需求参数公示

1.技术标准

采购内容：智能安全靶场1套。

（一）功能参数

1.总体要求

★（1）提供智能安全训练所需的决策式人工智能模型和生成式人工智能模型，以及配套数据集。

★（2）提供智能安全训练案例和配套训练指导书，训练案例内容包含：样本攻防、数据投毒攻击、模型窃取攻防、大模型攻防。

★（3）提供模拟攻击演练能力。基于系统提供的靶标及数据，培养演练人员可自主构造模拟攻击样本能力。

★（4）提供防御演练能力。基于系统提供的模拟攻击技术，培养演练人员可自主加固模型能力。

★（5）支持决策式模型靶标安全测试功能。支持样本攻击、数据投毒攻击、窃取攻击等决策式模型模拟攻击。

★（6）支持生成式大模型靶标安全测试功能。支持角色扮演攻击、注意力转移等生成式大模型模拟攻击。

2.模型基础应用训练

★（1）提供靶标模型使用的技能训练案例。包含算法上传、

训练数据选择、参数配置、训练结果展示、模型发布等。

★（2）提供靶标模型开发优化的技能训练案例。包含模型上传、模型代码编辑、模型信息编辑、参数配置等。

★（3）提供大模型靶标测试的技能训练案例。包含大模型靶标 prompt 编辑、大模型文本输入、大模型文本输出等。

3.样本攻防训练

★（1）提供智能算法样本模拟攻击技能训练案例。案例覆盖的模拟攻击方法包括但不限于白盒和黑盒攻击。评估方法能够量化模型在遭受这些模拟攻击时的表现，包括攻击成功率、模型准确率下降幅度以及样本的可迁移性等关键指标。

★（2）提供智能算法样本防御技能训练案例。案例覆盖的防御方法包括但不限于训练、输入去噪、函数增强等。

★（3）提供样本攻防训练案例对应的训练指导书、训练环境资源和评估结果。

4.数据投毒模拟攻击训练

★（1）提供智能算法投毒模拟攻击技能训练案例。案例覆盖的评估方法包括攻击成功率、模型性能下降程度、以及攻击对模型决策的影响等关键指标。

★（2）提供智能算法投毒防御技能训练案例。案例覆盖的防御方法包括数据验证、异常检测、可信度评估、分布一致性检查等。

（3）提供智能算法投毒攻防训练案例对应的训练指导书、训练环境资源和评估结果。

5.模型窃取攻防训练

★（1）提供智能算法模型窃取模拟攻击技能训练案例。案例覆盖的攻击方法包括黑盒查询、白盒查询等。

（2）提供智能算法模型窃取防御技能训练案例。案例覆盖的防御方法包括查询限制、噪声注入、输出扰动等。

★（3）提供智能算法模型窃取攻防训练案例对应的训练指导书、训练环境资源和评估结果。

6.大模型攻防训练

★（1）提供大模型算法模拟攻击技能训练案例。案例覆盖的攻击方法包括鲁棒性攻击、隐私性攻击等。

★（2）提供大模型算法防御技能训练案例。案例覆盖的防御方法包括大模型算法的防护与拦截能力。

★（3）提供大模型攻防训练案例对应的训练指导书、训练环境资源和评估结果。

7.应用场景

★（1）提供5种典型人工智能安全应用场景构建所需的模型、数据集等资源。至少包括人脸识别认证、内容审核、车辆识别、无人机识别、自动驾驶。支持内置样本数据以及自定义样本数据两种样本模式。

