

题目编号：XH-202609

具备自主决策能力的通用网络安全智能体 技术研究比赛方案

一、发榜单位

杭州安恒信息技术股份有限公司

二、题目名称

具备自主决策能力的通用网络安全智能体技术研究

三、题目介绍

为深入贯彻落实国务院《关于深入实施“人工智能+”行动的意见》（国发〔2025〕11号）部署要求，严格执行工业和信息化部、国家网信办、国家密码管理局等多部门2026年初联合印发的《人工智能安全治理专项行动方案（2026-2028年）》《网络安全产业高质量发展三年行动计划（2026-2028年）》中关于提升自主安全技术创新能力、构建智能化网络安全防御体系的核心任务，当前亟需推动AI智能体实现从感知理解向自主决策的关键跃迁。网络安全作为护航数字经济发展、保障国家“第五疆域”安全的战略基石，正面临攻击手段快速演进、防御体系智能化不足、专家经验难以规模化复用的核心挑战。

在此背景下，如何将前沿AI能力深度融入安全治理全流程，构建能够自主理解、决策与响应的新一代防御体系，已成为一项紧迫而关键的课题。为探索这一方向，本课题旨在引导

开发一套面向多变真实网络环境的自主决策通用智能体，能作为网络安全从业人员的智能增强伙伴，显著提升单兵作战与团队协同的效率与深度。该智能体需能理解复杂、开放式的安全任务，综合利用分析、推理与工具调用能力，独立完成从威胁感知、分析研判到响应处置的全闭环操作，以推动网络安全防御体系实现从自动化到智能化的根本性跃升。

为此，所构建的智能体需具备以下核心特性：

（一）自主任务理解能力：能够准确理解以自然语言、结构化文件、压缩文件、接口文档等多种常见形式描述的安全任务目标，自动解析任务需求并生成具备可操作性的执行计划。

（二）多场景自主决策与执行能力：能够适应渗透测试、应急响应、漏洞挖掘、逆向分析等各类网络安全实战场景，具备从态势分析、目标设定、任务规划到实时策略动态调整的闭环自主决策与执行能力。

（三）运行鲁棒性与决策可靠性：决策过程具备可解释性与可审计性，关键行动均有明确依据；能够在复杂的对抗与干扰环境中保持稳定运行，确保其行为逻辑具备较高的可复现性。

四、参赛对象

学生赛道：参赛对象为 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。

参赛对象可以团队或个人形式参赛，每个团队不超过 10 人，每件作品可由不超过 3 名指导教师进行指导。可以跨专业、跨学校、跨地域组队，但同一团队所有成员均应符合本赛道相关年龄、身份要求。每件作品只可由 1 所高等院校或科研院所作为参赛主体提交申报。

五、答题要求

参赛队伍应完成通用安全智能体研发，实现自主任务理解、多场景自主决策、自主执行等功能。

提交的作品形式应包括如下三部分内容：

1. 程序材料：包括但不限于智能体编译前的源代码、一键部署方案或完整部署手册、在线可测试的访问地址。

2. 文档材料：包括但不限于方案设计文档、开发文档、测试文档、用户手册、技术报告、方案介绍 PPT、阐述演示视频等。

3. 声明函：参赛方案原创性及保密性声明。

其他参赛者认为对参赛作品有辅助作用的材料均可作为附件提交，附件的质量和丰富度也会作为打分的参考之一。

六、作品评选标准

本选题初审由评审专家根据参赛团队提交作品材料进行综合打分，总分 100 分。作品评分标准具体如下：

评分 维度	分值	评分区间与标准描述
任务理解与执行设计	20 分	<p>（0-7分）：方案仅描述基础概念，任务理解与执行路径设计模糊，缺乏可操作性。</p> <p>（8-12分）：能理解常见安全任务，设计基本执行流程，但策略单一，对复杂场景考虑不足。</p> <p>（13-16分）：能清晰解析多模态任务输入，设计逻辑完整的闭环执行计划，具备场景适应性与策略调整思路。</p> <p>（17-20分）：展现出对复杂、开放式任务的深度理解，执行设计精巧，具备多策略选择和动态规划能力。</p>
系统架构与工程实现	20 分	<p>（0-7分）：架构混乱，代码与部署方案不完整，无法运行或极不稳定。</p> <p>（8-12分）：具备基本可运行的智能体架构，但代码质量一般，扩展性和鲁棒性较弱，部署复杂。</p> <p>（13-16分）：系统架构清晰合理，采用模块化/插件化设计，代码质量良好，部署方案可行，具备较好可维护性。</p> <p>（17-20分）：工程实现优秀，架构优雅健</p>

