昆山杜克大学校企协同创新平台_挑战项目详情

合作方信息			
机构名称：	Stevin		
机构类型：	初创公司		
官网 / 官方账号链接：	https://stevin.app		
联系人：	Yuandong Zhang - Founder & CEO of Stevin		
项目信息			
项目类别：	科技创新类		
预期开启时间	2026年12月	预计结束时间	2027年3月

机构背景：

Stevin 是一个面向团队协作的开源 AI 智能体平台，支持不同模型与聊天平台，可接入 Slack、Lark 和飞书，并通过 MCP 与插件连接现有工具和工作流程。平台支持按成员分别管理记忆与权限，让团队能够共同使用 AI 助手。目前，产品处于小范围内测阶段，希望通过技术评测与用户研究，进一步改善任务表现和团队使用体验。

项目概述：

团队 AI 智能体评测与用户研究

本项目将与 Stevin 开展合作，探索共享 AI 智能体在真实团队协作中的表现与使用体验。学生可围绕技术评测或用户研究开展工作，检验智能体能否准确完成协作任务、合理处理不同成员的信息与权限，并了解非开发背景团队如何使用、信任和管理共同的 AI 助手。项目成果可包括评测任务集、自动化评测流程、模型比较报告或用户研究与设计建议，为产品迭代提供依据。参与过程中，学生将锻炼 AI 评测、数据分析、用户研究与技术协作能力，无需具备既有的 AI 智能体项目经验。

可探索方向：

- **团队任务评测与自动评分：**将真实协作场景转化为结构化测试，明确预期结果与评分标准，并建立个人记忆、信息共享与权限测试集。比较并校准大语言模型自动评分与人工评分，接入 Langfuse 开展自动评测，形成评分一致性分析及模型比较报告。任务集初步目标为100个以上场景，并比较至少两种模型。
- **非开发背景团队的使用研究：**面向3–5个 DKU 团队开展4–6周实地研究，结合体验记录、使用日志与访谈，了解信任建立、任务委派、个人与共享记忆管理，以及 AI 主动协助的使用体验。成果可包括匿名化研究资料、主题分析与产品设计建议，并整理为短篇研究论文或海报。

合作方支持与资源：

产品与测试环境支持： Stevin 可提供独立测试空间、产品及管理后台访问权限，以及覆盖项目使用的模型与 API 额度，支持学生在真实但可控的环境中开展项目。团队也将开放 Langfuse 运行记录与评测环境，便于学生进行追踪、分析和评测。

导师与技术支持： Stevin 创始团队与工程师可提供项目启动培训、每周答疑、代码审查和研究设计指导，并通过双周进度交流和结项展示支持项目推进。

其他说明：

适合的专业与背景：

方向一较适合计算机科学、数据科学、应用数学等专业，或具备 Python 基础、对大语言模型评测感兴趣的学生。

方向二较适合行为科学、心理学、媒体与艺术等专业，或对人机交互（HCI）和混合研究方法感兴趣的学生。

欢迎不同年级学生参与，无需具备 AI 智能体相关的既有项目经验。建议每个方向以 2–4 人的小组形式开展。

协作与项目推进方式：

项目开始阶段，Stevin 将组织启动 workshop，介绍平台架构、评测工具和研究方法。项目过程中，合作方将与学生团队保持双周沟通，并在结项阶段安排成果展示。学生团队也将在 Stevin 提供的共享工作空间中协作，直接体验多成员共享 AI 助手在真实协作场景中的使用方式。

本项目以任务和阶段性成果为导向，不设固定的每周投入时长要求。学生可根据团队分工和项目进度灵活安排时间，但需按计划完成项目任务并参与必要的进度交流。

涉及参与者的研究将在知情同意基础上开展，数据在分析前匿名化，并按需遵循 DKU 的研究伦理审批流程；学生不接触客户生产数据。

DKU Co-Innovate Platform Challenge Information

Partner Information			
Organization Name:	Stevin		
Organization Type:	Start-up		
Website / Official Account Link:	https://stevin.app		
Contact Person:	Yuandong Zhang - Founder & CEO of Stevin		
Project Information			
Project Category:	Technology Innovation		
Expected Start Date	December 2026	Expected End Date	March 2027
Partner Brief Introduction: Yuandong and his team are building Stevin, a multi-player agent harness think Claude Tag, but open source, model-agnostic and platform-agnostic. Stevin serves a whole team, with per-person memory and permissions. Stevin is integrated into instant message platforms (slack/lark/feishu), and can be easily integrated into existing tools and workflows with the MCP and plugin system.			

