



第二届北京市大学生“人工智能+”创新大赛
THE 2ND BEIJING COLLEGE STUDENTS' AI+INNOVATION COMPETITION

智领新质 共创未来

第二届北京市大学生“人工智能+”创新大赛

赛题规则

赛题

“智语外呼”——语音外呼客服
数字员工实时语音交互与自主进化

京东科技信息技术有限公司

一、赛题背景

我国联络中心坐席规模约 440 万人，生态带动就业约 1000 万人，电话仍是高时效、强意图业务的主动触达主通道。2026 年京东全球科技探索者大会（JDD）发布“京东百灵数字员工”，其中外呼数字员工以“AI 销冠”定位，在某头部汽车品牌实现线索筛选、企微加粉、试驾邀约、门店排程全流程自动执行，效能提升 2.5 倍；在高客单电商挽回场景，挽单率达 10.41%，高于人工坐席 3—5 个百分点。行业侧实证同样充分：物业催缴外呼量提升 10 倍以上、人力成本下降 50% 以上，某社区物业费收缴率由 72% 提升至 94%；快递派送场景转人工量下降 70%，质检覆盖率由不足 1% 提升至 100%。但该场景仍存在三大痛点，其一，产业 AI 普遍“投入巨大、产出甚微”，多数应用停留在能力演示层面，无法交付可评估的业务结果；其二，真实外呼是“秒级延迟 + 随时打断 + 双讲抗噪 + 情绪对抗 + 合规红线”叠加的高难度交互，通用对话框架往往“能闲聊、不能办事”；其三，优秀坐席的经验沉淀在个人身上，难以转化为可复用、可自进化的数字资产。

本赛题把企业真实外呼业务搬进校园，用“能不能对结果负责”作为验收标尺，让学生在真实约束下完成语音识别与合成、实时音视频工程、任务型 Agent 编排、系统工具调用到自进化闭环的完整实践，为产业储备既懂算法、又能交付的复合型人才。

二、赛题考察点及技术创新

本赛题重点考察“实时语音交互、任务型 Agent 与自进化”三位一体的端到端系统能力，强调能否在真实业务场景中完成任务，而非单一算法指标。参赛方案应具备完整的实时语音交互能力，包括流式语音识别与端点检测、领域词表定制、自然语音合成、端到端时延优化、打断响应、回声消除、双讲抗噪及 WebRTC 弱网传输；能够实现意图识别、槽位填充、对话状态管理、长会话记忆、任务规划、幻觉抑制、合规话术校验、情绪安抚、转人工及异常兜底；支持通过 Function Calling 或 MCP 等方式调用名单、订单和工单系统，完成结构化结果回写，并保障调用的幂等性与一致性。同时，系统应具备全量会话转写、合规检测、指标看板、失败归因和知识缺口挖掘能力，并做到一键启动、日志完整、实验可复现，

能够对时延、成本和质量之间的权衡进行量化分析。

在此基础上，本赛题鼓励探索 RSI 自进化能力，将实时语音交互与持续优化闭环贯通：系统可自动生成测试用例和包含对抗性场景的模拟用户，结合规则评分与 LLM-as-Judge 构建可复现的评测基线，通过失败聚类与归因，自动优化提示词、业务话术、检索知识和工具调用策略，并借助版本管理、回归验证和安全回滚机制，确保进化后的整体指标优于原有基线。同时，方案还应注重泛化能力，避免对已知测试集过拟合，使其在未公开场景中仍能保持稳定表现。Agent 不仅要在真实外呼中“把事办成”，还要利用自身产生的对话数据持续提升下一版本的能力，这是从“AI 工具”迈向“数字员工”的关键一步，也是本赛题区别于常规语音对话赛题的核心创新点。

三、参赛要求

（一）**参赛对象**。北京地区普通高等学校在校本科生，不限制专业和年级；需以团队形式参赛，每支参赛队伍由 3—5 名队员组成，每支队伍的指导教师不多于 2 人。

（二）**能力建议**。鼓励跨专业组队，建议队伍整体至少覆盖以下 2 类能力：

- 大模型应用开发：Prompt 工程、Tool-use/Function Calling、RAG、Agent 编排
- 语音算法基础：ASR/TTS/VAD 原理，或具备调用与调优主流语音服务的能力
- Web 前端与实时音视频：WebRTC、JavaScript/TypeScript
- 后端与工程化：Python/Node.js、服务部署、日志与评测体系建设

（三）**作品边界**。参赛作品须由参赛队自主完成。允许使用开源框架与预训练模型，但须在技术方案书中标注来源与许可证；不得将成品化的“整机外呼机器人”商业产品直接用于完成核心任务。

