

汉坤法律评述

2026 年 9 月 14 日

北京 | 上海 | 深圳 | 杭州 | 武汉 | 海口 | 香港 | 新加坡 | 纽约 | 硅谷 | 伦敦

AI 算力平台责任共担模型：边界、场景与合同安排

作者：陈文昊 | 原舒仪 | 李诗

人工智能的发展离不开算力。AI 算力平台通过整合 GPU、NPU 等计算资源以及网络、存储和调度能力，为人工智能模型的训练、微调、部署和运行提供基础设施和支撑服务。随着 AI 加速进入企业生产经营，算力服务的客户正从专业 AI 企业、科研机构逐步扩展至更多行业企业，企业对算力的需求也从模型训练进一步延伸至推理和实际业务应用。Gartner 预计，2026 年全球 AI 优化型基础设施即服务（AI-optimized IaaS）支出将达到约 423 亿美元，同比增长 96.4%；PwC 则预计，2026 年全球数据中心年度资本开支约为 8,000 亿美元¹²。无论从企业应用还是资本投入来看，AI 算力及其基础设施都正在成为 AI 产业链中越来越重要的一环。

今天的 AI 算力平台也已经很难简单理解为“GPU 出租方”。有的平台主要提供服务器和计算资源，有的平台进一步提供训练环境、资源调度、模型部署和推理服务，还有一些平台已经向企业知识库、检索增强生成（Retrieval-Augmented Generation, RAG）、智能体（Agent）等 AI 开发和应用支撑功能延伸。平台提供的功能越多，对客户训练和应用运行的介入越深，相应的责任边界也越需要重新审视：训练任务中断时，需要判断平台究竟只承诺了算力资源，还是同时承担训练状态保存和恢复义务；企业资料被问答系统不当披露时，需要追踪源文件权限、检索权限和模型调用链条中哪个环节出现问题；智能体错误修改业务记录甚至发起交易时，还要区分业务权限由谁授予、人工确认机制由谁提供、最终执行系统由谁控制。

随着算力交易规模扩大、平台服务链条拉长，这些问题很可能逐渐从产品和合同设计问题转化为实际的诉讼、仲裁和监管问题，因此，在相关争议大量发生之前讨论平台与客户之间应当如何划分责任，无论对于 AI 算力平台运营者、投资人还是平台客户来说，都具有重要的现实意义。

传统云服务中的责任共担模型（Shared Responsibility Model）为此提供了一个有价值的起点，即一项服务可能由平台、客户以及第三方共同完成，各方分别对自己承担和控制的部分履行相应义务。但技术上的工

¹ Gartner, *Gartner Forecasts Worldwide AI-Optimized IaaS Spending to Grow 96% Through 2026*, 10 August 2026. Gartner 预计，2026 年全球 AI 优化型 IaaS 支出为 422.76 亿美元，同比增长 96.4%，并指出大型语言模型训练需求及 AI 在企业应用和工作流中的快速投入运行是相关增长的推动因素。

<https://www.gartner.com/en/newsroom/press-releases/2026-08-10-gartner-forecasts-worldwide-artificial-intelligence-optimized-iaas-spending-to-grow-96-percent-in-2026>。

² PwC, *Global investment in AI infrastructure to hit US\$31.6 trillion through 2050*, 2 September 2026. PwC 在其 Global Data Centre Outlook 相关发布中预计，2026 年全球数据中心年度资本开支约为 8000 亿美元，并预计到 2050 年将增长至每年约 1.8 万亿美元。

<https://www.pwc.com/gx/en/news-room/press-releases/2026/global-investment-in-ai-infrastructure.html>。

作分工、法律规定的主体义务与损害发生后的责任承担并不完全重合，真正的责任判断仍然要回到具体购买了什么服务、谁掌握关键控制能力、合同作出了什么承诺、各方如何实际履行，以及相关行为与损害之间存在何种关系。最高人民法院近期发布的《关于依法审理涉人工智能纠纷案件的意见》也体现了类似思路，强调结合具体应用场景、风险性质和控制能力等因素，对人工智能开发者、提供者、使用者等主体的责任作出合理划分³。本文沿着这一思路，从具体服务、运行与退出以及应用功能扩展三个层面讨论 AI 算力平台与客户之间的责任分配。本文所称“客户”，主要指直接购买或者使用 AI 算力平台服务的企业或者其他组织；使用客户基于平台开发的应用的个人或者组织，则称为“终端用户”。

