2026年的7940万个增长至2030年的22.16亿个，复合增长率达到129.8%。随着智能体任务从简单的信息检索、内容生成，向需要多轮推理、工具调用和跨系统协作的复杂任务演进，Token消耗被指数级放大。同期全球Token消耗量复合增长率达4822.6%，约为活跃智能体数量增速的37倍。中国增长同样显著，2026-2030年中国活跃智能体数量复合增长率达151.2%，同期中国Token消耗量的复合增长率高达3499.3%。

智能体数量与Token消耗的快速增长，正带动算力需求持续攀升。2026年中国智能算力规模预计达到2576.5EFLOPS（每秒可执行一百亿亿次浮点运算），同比增长87.9%，到2030年将进一步达到10524.3EFLOPS；2025-2030年智能算力规模复合增长率由上一版本预测的46.2%提升至50.3%。同时，中国人工智能（AI）算力市场规模将在2026年达到515亿美元，同比增长37%，并在2030年达到1501亿美元，2025-2030年复合增长率达31.8%。

具体来看，《报告》围绕智能体规模化带来的新一轮算力变革，从算力供需、Agent基础设施重构、行业AI应用成熟度及区域AI算力发展格局等多角度展开研究，分析了智能体如何推动人工智能在模型能力持续演进的基础上，加速走向规模化应用，并进一步重塑中国人工智能算力产业的发展路径。

《报告》指出，在2026年中国人工智能城市排行榜中，北京、杭州、深圳位列前三。其中，北京凭借完整的算力供给产业链、实力雄厚的模型与人工智能应用企业稳居首位；杭州依托AI创新企业集群和互联网产业基础，以及在大模型、具身智能等领域的持续发展，保持第二位；深圳作为AI服务器与算力芯片制造的重要基地，凭借深厚的算力供给产业基础，把握智能体爆发机遇，排名由第四升至前三。上海、广州、成都、南京、武汉、济南、苏州进入前十。

《报告》首次从智能体当前应用成熟度和长期生产力影响两个维度评估行业智能体发展，并指出，代码开发（Coding）和办公任务（Work）凭借更高的数字化程度、更清晰的任务边界和更强的结果可验证性，已率先进入规模化应用阶段。传统行业受业务复杂度与安全约束影响，多数仍处于探索试点阶段。制造、医疗健康、能源/电力等行业真正进入核心业务并形成“执行—反馈—优化”闭环的比例仍为0%；运营商虽然已有70.0%接入部分业务流程，但进入核心业务闭环的比例仍只有3.3%。

此外，《报告》预计，全球AI算力需求满足率将由2024年的79%下降至2027年的71%，2030年仅回升至77%，供需矛盾将长期存在。

单纯扩大算力规模已难以应对持续增长的智能需求。《报告》指出，智能体时代，AI基础设施正从规模扩张转向系统效率竞争。通过计算、存储、网络、数据中心、终端及模型、调度和Token运营的全系统协同，竞争核心从“拥有多少算力”进一步转向“能够多高效地持续产出智能”。