四、赛程赛制

（一）**总体赛制**。本赛题按“高校校赛推荐 — 企业分赛场选拔 — 北京市线下决赛”三个阶段组织实施，遵循大赛总体时间安排。企业选拔赛不设淘汰环节，按最终得分排名。

阶段	组织形式	关键任务
赛题发布与赛前培训	线上 (具体时间另行通知)	官方发布赛题、开放资料包与答疑群，组织 2 次线上技术说明会
高校校赛推荐	校内选拔	各高校自行组织校内选拔并推荐不超过 5 支队伍参加企业赛
企业分赛场选拔	线上/线下结合	提交作品、系统演示、答辩、统一指标评测，产生一、二、三等奖

(二) 比赛形式。

企业分赛场采用“作品提交 + 系统演示 + 现场答辩 + 统一评测”四段式：

- 作品提交：按赛题组要求的截止时间及材料清单提交代码、技术方案书、演示视频与自测报告等资料。
- 系统演示：每队 3 分钟，在 WebRTC 网页端真实跑通“发起外呼—多轮语音交互—任务闭环—结果回写”全流程，评委可现场打断及提问。
- 现场答辩：约 7 分钟，评委就架构选型、时延优化、合规设计与评测方法提问。
- 统一指标评测：企业使用未公开测试集统一跑分，结果计入基础分主要权重

总决赛现场采用“限时开发 + 演示 + 答辩”三段式，具体比赛内容及要求后续公布。

(三) 提交材料清单。

1. 技术方案书：含系统架构、技术选型、端到端时延预算、评测方法与结果分析
2. 全部源代码与可运行环境说明：提供一键启动脚本或 Dockerfile，确保评委可在标准环境下复现
3. 演示视频：不超过 5 分钟，须包含真实跑通的语音对话录音
4. 自测报告：含自建评测集说明、原始日志，以及 RSI 迭代前后的对比数据
5. 开源组件与预训练模型清单：逐项标注名称、版本、来源与许可证

(四) 赛前培训。

企业提供 2 次线上集中技术培训，并开放赛期专属答疑群，在合规范围内提供队伍参赛所需的技术资料与评测口径说明。

1. 培训时间与形式：赛题发布后组织第 1 次线上技术培训，企业分赛场作品提交前组织第 2 次；均采用线上直播并提供录播回放，全体报名队伍均可参加。

2. 培训内容：第 1 次集中培训讲解赛题目标、参考架构与技术选型建议；第 2 次集中培训讲解统一评测口径与指标测量方法，并集中答疑。

3. 答疑群与资料：自赛题发布之日起建立赛期专属答疑群，由赛题技术委员会轮值答疑；企业按"硬件或软件要求（三）企业提供"所列清单，在赛前统一向所有符合条件队伍发放脱敏资料包、Mock 接口与沙箱账号、评测口径说明等队伍参赛所需的资料。

五、赛事规则与评分标准

（一）总体规则。

“企业选拔任务”考察“能不能把语音外呼数字员工做出来并跑通”。企业选拔任务先行开展线上预选，参赛队伍完成任务后须提交技术方案书、演示视频及测试集录音音频，由评审团队进行评议，形成参与企业线下选拔队伍名单，最终形成推荐进入总决赛的队伍名单。

（二）企业选拔任务。

任务描述：基于 WebRTC 网页端构建一套完整的语音外呼客服数字员工，在给定业务场景下完成真实外呼任务闭环。

给定场景：服务类外呼场景，报名完成后公布详细背景资料及业务场景细节。

基础完成门槛（不满足者不进入排名）：

1. 浏览器端可与 Agent 完成不少于 3 轮连续实时语音对话，端到端首字响应时延 P50 不超过 2 秒

2. 至少完成 1 个完整外呼任务闭环：开场自报身份 → 身份确认 → 需求确认 → 方案或权益沟通 → 结果确认 → 结构化结果回写

3. 无严重合规违规（辱骂歧视、虚假承诺、诱导、泄露他人隐私、使用禁用话术）

分数构成与评价标准：

序号	评分维度	主要评价点
1	实时语音	打断响应与响应时延（P50/P95）、VAD 准确度、

序号	评分维度	主要评价点
	交互质量	双讲与抗噪能力、语音自然度与可懂度
2	任务完成能力	意图识别准确率、槽位完整率、任务闭环完成率、多轮纠错与主动追问能力
3	业务知识 与工具调用	知识问答准确率与可溯源性、工具调用正确率、结构化回写完整度与格式合规性
4	合规与全量质检	必说话术合规率、禁语拦截率、质检报告的完整度与可用性
5	工程质量与创新性	架构合理性与可复现性、时延与成本控制、技术亮点与创新
6	跨学科协作 与团队分工	团队专业背景覆盖度与分工合理性、技术方案书中协作机制与成员贡献说明、答辩环节对跨领域协同与决策取舍的阐述质量