一、从实际服务出发确定责任分工

（一）按功能明确分工，并识别实际服务主体

“AI 算力平台”并不是单一的服务形态，设施托管、GPU 租赁、托管训练、模型部署和模型接口服务，对应的平台行为和控制能力存在明显差异。在基础资源服务中，平台通常负责其提供和维护范围内的机房、网络、服务器、计算资源和底层运行环境，客户管理自己的业务应用、账号和数据；随着托管程度提高，平台还可能承担训练环境维护、任务调度、模型部署甚至推理服务。传统的 IaaS、PaaS 和 SaaS 分类有助于理解这些层次，但真正影响责任判断的，不是产品被归入哪一个抽象类别，而是具体功能由谁提供、谁负责配置、谁能够控制，以及谁对其正常运行作出了承诺。

因此，与其在合同中笼统写一句“平台负责基础设施、客户负责应用”，更有意义的是把责任拆到具体功能。对于训练任务，至少应当说明运行环境、算力调度、检查点生成和保存、持久存储、恢复操作分别由谁负责，哪些能力需要客户自行配置，哪些属于平台默认提供的服务，以及双方如何确认相关功能已经生效。客户即使自行管理服务器，如果同时购买了平台提供的训练调度、检查点保存和恢复能力，其服务器管理权限也不会使平台已经承诺的恢复义务当然消失；反过来，如果平台只提供 GPU 实例，并没有承诺管理客户训练框架和恢复逻辑，也不能因为任务运行在平台资源上，就把应用层的全部问题归于平台。

除功能分工外，还要识别实际服务链条。签约平台、GPU 资源供应商、机房运营者、云服务商和模型接口提供方可能并非同一主体，平台以自己的名义向客户承诺交付某项服务的，上游供应商发生故障并不当然免除其对客户的合同责任；如果某一主体仅提供交易撮合或者资源调度，则仍需结合合同结构和实际履行方式判断其责任。多方供给还可能改变数据流向，例如所谓“跨地域调度”究竟只是调整计算任务，还是会同时转移训练数据、模型权重、检查点和日志，或者开放异地人员远程访问，法律后果可能不同。涉及跨境处理时，应进一步核实实际接收方、处理地点及数据对象，而不能只看控制台显示的资源区域。

（二）从训练恢复场景看服务承诺应具体到什么程度

AI 训练中的恢复机制能够直观说明，提供 GPU 资源、提供检查点保存能力，与承诺训练任务能够恢复，是不同层次的服务。检查点（checkpoint）可以保存训练过程中的特定状态，使任务中断后从较

³ 最高人民法院《关于依法审理涉人工智能纠纷案件的意见》（法发〔2026〕10号），2026年9月7日发布。《意见》提出，应根据人工智能在不同应用场景下可能造成的损害以及风险的性质和大小依法准确认定法律责任，并结合不同模型的风险外溢与控制能力，对开发者、提供者、使用者等主体的法律责任作出合理划分；在判断行为人有无过错及过错程度时，还应综合考量具体应用场景、自主化程度、技术与信息透明度、潜在风险及影响范围、相关主体采取的预防措施及技术上的可能性等因素。

<https://www.court.gov.cn/zixun/xiangqing/511101.html>。

近状态继续，但其是否真正可用于恢复，还取决于保存内容、存储位置、保存频率以及恢复时能否正确加载。

Amazon SageMaker AI 的官方文档对此提供了一个例子：训练过程中生成的 checkpoint 可以保存在训练容器本地目录并同步至 Amazon S3，新训练任务启动时，S3 中的既有 checkpoint 又可以写回训练容器；对于自定义训练容器、训练脚本或者部分未直接支持 checkpoint 的框架，开发者还需要在训练脚本中设置相应的保存和加载逻辑，保存频率也可能需要通过训练脚本控制⁴。这一例子并不是要把 AWS 的产品机制作为所有平台的交付标准，而是说明“训练恢复”可能同时包含平台能力和客户配置，训练未能恢复时不能只看结果，而要判断哪个环节没有完成约定工作。

