附件

# “铸盾模都”2026年人工智能安全赋能专项行动任务清单

| 序号 | 2026年重点任务 | | | 企业提报材料 |
|---|---|---|---|---|
| 一 | 人工智能安全主体责任 | 1.加强网络和数据安全管理组织建设。 | 企业应明确网络和数据安全负责人，设立专门的安全管理机构，强化企业内部监督管理及资源保障。 | - |
| | | 2.健全人工智能安全管理制度体系。 | 企业应建立健全人工智能网络和数据安全管理制度，强化安全风险防范，形成全链条闭环的安全管理运营体系。 | - |
| | | 3.明确人工智能技术与应用服务安全责任主体。 | 企业在对外提供大模型、智能体、具身智能等人工智能技术与应用服务时，应明确相关参与主体的网络和数据安全责任与义务。 | - |

| | | | | |
|---|---|---|---|---|
| 二 | 人工智能安全保障基线 | 4.加强人工智能网络安全管理。 | 企业应做好网络安全防护措施的同步规划、同步建设、同步运行，基于不同阶段的安全风险，常态化开展资产梳理、代码审计、漏洞扫描、暴露面管理、安全加固等安全管理工作，健全全域安全防护能力建设，落实通信网络安全防护定级备案相关工作要求。 | - |
| | | 5.加强人工智能数据安全管理。 | 企业应严格规范管理训练数据、测试数据、用户输入数据及其处理活动，确保全生命周期安全合规。重点防范恶意数据或指令注入、虚假信息传播等安全风险，做好数据质量管控措施，建立相应溯源、审计及数据备份恢复机制，留存日志不少于6个月。企业应落实数据分级保护、重要数据识别等工作，定期进行数据资产梳理，开展重要数据及1000万人以上个人信息识别工作，形成并定期更新本单位重要数据目录，在8月31日前向市通管局备案。 | 企业应在8月31日前向市通管局备案重要数据目录。 |
| | | 6.加强人工智能用户个人信息安全 | 企业开展人工智能技术研发与提供服务、应用时，应以最小必要原则采集个人信息，采取匿名化、去标识化、脱敏及差 | - |

| | | | | |
|---|---|---|---|---|
| | | 保护。 | 分隐私等技术加强安全防护，并以清晰、准确、易于理解的方式告知用户个人信息处理规则、收集范围、用途及存储期限等，依法保障用户知情、更正、删除等合法权益。企业应依法依规定期对其处理个人信息情况开展个人信息保护影响评估与合规审计。向第三方提供或委托处理个人信息、重要数据的，企业应严格约定、落实双方安全责任并留存数据处理情况记录不少于3年。 | |
| | | 7.加强人工智能业务与内容安全管理。 | 企业应规范深度合成、生成式人工智能等应用，提升涉生成违法违规内容识别与拦截能力，做好生成合成内容标识工作并依法留存相关日志不少于6个月。企业应结合业务实际梳理人工智能业务清单，并在11月30日前将年度人工智能业务清单上报市通管局。 | 企业应在11月30日前将年度人工智能业务清单上报市通管局。 |
| | | 8.加强人工智能供 | 企业应建立健全第三方供应链安全管理制度。运用技术手 | |

| | | | | |
|---|---|---|---|---|
| | | 应链安全管理。 | 段，强化对第三方插件及第三方模型的供应链准入管控、安全评估和漏洞管理，并形成软件物料清单（SBOM）等供应链底账。加强对抗训练、提示词过滤和异常行为检测，防范开源社区、插件工具等供应链恶意投毒。 | |
| | | 9.加强智能体安全管理。 | 企业应严格划定智能体自主行为边界，明确执行权限范围，建立完整的智能体行为溯源记录机制，实现决策、执行、工具调用等智能体全流程管控。涉资金转账、数据删除等高风险操作，应落实二次确认或强制人工操作等管控措施。对调用终端底层、跨系统横向访问等敏感操作，应取得用户明确授权，防止未授权擅自执行。 | |
| | | 10.加强具身智能安全管理。 | 企业应强化设备身份认证、OTA 升级安全、运维安全及交互安全，健全碰撞检测与自动避让设计，强化异常行为检测，落实硬件运动阈值与设备失控紧急制停机制，规范涉自动化作业流程的人机安全接管机制，防范视觉欺骗、传感器干扰等物理对抗攻击。强化本地数据全生命周期及远程访问权限 | |

| | | | | |
|---|---|---|---|---|
| | | | 管控，对人脸、语音、环境等敏感数据实行端侧脱敏，规范存储、销毁流程，防止因设备丢失或报废导致数据泄露。 | |
| | | 11.强化人工智能安全应急处置能力。 | 企业应建立应急工作组与响应机制，制定有效可操作的安全事件应急预案，明确事件分级标准、响应处置流程等内容，每年至少开展一次应急演练并完成问题整改闭环。 | |
| 三 | 人工智能安全管理赋能 | 12.加强人工智能安全风险监测与威胁信息共享。 | 鼓励有关企业和机构主动开展人工智能技术与应用的漏洞挖掘、安全检测、安全伦理审查并及时上报至市通管局。市通管局定期综合情况发布人工智能安全风险预警与威胁通报，企业应及时处置相关安全风险隐患并反馈整改结果。 | |
| | | 13.加强行业安全防护能力。 | 市通管局持续深化人工智能技术在安全领域的创新应用，提供面向传统网络安全检测的人工智能网络安全自动化渗透智能体以及面向人工智能安全的测评能力，组织“磐石行动”人工智能专项攻防活动，为企业提供高效精准的安全改进建议。 | |