Project Overview:

Team AI Agents: Evaluation and User Research

Working with Stevin, students will explore how a shared AI agent performs in real collaboration scenarios and how teams experience working with it. Students may focus on technical evaluation or user research, examining whether the agent completes tasks correctly, handles individual information and permissions appropriately, and earns the trust of teams without a software development background. Outputs may include evaluation tasks, an automated testing pipeline, model comparisons or user research and design recommendations. The project offers experience in AI evaluation, data analysis, user research and technical collaboration. Prior experience with AI agents is not required.

Areas to explore:

- **Direction 1: Team-Agent Benchmark & Human-as-Judge Pipeline:** Students turn realistic team scenarios into structured evaluation tasks (e.g., someone in a group chat asks for a bug to be filed → Stevin files the GitHub issue and replies in the same thread), each with expected outcomes and a grading rubric. They also build a memory eval set that tests what Stevin should remember, share or keep private for each person. They then calibrate LLM-as-judge graders against human ratings and wire the whole suite into Langfuse, so it runs automatically on every model or prompt change.
  - **Deliverables:** a documented task set (target: 100+ scenarios) with rubrics; calibrated judge prompts with a human-agreement report; an automated Langfuse evaluation pipeline; and a benchmark report comparing at least two models.
- **Direction 2: How Non-Developer Teams Adopt a Shared AI Collaborator :**Students run a 4–6 week field study with 3–5 DKU teams (student organizations, research labs, administrative offices) that use Stevin in their daily chats. Using a mixed-methods design of diary study and experience sampling (EMA), usage logs and semi-structured interviews, they examine how trust forms, which tasks get delegated, how teams handle shared vs. personal memory and permissions, and when proactive behavior helps or annoys.
  - **Deliverables:** a study protocol (submitted to DKU IRB if required), an anonymized dataset with thematic analysis, design recommendations for Stevin, and a short research paper or poster.

Partner Support & Resources

Product and Compute Access: Stevin can provide full access to its product environment, including a dedicated sandbox workspace, product and admin console access, model and API credits covering project usage, as well as access to its Langfuse tracing and evaluation setup.
Founder and Engineer Mentorship: Stevin's founders and engineers can support the project through onboarding, weekly office hours, code review, and hands-on guidance on research design. The team will also support progress through bi-weekly check-ins and a final showcase.

Additional Notes:
Target Majors and Backgrounds:
Direction 1 is best suited to students in Computer Science, Data Science, Applied Mathematics, or those comfortable with Python and interested in LLM evaluation.
Direction 2 is best suited to students in Behavioral Science, Psychology, Media & Arts, or those interested in HCI and mixed-methods research.
Students from all years are welcome. No prior experience with AI-agent projects is required. An ideal team size is 2–4 students per direction.

Collaboration and Onboarding Format:
At the start of the project, Stevin will offer a kickoff workshop covering the platform architecture, evaluation tooling, and research methods. During the project, the team will hold bi-weekly check-ins with students and support a final showcase at the end. Students will also collaborate in a shared workspace provided by Stevin, giving them first-hand experience of how a multi-player AI collaborator works in practice.

This is a task-driven challenge with no fixed weekly time commitment. Students can manage their time flexibly based on team responsibilities and project progress, while ensuring that agreed deliverables and milestones are met and participating in necessary check-ins.

Participant research will use informed consent and anonymized data, following DKU

ethics-review procedures where required. Students will not access customer production data.