

题目编号：XA-202620

面向政企场景的大模型智能体安全关键技术 研究比赛方案

一、发榜单位

企业全称：中国雄安集团数字城市科技有限公司

企业类型：国有企业

企业地址：中国(河北)自由贸易试验区雄安片区容城县雄安市民服务中心雄安集团办公楼 226

二、题目名称

面向政企场景的大模型智能体安全关键技术研究

三、题目介绍

(一) 题目背景

随着大模型在政务办公、知识服务、业务辅助、流程办理、智能运维等场景加速落地，应用形态已由传统问答式系统演进为具备检索、记忆、工具调用、文件处理、跨系统联动和自动执行能力的智能体系统。该类系统在提升效率的同时，也带来了区别于传统软件和普通大模型应用的新型安全风险：一是恶意提示注入、间接指令污染和知识投毒可影响模型决策；二是插件、脚本、Skill 等扩展组件可能引入供应链风险；三是高权限工具调用、文件读写、系统命令执行和外部接口联动可能导

致越权操作与数据泄露；四是缺乏面向运行过程的持续监测、审计溯源和分级处置能力，难以支撑政企场景对安全可控、过程可审计、责任可追溯的要求。当前行业普遍缺少兼顾“评测—防护—审计—运营”的一体化关键技术，难以支撑智能体在高价值业务场景中的规模化应用。

（二）目标介绍

（1）面向复杂输入链路的攻击识别与风险评估技术——构建覆盖用户输入、网页内容、文档附件、知识库检索结果、历史记忆等多源输入的攻击识别机制，实现对提示注入、间接指令污染、越狱诱导、数据投毒等风险的自动发现、关联分析和分级评估。

（2）面向工具调用和任务执行的安全约束与审批控制技术——构建智能体“感知—决策—调用—执行”全流程安全控制机制，围绕文件操作、系统命令、浏览器访问、接口调用等高风险动作，研究细粒度权限约束、动态策略校验、关键操作审批与异常任务链阻断技术。

（3）面向插件、Skill 与脚本生态的供应链安全检测技术——构建扩展组件安全分析框架，研究针对插件、脚本和 Skill 包的代码行为检测、依赖关系分析、恶意逻辑识别与安全评级方法，降低通过第三方组件引入后门、恶意下载、隐蔽外联和高危调用链的风险。

(4) 面向政企场景的智能体安全评测与审计溯源技术——建立覆盖数据安全、内容安全、执行安全、供应链安全和合规风险的多维度评测体系，形成可量化的测试方法、风险指标与审计证据输出机制，支持攻击复现、问题定位、效果验证和持续优化。

(三) 选题意义

本选题可面向政务办公助手、知识检索问答、业务流程办理、运维协同助手等典型场景开展验证，兼容 OpenClaw 类智能体及其他具备工具调用能力的大模型应用。成果可用于提升政企单位大模型应用的安全防护能力、评测验证能力和运行治理能力，为高风险场景下智能体安全落地提供技术支撑。

四、参赛对象

学生赛道。2026 年 6 月 1 日以前正式注册的国内全日制非成人教育的普通高等学校在校专科生、本科生、硕士和博士研究生（不含在职研究生），以及全日制职业教育本科、高职高专在校学生，可通过学生赛道申报作品参赛。

五、答题要求

参赛选手须围绕上述一项或多项关键技术方向，提交完整的技术解决方案，具体包括：

1. 技术方案报告（PDF 格式，原则上不超过 30 页）。包含问题分析、技术路线、算法设计、实验方案、指标体系及预期效果。

2. 原型系统或核心算法代码。须可复现关键技术验证结果，并提供运行说明、部署说明或代码仓库链接。

3. 演示视频（原则上不超过 10 分钟）。展示原型系统核心功能、关键技术流程及测试效果。

4. 可选补充材料。包括测试数据说明、评测脚本、攻击样例、审计日志样例等，用于支撑方案可验证性。

六、作品评选标准

作品评审采用百分制，具体评审维度如下：

1. 技术创新性：25%。评审技术路线是否具有原创性和前沿性，是否提出面向智能体安全问题的新方法、新架构或新机制。

2. 实际效果：30%。评审原型系统或算法在真实或模拟场景中的表现，包括攻击识别准确率、误报漏报情况、风险阻断效果、审计溯源完整性等量化指标。

3. 方案完整性：20%。评审技术方案的系统性、逻辑性和可行性，实验设计是否合理，是否充分考虑政企场景约束和工程落地条件。

4. 应用价值：20%。评审技术成果的落地可行性、可推广性及对提升大模型智能体安全防护和治理水平的实际价值。

5. 展示表达：5%。评审方案报告、演示视频和现场答辩的清晰度、准确性与条理性。

七、作品提交时间

2026 年 5 月至 9 月上旬，各参赛团队选择榜单中的题目开展研发攻关，各高校、企业、科研机构等组织协调机构应组织学生和青年科技工作者参赛，安排专业人员给予指导，为参赛团队提供支持保障。