责任判断也应区分模型效果和服务履行。如果平台只承诺特定规格和可用性的计算资源，并未承诺模型达到某一准确率或者业务指标，那么模型效果不理想通常不能单独证明平台违约；但平台已经明确承诺训练进度保存或者 checkpoint 持久化服务，却因其负责的存储、同步或者调度功能失效导致训练状态丢失，则应进一步审查具体服务是否履行。模型训练结果具有不确定性，并不意味着算力、存储和恢复功能也失去客观的交付标准。

GPU 重新上线也不等于训练任务能够无损继续。短暂故障如果造成此前数小时甚至数日训练状态丢失，其实际影响可能明显高于资源不可用时间本身；如果可以从较新的 checkpoint 继续训练，影响则相对有限。因此，重要训练项目除约定资源可用性外，还可以明确 checkpoint 由谁生成、保存频率、存储位置、保留周期、恢复条件及技术支持范围，并在上线前用代表性任务进行一次中断恢复测试。服务水平协议（Service Level Agreement, SLA）中的费用补偿也应与具体承诺相衔接，资源不可用、训练状态丢失、checkpoint 损坏和数据泄露等情形未必适合适用同一种赔偿规则；具体争议中，我国《民法典》有关格式条款、免责和限责条款、可预见损失及减损义务等规则，也可能影响最终责任范围。

（三）把基础安全保障与增值服务区分清楚

平台可以根据客户需求设置不同层级的备份、安全监测和恢复产品，但商业上的服务分层不会改变法定义务或者已经作出的合同承诺。客户没有购买高级备份或者增强监测服务，并不意味着平台无需履行依法承担的基础安全保障，也不能据此否定已经包含在基础产品中的安全功能；反过来，客户购买了平台安全产品，也不意味着自身对业务账号、权限配置和数据使用的管理责任全部转移给平台。

工业和信息化部办公厅《关于加强互联网数据中心客户数据安全保护的通知》已经体现出按照实际服务划分责任的思路，要求 IDC 业务经营者根据合作模式和内容，在合同中明确各方数据安全保护责任；对于人工智能场景，该通知还提出，涉及训练数据集管理功能的，应提供保障客户自有训练数据集安全的能力，涉及算力调度和算力服务的，则应做好调度策略安全管理，并配备针对算力异常使用的监测预警和应急处置能力⁵。《网络安全技术 人工智能计算平台安全框架》（GB/T 45958-2025）也为人工智能计算平台安全建设提供了技术框架⁶。前述通知有其特定适用主体和业务范围，推荐性国家标准在

⁴ Amazon Web Services, *Checkpoints in Amazon SageMaker AI*。AWS 官方文档说明，SageMaker AI 可以将训练容器本地目录中的 checkpoint 同步至 Amazon S3，并在新的训练任务开始时将既有 checkpoint 写回训练容器；对于自定义训练容器、训练脚本或者其他相关框架，开发者可能需要自行配置 checkpoint 的保存和加载逻辑，保存频率也可能需要通过训练脚本控制。
<https://docs.aws.amazon.com/sagemaker/latest/dg/model-checkpoints.html>。

⁵ 工业和信息化部办公厅：《关于加强互联网数据中心客户数据安全保护的通知》（工信厅网安〔2025〕5号）。该通知要求 IDC 业务经营者根据合作模式和内容明确各方数据安全保护责任，并针对人工智能训练数据集管理、算力调度及算力服务提出相应安全要求；同时规定，在实施可能影响客户数据安全的高危操作和对外提供客户数据前，应告知客户并取得授权。
https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2025/art_1bbf7c744c994183abb3ad6148658960.html。