（2）人脸识别认证场景。模拟人脸识别认证、人脸验证等功能，支持人脸识别认证训练。

（3）内容审核场景。模拟社交媒体文章、重要发言等功能，支持内容审核训练。

（4）车辆识别场景。模拟支持围绕特种车辆识别、特殊环境下车辆识别等典型车辆监测、追踪功能。支持车辆识别算法训练。

（5）无人机识别场景。模拟多种典型机型无人机识别、极端环境下无人机识别等功能。支持无人机识别算法训练。

（6）自动驾驶场景。模拟各类典型人车物识别、特殊环境下障碍物识别等功能，支持针对自动驾驶感知模型进行定向模拟攻击与防御训练。

8.效果评估

★（1）提供相关的评估维度、评估指标和评估结果。

（二）技术指标要求

1.总体要求：

★（1）提供决策式人工智能模型种类≥10 种。至少包括：VGG、ResNet、inception、YOLO、faster RCNN 等。

★（2）提供生成式大模型种类≥3 种。包括：Qwen2-instruct7B、ChatGLM36B、SecGPT 或其他同类轻量化大模型。

★（3）支持数据集种类≥4 种。至少包含图像、文本、语音、表格。其中图像数据集≥10 个，总计样本数≥5 万；文本数据集≥4 个，总计样本数≥2 万；语音数据集≥2 个，总计样本数≥2 万；表格数据集≥2 个，总计样本数≥2 万。

（4）文本、表格数据集支持 txt、csv、json、xml 等文件格式；图像数据支持 jpg、png 等文件格式；压缩文件支持 zip、tar 等文件格式；语音数据集支持 mp3、wav 格式。

2.样本攻防

★（1）提供决策式人工智能样本攻击算法≥10 个，覆盖白盒与黑盒算法。包括 FGSM、PGD、CW、DeepFool、BIM、TextBugger、Boundary、ZOO、HSJA 和 NES 等。

★（2）决策式人工智能样本防御包含训练、样本检测和样本生成。

（3）提供训练方法≥5 种，包括 PGDDK 训练、TRADES、Free Adversarial Training、YOPO 和 Fast Adversarial Training。

（4）提供样本检测方法≥3 种，至少包括 Spatial Smoothing、Feature Squeezing。

3.数据投毒模拟攻击

★（1）决策式人工智能模型投毒模拟攻击包括数据中毒型后门攻击和模型注入型后门攻击。

（2）数据中毒型后门模拟攻击方法≥4 种，包括 BadNets、

Trojan、Feature Collision 和 Triggerless。

（3）模型注入型后门模拟攻击方法≥4 种，包括 Dynamic Backdoor、Physical Backdoor、Neuron Interference 和 Model Poisoning。

★（4）提供的基础模拟攻击方法≥5 种，包括单步梯度下降、多步梯度下降、基于优化、动量多步梯度下降和后门投毒。

★（5）提供决策式人工智能模型投毒防御算法≥3 个，覆盖白盒与黑盒算法。至少包括 NC、STRIP 等。

4.模型窃取攻防

★（1）支持的窃取模拟攻击类型≥3 种，包括模型模拟、模型输出逆向重构和模型权重逼近。

（2）防御措施≥3 种，包括 API 访问限制、模型水印和查询噪声。

5.大模型攻防

★（1）支持生成式大模型模拟攻击算法≥5 个。至少包括 DAN 攻击、角色扮演攻击、注意力转移攻击等。

★（2）支持生成式大模型防御算法≥3 个。至少包括自处理防御、自提醒防御等。

6.训练场景

★(1)人脸识别认证场景。提供 DeepFace、FaceNet、VGGFace、SphereFace、ArcFace 等经典人脸识别靶标模型。提供 LFW、YaleB、

CelebA、MegaFace、VGGFace2、CASIA-WebFace 等人脸识别常用开源数据集。

★（2）人脸识别认证场景。提供模拟攻击训练技能≥6 项。提供躲避攻击和假冒攻击样本检测训练技能≥6 项。包含 BIM、DIM、TIM 攻击方法和 HGD、TVM 防御方法。

★（3）内容审核场景。提供内容审核靶标模型和配套内容审核数据集。

★（4）内容审核场景。提供文本模拟攻击训练技能≥5 项。提供文本防御训练技能≥5 项。包含字符级别白盒攻击算法 HotFlip 和查询攻击算法 DeepWodBug、TextBugger、TextFooler、Genetic 等攻击方法和 Adversarial Training 、SEM 、semi-character-RNN 和 DISP 等防御方法。

★（5）车辆识别场景。提供 YOLO 系列、FasterR-CNN、SSD 等车辆识别靶标模型。提供 KITTI、UA-DETRAC、BDD100K 等车辆识别数据集。

