

研究助理岗位招募

云基础设施 机密计算与可信人工智能

西交利物浦大学物联网学院

西交利物浦大学物联网学院现面向四个互补技术方向招募研究助理（Research Assistant, RA）。工作将支持可信人工智能与网络安全韧性中心（Trustworthy AI and Cyber Resilience Center）的发展，以及面向产学研共创、校内项目实践与科研验证的大学沙箱建设。

研究团队由黄迪江教授（项目负责人 PI）和董彧佶博士（共同项目负责人 Co-PI）带领。研究助理将与学术导师、技术导师合作，并在约定项目中与企业工程师共同开展研发。

工作目标

我们重视可实际使用的系统。西浦沙箱将结合云原生基础设施、受控的数据与模型服务、硬件支持的机密计算，以及具备可重复安全评测能力的 AI 应用，为研发和测试提供受控环境。成果进入企业生产环境前，须经过相应验证并获得企业授权。

选择主要申请方向

- A 云原生架构与沙箱工程
- B 机密计算软硬件集成
- C AI 多智能体应用与模型后训练
- D 可信人工智能与网络防御工程

申请者请选择一个主要方向，可同时注明一个次要兴趣方向。我们重视所选方向的技术深度和实际实现能力，不要求任何申请者同时掌握四个方向或下文列出的所有工具。方向 A 需要已有基础设施部署经验；其他方向可通过扎实的相关项目证明胜任能力。

共同要求

- 通过课程、科研、开源或工程项目证明编程与调试能力，能够理解并解释自己提交的代码。
- 能够使用 Git、编写自动化测试与技术文档，阅读英文技术资料，并清晰说明进展、技术选择和阻碍。
- 负责任地处理保密信息，仅使用已授权的数据、凭据、计算资源和安全测试环境。
- 能够负责边界明确的模块，对重要决策主动寻求评审，并与基础设施、应用和安全方向成员协作。

方向 A 云原生架构与沙箱工程

在资深技术人员评审和指导下，设计并实现西浦私有云沙箱，负责范围明确的基础设施技术栈。本方向需要实际部署与排障能力，不能仅停留在架构图和方案说明。

主要职责

- 将科研与企业需求转化为模块化架构，覆盖计算、网络、存储、身份与应用部署，记录备选方案及运行维护方面的取舍。
- 通过可重复配置和基础设施即代码部署 Linux 服务器或虚拟机及 Kubernetes 环境，并在指定硬件支持的前提下集成 GPU 调度。
- 实现项目命名空间、角色权限、网络策略、资源配额、机密信息服务集成，以及获准使用的数据集和模型存储；按需对接现有校内服务。
- 建立镜像与部署流水线，提供版本化配置、监控、日志、备份恢复和回滚能力，定位网络、存储与资源瓶颈。
- 为研究人员、实习生和企业协作者提供清晰的接入流程，形成部署包，并针对约定的私有云或公有云目标开展有限范围的可移植性测试。

必备能力与经历

- 具有 Linux 管理和 TCP/IP 网络实践经验，实际部署并调试过 Kubernetes 上的容器化应用。
- 理解存储、虚拟化、容器镜像、身份权限和服务暴露方式，能够使用 Python、Go 或 Shell 完成自动化。
- 具有超出照搬教程的环境建设与维护经历，能够说明一次故障诊断、恢复过程和相应文档。

优先考虑的经验

Helm、Terraform 或 OpenTofu、Ansible、GitOps、GPU 调度、Prometheus/Grafana、OpenStack，或 AWS/GCP/阿里云经验。掌握一种相关技术栈即可。