⁶ 国家标准 GB/T 45958-2025《网络安全技术 人工智能计算平台安全框架》，发布日期为 2025 年 8 月 1 日，实施日期为 2026

具体责任判断中的作用也需要结合适用依据和合同约定分析，二者都不宜被简单理解为适用于所有平台的统一赔偿规则。

在产品和合同设计中，更可操作的方式是把基础服务已经包含的保护措施、增值产品新增的能力以及客户仍需自行完成的配置分别说明清楚；如果某项安全功能允许客户主动关闭，应说明其作用、关闭后果和替代措施，并保存操作记录。事故发生后，也不能仅凭客户曾关闭某一安全开关就直接推导全部责任，如果数据泄露实际源于平台负责的租户隔离或者底层访问控制缺陷，而客户关闭的是与该缺陷无直接关系的告警功能，两者与损害之间的关系仍需分别判断；如果客户关闭的恰恰是能够有效阻断或者显著减轻损害的保护措施，其行为则可能进一步影响损失承担。

二、将安全义务落实到运行、应急和退出

采购合同和上线验收解决的只是服务开始时的责任边界，AI 算力服务进入持续运行后，平台还要进行日常运维、异常处置，合作结束时又涉及数据、模型和其他数字资产的迁出与删除，因此责任分工还需要覆盖整个服务生命周期。

（一）在必要运维与数据使用之间划定边界

为了保障服务稳定运行，平台通常需要掌握资源利用率、设备状态、账号异常和安全告警等信息，但监测 GPU 利用率与读取客户训练样本、提示词、模型权重和模型输出，并不是同一层面的数据处理活动。技术上能够访问某类数据，并不意味着平台因此取得查看、复制或者用于其他目的的当然权限。AI 算力场景中的训练数据可能包含个人信息或者企业内部资料，模型权重和 checkpoint 可能具有独立商业价值，推理输入输出也可能直接反映客户业务内容，因此平台对这些数据的访问和使用应与具体服务目的相匹配。

例如，平台按照客户指示处理客服对话，用于向客户提供模型推理服务，在符合其他适用条件的情况下，双方可能形成个人信息委托处理关系；如果平台又自行决定把这些对话用于训练或者改进自己的通用模型，则可能已经超出原有委托处理范围，需要重新判断处理目的、处理依据以及平台自身在相关活动中的法律身份。双方不宜仅以“委托方”“受托方”两个标签覆盖全部数据活动，而应区分运维、推理、评测、训练和模型改进等用途，判断平台究竟是在按照客户指示处理数据，还是已经对部分处理目的和方式作出独立决定。

“不用于训练”也不能等同于“不保存任何数据”，为故障排查、安全审计或者计费保留必要日志，并不当然与不训练承诺矛盾，但必要留存也不应演变为无限期保存完整提示词、训练文件或者模型输入输出。双方可以分别约定训练数据、推理输入、模型输出、运维日志、checkpoint 和备份的用途、访问范围及保存期限，平台内部人员确需接触客户数据时，通过工单审批、权限限制和操作记录控制访问；客户则应在数据上传前确认自身具有相应使用权限，并根据岗位和业务需要配置、更新访问权限。

（二）用应急流程衔接及时止损与授权操作

安全事件发生后，平台可能需要迅速暂停训练任务、限制账号访问、隔离资源或者实施恢复，这些措施既可能是防止损害扩大的必要手段，也可能影响客户业务甚至改变数据和模型状态。因此，与其在事故现场临时讨论“平台能不能操作”，更适合在合作开始时就区分哪些属于平台控制范围内可以或

者应当立即实施的保护措施，哪些会深入客户环境或者改变客户资产，需要进一步取得授权。

工信部前述 IDC 客户数据安全保护通知要求，在实施可能影响客户数据安全的高危操作以及对外提供客户数据前，应当告知客户并取得授权⁷。双方可以据此在合同附件或者应急预案中明确联系人、事件分级、平台可直接实施的必要措施以及需要客户确认的高影响操作，并尽量把授权落实到触发条件、操作对象和操作类型。