★（6）车辆识别场景。提供图像训练攻击训练技能≥6 项。提供防御训练技能≥5 项。包含：针对车辆识别模型可见光数据的 FGSM、PGD、MIM、C&W、DEEPFOOL、BadNet 等攻击方法和图片压缩、图片放缩、PGD、Neural Cleanse、Adversarial Training 防御方法。

★（7）无人机识别场景。提供 Drone-YOLO 无人机识别靶标

模型。提供 Drone Dataset、Drone-type dataset、YOLO Drone Detection Dataset 等无人机识别数据集。

★（8）无人机识别场景。提供图像训练攻击训练技能≥7项。提供防御训练技能≥4项。包含针对无人机识别模型可见光数据的 VMIM、FGSM、PGD、MIM、C&W、DEEPFOOL、BadNet 等攻击方法和图片压缩、图片放缩、PGD、Neural Cleanse 防御方法。

★（9）自动驾驶场景。提供 UniAD、VAD、DriveGPT4 等自动驾驶靶标模型。提供 Cityscapes、Imagenet（ILSVRC）、COCO、PASCAL VOC、CIFAR、MNIST 等自动驾驶数据集。

★（10）自动驾驶场景。提供攻击训练技能≥6项。提供防御训练技能≥2项。包含针对自动驾驶模型 BadNet、FGSM、BIM、PGD、Square Attack、NES 等攻击方法和 FGSM-AT、PGD-AT 等防御方法。

2、商务要求

★（一）服务时间、服务地点和方式

1.服务时间：合同生效后，中标供应商在 30 日内完成阶段性试用版本的安装及调试，在 90 日内完成安装及调试。

2.服务地点：湖南省长沙市，采购单位指定地点。

3.服务方式：中标供应商提供软件的各项技术性能指标必须达到合同和技术文件规定的要求。

（二）生产及安装调试等要求

1.中标供应商负责所有软件的供货与安装、调试，解决系统联调中相关技术问题。

2.中标供应商应配合靶场管理平台项目的中标供应商，完成相关的集成工作。

（三）售后服务

★1.质量保证期：自产品验收完毕之日算起，所有产品质保至少 5 年。质保期内中标供应商提供免费的维护服务、系统升级服务、上门服务，提供 7*24 小时电话、线上技术支持。质保期满后，终身免费技术支持维护。

2.中标供应商负责对采购单位使用、维护人员进行软件的安装、调试、使用和运维培训。至少免费培训 4 名操作维护人员；具体培训计划（时间、地点、人数、内容）由中标供应商提供，其费用应包括在投标价格内。

3.中标供应商应有常用配件库和快速服务保障能力。

4.出现产品、使用问题采购单位提出后，中标供应商在 30 分钟内运维人员远程故障响应，24 小时内提供解决方案，48 小时内提供相关的维修、升级等解决问题的服务。质保期内如系统故障导致系统停用时间超过 7 天，质保期按照故障时间顺延，期间产生的损失由中标供应商承担。

5.定期提供例行检查和维修服务。

★（四）知识产权和保密要求

1.对采购单位提供的人员、地址、采购情况等信息要保守秘密，不得向外界透露。中标通知书发出后，采购单位将与中标供应商签订保密协议。

2.基于项目合同履行形成的知识产权和其他权益，其权属归采购单位所有，法律另有规定的除外。

3.中标供应商在采购和履行合同过程中所获悉的采购单位属于保密的内容，中标供应商具有保密义务。

4.中标供应商应保证采购单位在使用该货物或其任何一部分时，不受第三方侵权指控。同时，中标供应商不得向第三方泄露采购单位的技术文件等资料。

★（五）付款及结算方式

合同签订后，采购单位在30个工作日内向中标供应商支付30%合同预付款。货物交付并验收合格后，中标供应商提供全额发票，采购单位在30个工作日内向中标供应商支付65%合同款，剩余5%合同款为质量保证金，质保期满且无质量问题，采购单位在接到中标供应商的质量保证金返还申请后30日内无息全额支付。