2026 年 9 月 15 日前，各参赛团队要向发榜单位完成作品提交，具体要求详见本方案第八点第（二）款，并严格遵照发榜单位明确的提交规范执行。

2026 年 9 月 30 日前，由发榜单位完成初审，确定入围终审擂台赛的晋级作品和团队。

2026 年 10 月，发榜单位安排专门团队提供帮助和指导，各晋级团队完善作品。

2026 年 11 月，组织终审擂台赛，角逐“擂主”。

八、参赛报名及作品提交方式

（一）报名方式

（1）参赛选手登录“挑战杯”官网 www.tiaozhanbei.net，在“揭榜挂帅”擂台赛报名入口注册账号，登录大赛申报系统在线填写报名信息。报名信息提交后，下载打印系统生成的报名表。

（2）申报人在报名表对应位置加盖所在学校或所在单位公章。

（3）将盖章版报名表扫描件上传至报名系统，等待系统审核。请参赛选手注意查看审核状态，如审核不通过，需重新提交。

(4) 系统开放报名时间为 2026 年 5 月 30 日—6 月 30 日，逾期后系统将自动关闭报名功能。

(二) 作品提交方式

参赛团队请按以下要求向本单位提交作品：

(1) 提交内容：技术方案报告（PDF）、代码仓库链接、演示视频（不超过 10 分钟），以及 1 份经报名系统审核通过的参赛报名表。报名表所有信息须与系统内填报内容完全一致。

(2) 提交邮箱：caoruyue@chinaxiongan.com.cn

(3) 邮件命名规范：[揭榜挂帅·心理智能体]+学校+团队负责人姓名。

(4) 截止时间：2026 年 9 月 15 日 24:00。

九、赛事保障

1. 出题单位拟提供典型政企场景问题说明、应用背景资料和榜题解读材料，帮助参赛团队准确理解攻关目标。

2. 视比赛组织情况，拟开放部分模拟测试场景、风险样例或评测指标说明，支持参赛团队开展技术验证。

3. 拟组织技术咨询和阶段交流，对榜题边界、场景需求、作品提交要求等进行答疑说明。

4. 对优秀作品，出题单位可结合实际需求推动进一步测试验证、成果转化或场景对接。

十、设奖情况及奖励措施

1. 设奖情况

只设学生赛道，擂主 1 个，特、一、二、三等奖各 5 个。

2. 奖励措施

擂主：10 万元

特等奖：1 万元

一等奖：0.5 万元

二等奖：0.2 万元

三等奖：0.1 万元

3. 奖金发放方式

比赛结束后，单位比赛专班工作人员与获奖团队取得联系，填写奖金申请表，待获奖团队提供银行卡详细信息后 1 个季度内，将奖金一次性发放至获奖团队提供的银行卡中。

十一、比赛专班联系方式

1. 专家指导团队

顾问专家：赵文超 联系电话：15933016767

顾问专家：史诗 联系电话：19801791124

2. 赛事服务团队

联络专员：周倩老师，手机：18713555032

联络专员：杨志华老师，手机：19518319863

3. 联系时间

比赛期间工作日（9:00-17:00）

4. 申报联系人

姓名：曹如月，职务：科技创新组组长，联系电话：
18801060652

微信号：wxid_v2wif9di03o722

邮箱：caoruyue@chinaxiongan.com.cn

附：发榜单位简介

中国雄安集团数字城市科技有限公司成立于 2017 年 12 月，国有控股公司，与雄安新区同步诞生、同步成长。公司肩负将雄安打造成为“具有深度学习能力、全球领先的数字城市”的战略使命，以“全球领先的新型数字城市运营服务商”为愿景，坚持数字城市与现实城市同步规划、同步建设，让数字雄安“无处不在、处处可用”。

目前，公司控参股 11 家企业，在职正式员工 82 人，已构建以科技创新为内核，覆盖数字政府、数字民生、数字孪生、企业数字化转型、数字基础设施、网络安全等领域的基础业务生态。同时，前瞻布局人工智能+、数据要素、低空经济、空天信息、智慧生活等战略新兴赛道，为新区高质量发展持续注入新动能，致力于为全球未来城市建设提供“可感知、会思考、可进化、有温度”的数字雄安方案。公司承担科研课题 24 项，其中国家部委课题 9 项、省厅课题 10 项、新区课题 5 项，拥有国家高新技术企业、省级企业技术中心、河北省专精特新中小企业资质，具备 CMMI5、DCMM3、ITSS 二级、ISO27001、ISO20000 等完备资质，信息化与智能化建设能力行业领先。