授权与操作责任也需要区分。假设训练任务最初因客户代码异常中断，客户授权平台工程师介入恢复，但工程师在操作中又覆盖或者损坏原本可用的 **checkpoint**，那么最初中断与恢复操作造成的新增损失需要分别评价；客户授权解决的是平台是否有权执行操作，并不意味着客户当然接受不当操作造成的后果。对于可能覆盖、删除或者改变模型状态的操作，可以根据项目重要性预先设计副本验证、变更记录和回滚路径。

事件通知和监管报告也不应等待最终赔偿责任查明以后才启动。《国家网络安全事件报告管理办法》要求网络运营者通过合同等方式，要求为其提供网络安全、系统运维等服务的组织或者个人及时报告监测发现的网络安全事件并协助履行报告义务；对于规定时限内暂时无法判定有关情况的，可以先报告已经掌握的基本信息，再及时补报⁸。因此，平台和客户宜提前约定内部告知时限、事件信息共享、对外报告主体和证据交接安排；涉及个人信息安全事件时，还需结合《个人信息保护法》等规定判断补救、个人通知和监管报告义务。

（三）在退出时完成资产交接与数据处置

AI 算力服务结束时，客户可能需要迁出模型、**checkpoint**、训练日志、配置文件等数字资产，平台和上游服务商则可能仍保存快照、缓存、备份或者运行日志。如果这些问题直到停服时才开始讨论，很容易影响业务连续性，因此，迁出安排最好在采购阶段就明确，包括可导出的资产范围、文件格式、办理期限、费用和必要技术支持，并在关闭旧环境前验证模型权重、**checkpoint** 等关键文件的完整性和可用性。

数据删除同样要对应真实技术架构，因为删除原始文件并不意味着由其生成的向量、文本片段、索引、缓存或者备份会自动同步删除，知识库和 RAG 场景尤其如此。双方可以明确原始资料、索引、副本和备份的处理方式、完成期限和确认记录；依法确需继续保存的，应说明保留依据、用途和访问范围。与此同时，“资产能够导出”并不意味着客户当然有权不受限制地继续使用，基础模型许可范围、微调权重或者适配器的权属和使用条件、保密义务以及跨平台迁移限制，也应在合作开始时一并核对。退出管理因此既涉及数据安全和业务连续性，也涉及合同和知识产权安排。

三、随着应用功能扩展重新审视责任范围

随着平台从基础算力向模型服务、企业知识库、RAG 和智能体延伸，其影响的已经不只是“计算是否正常”，还包括哪些信息能够进入模型、模型内容如何呈现，以及模型输出能否直接触发外部业务操作。责任共担模型也需要随服务功能向上延伸，但基本判断方法仍然不变，即确认相关功能由谁提供、谁控制、谁

⁷ 同脚注 5。

⁸ 《国家网络安全事件报告管理办法》第五条要求网络运营者通过合同等形式，要求为其提供网络安全、系统运维等服务的组织或者个人及时报告监测发现的网络安全事件并协助报告；第七条规定，对于规定时间内不能判定事发原因、影响或者发展趋势等情况的，可以先报告已经掌握的部分基本信息，其他情况及时补报。

https://www.cac.gov.cn/2025-09/15/c_1759583017717009.htm。

作出了相应承诺。

（一）当第三方模型风险应当对应具体的集成和维护行为

不少 AI 平台会调用第三方或者开源模型，但“模型来自第三方”本身不能回答发生问题后由谁负责，需要区分模型固有能力边界和平台自身的集成、部署或者维护缺陷。例如，通用模型无法稳定回答高度专业的问题，可能首先涉及模型能力边界、服务说明以及客户是否将其用于适当场景；如果系统异常实际源于平台负责维护的模型加载组件存在已知漏洞，或者平台配置错误导致部署环境不稳定，则属于另一类问题。即使两者最终都表现为“模型调用失败”或者“结果错误”，也不能仅因为底层模型不是平台自行训练，就排除平台在其负责环节的责任。

在采购和上线阶段，可以相应明确模型来源、版本、许可条件、升级机制以及集成维护责任，由承担部署和维护工作的一方记录模型版本、重要依赖、已知限制及更新安排，客户则结合业务场景验证模型的准确性、稳定性和适用性，并针对高影响用途安排必要的人工复核。平台说明大模型可能出现错误或者幻觉，并不会使其已经承诺的部署、接口可用性、版本管理和安全维护义务消失；客户完成上线测试，也不意味着接受未来所有因平台维护缺陷产生的风险。