（六）报价要求

本项目以人民币报价（含税），投标供应商的投标报价不得高于本项目最高限价，否则为无效投标。本项目采用合同总价方式，报价总价一次性包干，包含供应、运输、安装调试、技术培

训、售后服务、备品备件和伴随服务等价格，价格不因实施期间市场变化及政策调整因素而变化。

（七）验收方式

1.在完成全部采购及服务内容后，中标供应商应对交付物进行全面自检，符合交付条件后，由中标供应商向采购单位提出验收申请，采购单位按照合同、招标文件及中标供应商投标文件要求以及学校验收相关规定组织专家对本项目进行验收。具体组织程序、验收标准和方法，按采购单位规定程序执行，中标供应商配合。

2 采购单位在验收中，如果发现有与合同规定不符的，应在10 个工作日内向中标供应商提出书面异议，不签发验收单。

3.中标供应商在接到采购单位书面异议后，应在 10 日内予以纠正，并承担由此发生的一切费用和损失。再次验收所产生的费用由中标供应商承担。如果再次验收仍不合格，采购单位有权取消或解除采购合同，由此造成的损失，由中标供应商承担。

4.采购单位在中标供应商按合同规定交货或安装、调试后，无正当理由而拖延接收、验收或拒绝接收、验收的，应承担因此给中标供应商造成的直接损失。

5.采购单位对货物进行检查验收合格后，应当收取发票并在《交货验收单》上签署验收意见及加盖单位印章。

（八）实施人员要求

本项目技术负责人取得高级职称且在本投标供应商工作3年（含）以上。拟派项目技术负责人须为投标供应商正式员工，提供技术负责人身份证复印件、相关证书复印件或网上查询截图及对应年限的由投标供应商缴纳社保证明材料的复印件(代缴社保证明材料不予认可）。投标供应商在软件开发期间投入本项目人员（不含项目负责人）不少于5人。开发项目团队和人员需为投标供应商正式员工。

（九）现场演示要求

★1.本项目评审现场进行案例系统演示，未演示的投标无效。

2.依据技术评审表中案例系统演示要求进行演示，系统演示及所需要的相关设备由投标供应商自行准备，演示设备禁止连接国际互联网；演示中使用的数据由投标供应商自行准备；演示总时间不超15分钟。

3、资质要求

（一）具有企（事）业法人资格（有行业特殊情况的银行、保险、电力、电信等法人分支机构，会计师、律师等非法人组织，行业协会等社会团体法人除外）；

（二）国有企业；事业单位；军队单位；成立三年以上的非外资（含港澳台）独资或控股企业；国内市场无类似或可替代产品的企业除外。

（三）具有良好的商业信誉和健全的财务会计制度；

（四）具有履行合同所必需的设施设备、专业技术能力、质量保证体系和固定的生产经营、服务场地；

（五）有依法缴纳税收和社会保障资金的良好记录；

（六）参加军队采购活动前3年内，在经营活动中没有受到刑事处罚或者责令停产停业、吊销许可证或者执照、较大数额罚款（200万元以上）等重大违法记录；

（七）未被中国政府采购网（www.ccgp.gov.cn）列入政府采购严重违法失信行为记录名单，未在军队采购网（www.plap.mil.cn）军队采购暂停名单处罚范围内或军队采购失信名单禁入处罚期和处罚范围内，以及未被“信用中国”（www.creditchina.gov.cn）列入严重失信主体名单或国家企业信用信息公示系统（www.gsxt.gov.cn）列入严重违法失信名单（处罚期内）；

（八）投标企业应当具备服务履约的能力；

（九）单位负责人为同一人或存在直接控股或管理关系的不同供应商，不得同时参加同一包的采购活动。生产场经营地址或注册登记地址为同一地址的不同生产型企业，股东和管理人员

（法定代表人、董事或监事）之间存在近亲属或相互占股等关联关系的不同非国有销售型企业，也不得同时参加同一包的采购活动。近亲属指夫妻、直系血亲、三代以内旁系血亲或近姻亲关系；

（十）法律、行政法规规定的其他条件。