预期交付成果

可重复部署的沙箱代码库、架构与配置记录、权限和接入说明，以及隔离、备份恢复和回滚测试证据。初步验收应包括部署并使用一个约定的示例应用。

工作边界

RA 向校内 IT 和设施团队提供技术需求；电气施工、制冷设计、机房改造、采购审批及校级 IT 管理权限由相应专业人员或部门负责。

方向 B 机密计算软硬件集成

围绕硬件能力、固件配置、虚拟化、运行时服务和应用，实施并评测机密计算 workflows，为获准使用的数据与模型任务建立有证据支持的保护链。

主要职责

- 评估指定 CPU/GPU 平台、固件、驱动与运行时前提，部署前记录受支持配置及其安全限制。
- 在支持的硬件上部署机密虚拟机或机密容器，获取远程证明，并依据明确的工作负载及平台策略验证证明材料。
- 将工作负载身份和证明验证结果与密钥释放服务集成，使用成熟密码方案实现加密、凭据处理和产物保护，不自行设计密码算法。
- 梳理应用生命周期中敏感输入、密钥、中间结果与日志的流向，明确明文出现的位置及可信计算基的组成。
- 为远程证明、策略拒绝、密钥释放、日志和恢复编写正反向自动化测试，并与条件受控的基线比较代表性任务性能。

必备能力与经历

- 扎实理解操作系统、虚拟化、内存保护和基础网络，具备 Linux 调试能力。
- 能够使用 Python、C/C++、Go 或 Rust 编程，了解加密、数字签名、证书及密钥管理基础。
- 完成过有一定规模的系统或安全实现项目。部署过可信执行环境（TEE）者优先，也接受基础扎实且能复现技术实验的申请者。

优先考虑的经验

Intel TDX/SGX、AMD SEV-SNP、受支持的 GPU 机密计算、Confidential Containers、TPM 或密钥管理服务经验。深入掌握其中一个平台即可。

预期交付成果

受支持配置记录、可重复部署脚本、远程证明到密钥释放的演示、信任边界说明、负向测试证据，以及约定工作负载的实测性能报告。

工作边界

软硬件工作指配置、集成与验证，不包含芯片设计。涉及 GPU 时须分别验证 CPU 与 GPU 信任条件。机密计算本身不能消除恶意应用逻辑或经允许输出造成的泄漏。

方向 C AI 多智能体应用与模型后训练

基于已有大语言模型（LLM）构建面向实际任务的应用，并在有证据支持时对选定模型进行适配与微调。工作覆盖可靠工具调用、评测和部署，而不止于一个聊天演示。

主要职责

- 将约定的科研或企业流程转化为输入、输出和成功标准明确的应用，按需实现 API、后端服务及可用的交互界面。
- 开发包含工具调用、共享状态、任务路由、恢复和重要操作人工审批的智能体流程；只有相较简单基线确有改进时才采用多智能体。
- 建立检索增强生成（RAG）和数据连接能力，实施权限感知检索、来源标注及项目文档的受控处理。
- 整理已授权训练数据，在选定模型上实施监督微调或 LoRA/QLoRA 等参数高效适配，并记录数据集、模型与实验版本。
- 评测任务成功率、事实依据、工具调用错误、延迟与资源成本；使用独立留出数据比较基础模型、检索方案与微调方案，防止评测数据泄漏。
- 将应用和模型服务封装部署至沙箱，配套权限、凭据管理、执行追踪、回归测试和回退至已知版本的能力。

必备能力与经历

- 扎实的 Python、API 集成和软件调试能力，完成过 AI/机器学习应用或可复现的训练项目。
- 理解神经网络训练与推理、数据集处理和评测方法，能使用 PyTorch 或同类框架。
- 能够解释何时需要检索、工具调用、多智能体或微调，并通过测试验证设计，而不是只展示一次成功结果。

优先考虑的经验

Hugging Face Transformers/PEFT、模型服务框架、向量检索、LangGraph 或其他智能体框架、监督微调、偏好优化或 GPU 显存优化经验。欢迎采用同类工具。