可使用的技术工具：开发语言与框架不限（推荐 Python 或 Node.js）；ASR、TTS、大模型均可使用开源方案或云端 API（费用自理）；企业提供脱敏业务资料、Mock 业务系统接口与评测口径。数据边界：所有业务数据均为脱敏或合成数据，参赛队不得使用真实用户个人信息，不得自行爬取数据或向真实号码发起外呼。

（三）作品内容限制。

1. 允许使用开源代码、开源框架与预训练模型，须在技术方案书中标注来源与许可证；不允许直接使用成品化商业“外呼机器人”产品完成核心任务。
2. 禁止硬编码评测集答案或针对已知测试用例做特判，评审将使用未公开测试集，并在现场随机抽测。
3. 成果涉及语音、文本等内容时，须符合知识产权、数据安全、个人信息保护、AI 生成合成内容标识及公序良俗要求，不得包含涉密信息。
4. 赛后须提交全部源代码以供查重，查重不通过者取消成绩。

（四）排名与同分处理。

按综合得分由高到低排序。同分情况下依次比较：① RSI 进化有效性得分；② 实时语音交互质量得分；③ 现场加试一轮对抗性场景表现。

六、设备要求及资源支持

（一）参赛队自备。

- 开发用电脑：建议 8 核 CPU / 16GB 内存以上。

- 音频设备：带降噪麦克风的耳麦或耳机。
- 算力与额度：所选 ASR/TTS/大模型服务的本地推理显卡或云端 API 额度（费用自理）。

（二）软件环境

- 浏览器：须支持 WebRTC 与麦克风权限
- 推荐技术栈（非强制）：Python 或 Node.js 服务端；WebRTC + JavaScript/TypeScript 前端；Whisper/FunASR 类 ASR、CosyVoice/GPT-SoVITS 类 TTS，或等价的云服务；任意开源或商业大模型
- 不指定统一平台与数据集。

（三）企业提供支持

- 脱敏业务场景资料包：话术规范、业务规则、常见问题与合规红线说明
- Mock 业务系统接口与沙箱账号：支持名单查询、订单查询、工单登记等模拟调用
- 统一评测用例集与评测口径说明（仅在评测阶段开放）
- 统一时延测量工具与指标定义文档
- 参考架构说明与技术文档

（四）声学 & 拾音要求

- 演示区背景噪声不高于 50 dB
- 赛场统一提供标准测试噪声源（街道噪声、车载噪声、多人背景声），用于抗噪与双讲能力测试
- 赛场统一提供带降噪麦克风的耳麦供演示使用；参赛队可自备同类型设备，须赛前向赛场报备

七、安全及风险提示

（一）数据安全。赛题提供的业务资料、名单与对话数据均为脱敏或合成数据，仅限本次大赛使用，不得外传、不得用于任何商业用途；现场测评数据赛后统一回收销毁。

（二）内容安全。Agent 输出须受合规话术约束，禁止生成辱骂、虚假承诺、诱导及违法违规内容；一经发现，当场扣减该项得分，情节严重者取消参赛资格。

（三）**操作安全**。使用云端大模型或语音 API 时须做好限流与费用控制；现场严禁私自搭建外网穿透、代理，或进行大规模并发外呼。

（四）**合规与伦理**。遵守《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》《个人信息保护法》及相关安宁权规定；演示内容须尊重文化差异与公序良俗。

（五）**现场风险**。长时间语音演示请佩戴耳机并控制音量，避免听力损伤；比赛期间如出现系统故障、硬件异常或网络中断，应立即告知现场技术负责人协同处置，不得自行拆装赛场设备。

八、附加说明

（一）**激励奖励**。本赛题设一、二、三等奖及优秀奖，获奖队伍可获得京东科技提供的现金或实物奖励、京东实习与校招绿色通道、京东百灵数字员工平台技术认证证书；优秀项目可进入京东内部孵化对接通道。

（二）**知识产权及成果转化**。参赛作品的知识产权归参赛队伍所有，京东科技享有优先合作与成果转化权；赛后 30 日内组织优秀作品与京东业务部门对接，推动优质创新成果在企业真实场景中落地验证。企业提供的背景资料、数据与接口仅限本次大赛使用，不得外传。

（四）**其他**。本赛题的参赛对象、组队要求、评审纪律等与大赛总体要求保持一致。本说明未尽事宜，由赛题技术委员会与大赛组委会协商确定。本规则如与大赛组委会的其它规定不一致时，以大赛组委会规定为准。

“智语外呼”一语音外呼客服数字员工实时语音交互与自主进化

赛题答疑群