评分 维度	分值	评分区间与标准描述
		壮，代码规范安全，部署方案便捷高效，工具集成接口标准，可扩展性强。
决策逻辑与可解释性	20 分	<p>（0-7分）：决策过程为“黑盒”，缺乏解释与依据，行为不可预测。</p> <p>（8-12分）：能提供简单决策日志，但可解释性不足，逻辑链条模糊，复现困难。</p> <p>（13-16分）：决策过程有较清晰记录和推理链，关键行动有据可查，具备一定可审计性与鲁棒性。</p> <p>（17-20分）：具备完善的决策可解释性设计，行为依据明确，逻辑可追溯、可复现，能有效应对干扰。</p>
工具协同与扩展能力	20 分	<p>（0-7分）：工具调用方式原始，无协同设计，或无法有效利用工具。</p> <p>（8-12分）：工具调用为硬编码或简单拼接，集成度低，跨工具协同需大量人工干预，扩展新工具较为困难。</p> <p>（13-16分）：架构支持工具集成，具备插件化管理能力，能实现一定的工具串联与自动化。</p>

评分 维度	分值	评分区间与标准描述
		(17-20分)：采用高度插件化、模块化架构；工具接入规范、便捷，具备强大的跨工具链自动化编排与协同能力；扩展性极佳。
创新与 附加价值	20分	<p>(0-7分)：无明显创新，仅为现有技术的简单应用。</p> <p>(8-12分)：在某一环节有改进或特色设计，具备一定应用价值。</p> <p>(13-16分)：在核心能力上有明确创新，能带来效能的潜在提升。</p> <p>(17-20分)：提出并验证了具有重要价值的创新理念或技术，显著超越赛题基础要求，具备突出的实用或理论价值。</p>

本选题终审决赛由两个环节组成，总分 100 分，具体如下：

第一环节为人机协同实战赛，设置超出纯人工处理能力范围的赛题量，要求各团队以“3 名队员+自研 AI Agent”形式现场协同解题。为保障公平并聚焦工程能力评价，所有智能体需要在指定的受控环境中部署运行，且提前报备使用的模型 API（需为国内备案通过的大模型服务 API），通过主办方指定的 AI 安全网关在监控下接入相应的大模型 API。本环节为客观分，

由比赛平台自动判分，本环节分数占总分的 60%；

第二环节为线下集中答辩，参赛团队须依据作品评分标准（同上述初审评分标准），就系统设计、技术创新与工程实现等内容进行陈述并回答评委提问，现场答辩成绩占总分的 40%，每支队伍的最终总分由上述两环节成绩加权得出。

七、作品提交时间

2026 年 5 月至 9 月上旬，各高校组织学生参赛，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 5 日前，各参赛团队完成作品提交，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 20 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

请已在官网报名成功的团队，于 9 月 5 日前将盖章的参赛申报表 pdf、作品所有相关材料发送至发榜单位邮箱 edu@dbappsecurity.com.cn。参赛团队需将作品相关材料上传至百度云盘，并将云盘链接和提取码、以及云盘文件截图（含上传时间）打包发送至邮箱。邮件主题请严格按照格式“申报人所在单位－申报人姓名－作品名称－联系电话”填写。（例如：XX 大学－张 XX－XX 方案－手机号）提交具体作品时，务必一并提交 1 份报名系统中审核通过的参赛报名表（所有信息与系统中填报信息保持严格一致）。以上材料无需在“挑战杯”官网提交。

