附件

# “铸盾模都”2025年人工智能安全赋能专项行动任务清单

| 序号 | 2025年重点任务 | | | 企业提报材料 |
|---|---|---|---|---|
| 一 | 人工智能安全主体责任 | 1.加强人工智能安全管理组织建设 | 企业应明确网络和数据安全负责人，压实主体责任；涉及关键基础设施运营或重要数据处理的，还应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。 | - |
| 二 | 人工智能安全保障 | 3.加强人工智能网络安全 | 企业应结合不同阶段的安全风险，规范开展资产安全管理、漏洞安全管理、模型更新安 | - |

| | | | | |
|---|---|---|---|---|
| | 基线 | | 全等工作，不断健全安全防护措施、持续完善安全能力建设，落实通信网络安全防护定级备案相关工作。 | |
| | | 4.加强人工智能数据安全 | 企业应规范预训练和优化训练数据及其处理活动的安全要求，在通用数据安全基础上，做好训练数据安全防护、安全标注以及重要数据识别等工作。 | – |
| | | 5.加强人工智能内容安全 | 企业应规范深度合成、生成式人工智能等应用，规范内容生成模板、规则配置，健全信息发布全流程审核机制，提升涉生成违法违规内容识别与拦截能力。 | – |
| | | 6.加强人工智能业务安全 | 1.企业应结合业务实际梳理人工智能业务清单并上报市通管局，同时在人工智能业务上线前开展安全评估。<br>2.涉及到自主研发大模型以及部署开源大模 | 企业于11月30日前提交年度人工智能业务清单。 |

| | | | | |
|---|---|---|---|---|
| | | | 型并对外提供服务的企业，应自行或委托第三方机构对使用大模型的业务开展安全评测。<br>3.市通管局定期组织对人工智能业务的安全评估评测情况进行抽测。 | |
| | | 7.强化人工智能应急管理与能力 | 企业应建立网络与数据安全应急工作组织与响应机制，制定有效可操作的安全事件应急预案，每年开展一次应急演练并针对发现问题开展闭环整改。 | – |
| 三 | 人工智能安全管理赋能 | 8.提供人工智能安全评测工具 | 市通管局组织提供可复制、易推广的评测工具和开源代码。 | 鼓励企业主动提供可开源的人工智能安全工具、测试数据集等。 |
| | | 9.建立人工智能安全风险监测巡查和处 | 1.鼓励有关企业和机构主动开展安全风险监测上报。 | 1.鼓励有关企业和机构主动开展 |

| | | | | |
|---|---|---|---|---|
| | | 置机制 | 2.市通管局对人工智能技术和应用开展常态化安全风险监测巡查，并综合巡查结果与企业上报情况定期发布人工智能安全风险预警，开展威胁通报。<br>3.企业应开展内部常态化安全巡检，针对发现和通报的安全隐患及时开展整改处置。 | 安全风险监测上报。<br>2.企业应提供接受风险监测巡查所必要的技术支持，包括但不限于开放智能应用的API接口、测试用账号等。 |
| 四 | 人工智能安全行业发展 | 10.提供安全合规及应用实践指导 | 1.市通管局组织开展“磐石行动”人工智能安全（专项）攻防演练活动。<br>2.市通管局基于演练成果系统梳理优秀防御实践。 | - |
| | | 11.加快培育人工智能安全生态 | 1.市通管局面向重点企业开设人工智能安全风险防范与治理能力提升专题培训。 | 鼓励企业上报在人工智能内生安 |

| | | | | |
|---|---|---|---|---|
| | | | 2.市通管局组织开展“铸盾模都”2025年人工智能安全赋能典型案例评选，支持和培育在人工智能内生安全和人工智能赋能安全等领域有突出成效的企业。 | 全和人工智能赋能安全等领域中的典型案例。 |