预期交付成果

一个范围明确、可重复部署的应用，配套评测数据集和报告、基线对照及运维说明。若微调有明确必要性，还需提供版本化适配流水线及模型或适配器评测。

工作边界

重点是已有模型的后训练与应用开发，不要求从零预训练基础模型。RAG 不更新模型权重；采用多智能体本身不能证明准确性或可靠性更高。

方向 D 可信人工智能与网络防御工程

为 AI 应用及其支撑系统建设实用的保护与评测能力，同时分析模型和数据风险及传统网络安全风险，将发现转化为可重复测试和有证据支持的整改建议。

主要职责

- 为选定应用建立威胁模型，识别用户、模型提供方、数据所有者、智能体工具及基础设施管理员，明确资产、信任边界和允许的测试范围。
- 使用已授权数据集和受控实验评测提示注入、越权工具调用、敏感输出泄漏、不可靠回答及相关模型攻击风险。
- 通过代码审查、依赖与容器检查、权限测试及日志分析评估系统控制，复现问题并验证有明确边界的整改。
- 基于已批准的遥测数据建立检测和调查流程，包括事件关联与 AI 辅助研判，用已知测试案例评估检测质量和误报。
- 在沙箱开展事件响应与恢复演练，包括遏制、证据保存和恢复，报告中明确区分已观察问题、评测限制与未经测试的主张。

必备能力与经历

- 在应用或云安全、网络防御、系统安全、机器学习或 LLM 安全中的至少一个领域具备实践基础。
- 能够使用 Python 或其他合适语言编程、分析日志、复现故障，并解释安全结论的证据。
- 理解身份认证、权限控制、常见攻击面及实验评估，能够严格遵守授权与数据边界。

优先考虑的经验

威胁建模、SIEM/检测工程、事件响应、软件供应链安全、对抗机器学习、隐私评估、差分隐私、联邦学习或嵌入式/AIoT 安全经验。

预期交付成果

面向具体应用的威胁模型与测试矩阵、可复现防御测试脚本、问题及整改报告，以及一个经过验证的检测或恢复演示。说明已知限制，不声称全面安全或已获认证。

工作边界

安全测试仅限获准资产与约定沙箱范围。外部扫描、企业生产测试及敏感数据访问需另行授权；本岗位不承担全天候安全运营中心或企业整体合规负责人职责。

工作方式与申请

团队协作

- 团队提供科研方向与技术指导。RA 负责指定组件实现，参与代码评审，并用可重复证据展示结果。
- 云平台 and 硬件集成需要定期在西浦太仓校区现场开展；其他工作可协商灵活安排。申请者应说明可投入时间及现场工作的可行性。
- 在相关安排允许时，企业合作可包括校内联合开发与实习实践。学术指导、企业导师、数据与知识产权归属及访问权限按项目约定。
- 职责与初期里程碑依据方向、经验及实际资源共同确定，不要求单个 RA 独立完成整套生产平台的架构、部署和安全保障。

工程与科研规范

- 维护可评审代码、依赖和配置记录、测试及简明操作说明，区分演示、本地测试和实际部署证据。
- 可以使用 AI 编程工具辅助工作，但必须理解、测试并解释提交的代码。不得向未经批准的服务上传企业数据、凭据或未公开代码。
- 科研产出可包括技术报告、演示系统、开源贡献、论文或知识产权，具体遵循实际贡献、导师指导与项目协议。

申请方式

请将申请材料同时发送至以下联系人：

Jie.Song@xjtlu.edu.cn

- 个人简历，说明教育背景、相关课程与项目，注明主申请方向 A、B、C 或 D 及可选次要方向。
- 简短陈述一个由你实际参与构建的系统或应用，说明个人职责、解决的难题及与申请方向的匹配程度。
- 如有，请提供代码库、代码样例、部署记录或技术报告。涉及保密内容时，可提供脱敏样例或进行非公开讲解。
- 说明可投入时间、可能到岗日期及现场工作安排，并告知相关学习或工作义务。

建议邮件主题：RA 申请 — 方向 [A/B/C/D] — [姓名]

选拔方式

选拔可包括项目讲解，以及与申请方向相适应、范围明确的实践讨论或操作题。重点考察实现能力、调试过程、技术判断与清晰表达，不以熟悉所有列出的工具为要求。

聘任安排

聘任类别、招聘人数、期限、工时、薪酬及到岗日期由项目与学校聘任流程确认。本说明不构成特定薪酬、合同期限或实习名额的保证。