（二）先确认监管身份，再落实内容治理和标识分工

模型服务还可能触发独立于合同责任之外的监管义务。《生成式人工智能服务管理暂行办法》的适用，需要结合是否向境内公众提供生成文本、图片、音频、视频等内容的生成式人工智能服务等因素判断，其对“生成式人工智能服务提供者”的定义也包括通过提供可编程接口等方式提供相关服务的组织和个人，因此，“平台只是提供 API”不足以当然排除相关服务提供者身份；反过来，未向境内公众提供生成式人工智能服务，仅用于企业内部研发或者生产经营的，也不应仅因使用大模型就当然适用同一套义务。《互联网信息服务深度合成管理规定》还分别针对相关服务提供者和技术支持者设置相应责任，同样需要根据具体业务角色判断。

因此，在模型产品上线或者服务模式发生重要变化时，可以结合服务对象、内容类型、提供方式和实际控制环节先判断监管身份，再确定内容治理、算法备案、生成式人工智能服务相关备案或者登记、评估以及生成合成内容标识等事项由谁履行、另一方需要提供何种配合。涉及生成合成内容标识要求时，还应进一步落实到生成、接口传输、展示、导出和传播等环节。合同可以约定双方如何协作，但不能通过商业约定把依法应由某一主体自身履行的监管义务当然转移给另一方。

（三）RAG 需要让检索结果服从终端用户的访问权限

企业知识库和 RAG 带来的一个典型风险是，模型回答可能完全准确，但其依据的资料本来不应向当前终端用户披露。RAG 通常先从企业文档、数据库或者其他知识源检索内容，再把相关片段交给模型生成回答，其中至少存在两个不同的权限问题：后台检索服务是否有权读取资料，以及当前终端用户是否有权获得资料，两者不能混为一谈。

第一类风险来自原有权限设置过宽。例如，本应只供特定项目组使用的经营材料被错误设置为全员可见，在传统文档系统中其他员工未必容易发现，但接入企业 AI 问答后，用户通过自然语言提问就可能迅速检索和归纳相关内容，从而放大原有过度共享问题。此时，更准确的责任事实是源系统权限超出了实际业务需要，而不是模型突破了访问控制，客户对数据分类、文档权限和人员授权的管理仍然是责任判断的重要因素。

另一类风险则发生在 RAG 系统本身没有正确执行终端用户权限。后台检索服务为了建立索引或者

检索资料，可能具有读取大量文档的技术权限，但如果向模型提供检索结果时没有继续按照当前用户的身份和访问范围进行筛选，原本受限的资料就可能进入模型并出现在无权用户看到的回答中。此时需要判断权限过滤功能究竟由平台还是客户提供，身份信息如何传递，权限变化由谁负责同步，以及平台是否曾承诺相关检索功能能够继承或者执行客户的访问控制要求。

因此，如果源文件本身由客户错误设置为全员可见，重点通常在客户的数据治理和权限管理；如果平台承诺按照终端用户权限进行检索，却因自身功能实现或者维护问题没有正确执行，则需要审查平台的功能交付；如果客户只是租用基础 GPU 并自行搭建整套 RAG 应用，检索层缺陷也不能因为应用运行在该平台上就当然归属于基础设施提供方。RAG 安全控制应尽量前移到资料进入模型之前，上线验收可以使用有权和无权账号提出相同问题，并在撤销权限后再次测试，检查回答和引用内容是否相应变化。对法务和合规人员而言，真正需要确认的并不是底层检索技术采用了哪一种实现，而是合同中所称的“权限控制”是否在完整业务链路中真正生效。

（四）智能体应将业务授权落实到实际执行环节

如果说 RAG 主要解决“什么信息可以进入回答”，智能体进一步带来的问题就是“系统能够执行什么行为”。当智能体可以调用电子邮件、CRM、工单、数据库或者支付接口后，模型输出可能直接导致邮件发送、业务记录修改、文件删除甚至交易发起，此时仅把问题概括为“模型幻觉”并不足够，因为从模型产生建议到最终动作执行，中间还包含工具选择、权限授予、人工确认和下游系统执行等多个环节。