九、赛事保障

1. 成立专门指导人员

本单位将为此次比赛组建专家指导团队，介绍技术细节要求、定期解答疑问。参赛团队可在参赛期间的工作时段通过电

话进行咨询。

2. 终审擂台赛阶段练习测试平台提供

为确保入围终审决赛的各参赛团队能够充分备赛、优化智能体性能，本单位将在终审决赛晋级名单公布后，统一提供与终审决赛环境一致的仿真测试平台，并向每支入围决赛队伍开放累计 120 小时的测试时长。该环境支持参赛团队开展人机协同训练与系统调优，以更好地适应决赛阶段的实战场景与协同要求。

3. 企业参观实践：本单位在参赛团队完成相关审核等程序后可提供参观企业的机会。

十、设奖情况及奖励措施

1. 设奖情况

根据评分规则，综合评定参赛队伍成绩。设擂主 1 个（从特等奖中产生），特等奖 5 个，一等奖 6 个，二等奖 8 个，三等奖 10 个。奖项不重复，奖金按队伍所获最高奖项授予。

最终授奖数量视作品申报数量和质量情况，报组委会同意后动态调整。

2. 奖励措施

“擂主”：奖金 10 万元（税后）/个，并向团队全部成员优先提供实习实践、就业岗位、人才引进等机会。

特等奖：奖金 2 万元（税后）/个，并向团队主要负责人（1 个）优先提供实习实践机会、就业岗位机会。

一等奖：奖金 1 万元（税后）/个，并向团队主要负责人（1 个）优先提供实习实践机会、就业岗位机会。

二等奖：奖金 0.5 万元（税后）/个，并向团队主要负责人（1 个）优先提供实习实践机会、就业岗位机会。

三等奖：奖金 0.2 万元（税后）/个，并向团队主要负责人（1 个）优先提供实习实践机会、就业岗位机会。

全部获奖团队中应届毕业生参与杭州安恒信息技术股份有限公司招聘时，符合应聘条件者，直接进入面试环节，同等条件下可优先录用。

3. 奖金发放方式

以上奖金以汇款方式兑现。比赛结束后，本单位比赛专班工作人员将与获奖团队取得联系，获奖团队填写奖金申请表，提供银行卡详细信息并经公司审批后 3 个月内，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

为积极推进和更好筹备本次赛事，杭州安恒信息技术股份有限公司将成立比赛专班：

1. 专家指导团队

顾问专家：丁老师，联系电话：17557280625

顾问专家：叶老师，联系电话：15958032775

负责比赛期间技术指导保障。

2. 赛事服务团队

联络专员：王老师，联系电话：15068862417

联络专员：陈老师，联系电话：13020977666

负责比赛期间组织服务及后期相关赛务协调联络。

3. 联系时间

比赛期间工作日（9:00-17:00）

附：发榜单位简介

杭州安恒信息技术股份有限公司（简称“安恒信息”）成立于 2007 年，2019 年科创板上市（股票代码：688023），注册资本 10178.447900 万，是国内网络安全、数据安全和数据要素领军企业之一。现有员工近 4000 人，在全国设有 2 大总部，6 大产业基地，30 多个分公司及办事处，拥有数百位全国一线的核心安全专家以及具有创新力和自主知识产权的网络安全产品线，服务客户超 10 万家，公司 2025 年营收超 21 亿元。

公司坚持研发驱动，技术人员占比约为 65%，每年研发投入约占营收 30%。截至 2025 年底，累计申请专利 3049 项，主导/参与国家标准 70 余项。安恒信息以 AI 驱动服务革新，业内首发“恒脑”安全大模型及国内首个安全智能体，荣获全国人工智能应用场景创新挑战赛特等奖，并入选世界互联网大会领先科技成果集。

依托国家地方联合工程研究中心及国家级博士后科研工作站，公司承担国家级、省市级科研项目 50 余项。安恒信息屡获全球网络安全创新 500 强、中国品牌 500 强、卓越上市公司、全国数字贸易百强等殊荣，并获评国家高新技术企业、国家级核心安保单位、科创板创新 30 强企业、浙江省科技领军企业、浙江省产教融合试点企业。安恒始终致力于通过“AI+安全”双引擎，构建校企协同育人新生态，赋能数字经济安全高质量发展。