例如，客户主动给智能体授予修改全部客户记录的权限，并允许其无需人工确认即可执行，与平台已经承诺提供“删除记录前必须人工确认”的控制机制却因程序缺陷没有触发确认，是两个性质不同的责任事实；如果外部工具或者第三方系统又超出请求范围执行操作，还需要进一步分析相应工具提供方的行为。责任判断需要沿着实际执行链条，逐一确认谁决定智能体能够调用哪些工具、谁配置业务权限、谁提供人工确认机制，以及最终业务系统由谁控制。

从安全实践看，OWASP（Open Worldwide Application Security Project，一个用于辅助说明 AI 系统安全控制实践的行业权威参考，但非法律责任确定的直接依据）针对大语言模型系统“Excessive Agency”风险提出了限制可调用功能、给予最小必要权限、对高影响操作引入人工批准，以及由下游系统继续执行授权校验等措施⁹。这些原则与责任划分的思路实际上是一致的：谁决定智能体可以调用哪些工具，谁配置业务权限，谁提供人工确认，谁控制最终业务系统的执行，就分别掌握了不同的风险控制节点。对于付款、删除关键数据、向外发布信息等高影响操作，可以根据业务风险设置人工确认和额度限制，由最终执行系统继续校验身份和权限；合同和验收文件则进一步明确由谁提供这些控制、谁维护业务规则，以及授权记录、工具调用参数和执行结果如何保存。

结语

AI 算力平台与客户之间的责任边界，并不是只能等到事故发生以后才能回答的问题。从采购阶段明确平台实际提供的功能、服务主体和交付标准，到上线时验证训练恢复、数据访问、RAG 权限和智能体执行控制，再到运行阶段保存必要的使用和应急记录、退出阶段完成模型、checkpoint 以及其他数字资产的交接

⁹ OWASP GenAI Security Project, *LLM06:2025 Excessive Agency*. OWASP 将过多功能、过大权限和过高自主性列为 Excessive Agency 的典型根源，并建议限制智能体可调用的功能、给予最小必要权限、对高影响操作引入人工批准，以及由下游系统继续进行授权校验。

<https://genai.owasp.org/llmrisk/llm062025-excessive-agency/>.

和数据处置，实际上都是在把抽象的“责任共担”转化为可以被观察和验证的服务事实。

对平台而言，重要的是让服务说明、合同承诺和实际产品能力保持一致，避免一方面通过产品功能深度参与客户训练和 AI 应用运行，另一方面又以概括性的“应用由客户负责”排除全部责任；对客户而言，也需要管理好自身掌握的数据、账号、权限、模型使用场景和业务授权，不能因为购买了托管或者平台化 AI 服务，就认为所有风险已经转移。双方还可以围绕具体产品形成持续更新的责任附件，在新增上游供应商、改变数据用途、调整资源和数据地域、增加模型服务、接入知识库或者开放智能体执行能力时重新核对责任分工。

责任共担模型真正有价值的地方，并不是在风险发生以后帮助某一方证明“这不是我的责任”，而是在服务设计、采购、上线和运行阶段提前回答三个更具体的问题：这项功能由谁提供，关键控制掌握在谁手中，出现问题以后应当根据哪些事实判断责任。只有把这些问题说明白，AI 算力平台的责任共担才可能真正成为日常可以执行、发生争议后也能够由法院或者仲裁庭核验的责任分配机制。

特别声明

汉坤律师事务所编写《汉坤法律评述》的目的仅为帮助客户及时了解中国或其他相关司法管辖区法律及实务的最新动态和发展，仅供参考，不应被视为任何意义上的法律意见或法律依据。

如您对本期《汉坤法律评述》内容有任何问题或建议，请与汉坤律师事务所以下人员联系：

陈文昊

电话： +86 21 6080 0359

Email: wenhao.chen@hankunlaw.com