

让产品每次升级，都有一套持续保障。

把 App、后台和 AI 已有的记录接起来，让 Agent 持续测试、检查内容、调查故障；问题有人接，修复有验证。

这次要补上的

从各模块自己的检查，走向覆盖产品全流程的质量与运行保障。

团队怎样配合

Bear、Ice 保留原有工作方式；我们承担接入、检查、分析和复验，Gary 统筹产品与项目。

读到最后你会知道

现在哪里不足、要建什么、先做哪些、以后每天如何工作。

- 01 现在有什么？
已有基础与实际缺口
- 02 要建什么？
一张图看懂体系全貌
- 03 怎么一步步建？
阶段、依赖与交付
- 04 建成后怎样？
三种日常工作场景

01 产品已经在运行，检查能力还比较分散。

我们不是从零开始。现有记录、校验和告警能提供基础，但还不足以持续证明：整个产品可用，内容可靠，修复确实有效。

● 已有基础 ● 需要补齐 ● 按条件扩展

ICE / APP

用户看到与操作的产品

已有 App 和安装包入口。当前测试包、连接环境与自动验收能力仍需确认。

BEAR / 业务后台

内容与业务记录

新闻、问答结果、用户反馈、版本与运营规则，已能在 Admin 查看。

NADA / AI 模块

生成与判断证据

已有消息、模型调用、翻译审计及部分校验、回测和告警能力。

运行基础：外部数据接口、服务器、数据库、存储与网络，共同支撑上面的产品。

已验证一部分连接：后台问答能关联 AI 消息；新闻能关联翻译任务。还不是 App 全链路验收。

质量缺口

成功记录不等于内容正确；持续语义质检、共享测试环境、跨端验收和修复复验仍需补齐。

运行缺口

现有日志与告警需要集中关联；备份恢复、安全覆盖和受控故障处理需要逐项建立证据。

为什么现在做：目前真实用户反馈样本有限，不能等用户替我们发现问题。测试产生可控场景，质检持续检查真实产出。

现状依据 2026-10-16 的只读探查与已有项目资料；不代表对所有服务当前状态的重新核验。

02 建设一层覆盖整个产品的保障体系。

上接 App、后台、AI 和基础设施；中间由测试、质检与运维 Agent 工作；下接问题处理、独立复验和人的决策。测试环境、权限和证据支撑全程。

InsigtX 质量与运行保障·目标全景

这是拟建体系。图中表示职责与信息流，不代表各模块均已接通。



箭头表示证据和工作流转，不代表自动获得发布权限。只有经过授权、边界明确的操作，才逐步开放自动执行。

Gary：产品与项目统筹

与 Max 对齐方向，梳理需求、输出原型，确认产品预期；协调优先级、预算权限、跨端事项和发布决策。

Nada：建设检查与运维能力

负责 AI 与本体系的接入、执行、分析和复验。用统一证据减少各方反复定位问题的成本。

Bear、Ice：保留原工作方式

提供必要权限和一次性交接材料；初期沿用原有修复发布方式，后续可审核 Agent 提交的修复 PR。

这套体系的价值：检查覆盖整个产品，问题能跨端定位，修复必须验证；不再仅依靠某一个模块的“运行成功”。

03 先接起来，再跑通，最后逐步自动处理。

每阶段都交付能工作的能力，并验证它是否有效。测试环境提前准备；基础安全和备份不等待用户增长。

阶段 01 / 先能持续检查

接入与只读质检

连接已有记录，校准质检员，形成统一问题清单；盘点安全与备份。

需要配合

必要只读权限与证据；Gary 确认预期，Nada 建设。

完成标志

问答与新闻可检查、可举证、可追踪；已知错误能检出，正常结果不强行报错。

阶段 02 / 再能跨端验收

隔离环境与 App

独立小型测试服务器、测试包及身份；接通后台、AI 和 App 的完整流程。

需要配合

Bear 交接部署材料，Ice 提供测试包，Gary 协调资源。

完成标志

隔离成立，版本可追溯；一个跨端问题完成发现、修复和独立复验。

阶段 03 / 形成日常保障

持续运行与恢复试点

定时检查、发布验收、运行监测和恢复演练；试点一项受控处置。

需要配合

明确处置对象、权限、停止条件和预算。

完成标志

计划任务有明确状态；恢复可验证，条件不足时正确停止。

阶段 04 / 随项目成长

扩大覆盖与自动化

增加模块与执行端，按风险抽样；有明确需求后评估容器及 K8s。

启动依据

覆盖缺口、检查积压、资源压力或重复维护负担。

完成标志

新增能力有量化收益、成本和回退依据，通过试点再扩大。

并行准备 → 阶段 01 开始时就准备阶段 02 的环境与材料。AI 自有样板可先推进；覆盖后台与 App 仍需要权限、测试包、验收预期、版本与费用条件。

多久、多少费用：先给量级，再用实测收准。

时间量级：一期 AI 自有样板约 3–5 周，范围为自有证据的只读接入、质检校准与最小问题闭环，不依赖 Bear / Ice 交接；以样本、标准、权限和试测额度获批为前提。二期依赖三方环境到位时间；环境、测试包、身份及其他验收条件齐备后，约 4–8 周。以上为方案估算，非交期承诺。

质检费用参考占位：按 200 条 / 天、每条 US\$0.005–0.03 计算，约 **US\$1–6 / 天**。单价尚未实测，先测小批成本，再定频率和额度；不含服务器、App 执行端、测试回答生成、搜索及额外复核重试费用。

本轮仍是方案体验稿，以上不是已批准的开工、采购或支出。

04 建成以后，检查会成为日常工作的一部分。

有没有新版本、有没有用户投诉，体系都在工作。Agent 承担持续检查与调查，人集中处理产品判断、修复审核和发布决策。

有新版本时

从“我测过了”到“这个版本有验收证据”

提交候选

自动整理变更与影响面

→执行测试

隔离环境验证实际行为

→独立验收

内容与功能分别给证据

→负责人决策

Gary 协调，核对发布版本

→发布后验证

确认版本和业务结果

结果：失败退回修复，没跑标受阻；验收通过不等于自动获准发布。

日常没有发版时

持续检查新闻、回答和运行记录

按计划取样

异常优先，也检查成功记录

→核查证据

原文、问题、回答与版本

→生成问题记录

给出片段、影响和依据

→负责人修复

沿用原有工作方式

→跨端复验

结果回到同一问题 ID

例如：新闻解析回答“没有查到此证据”，检查文章是否传入、检索是否执行、已有材料是否被使用；确实无证据时不强行判错。

出现故障时

从告警到调查，再到受控恢复

发现异常

监控、接口或用户反馈

→Agent 调查

关联上下游与近期变更

→核对权限

对象、条件与执行范围

→处理或升级

获准动作执行，其余交人

→验证恢复

业务恢复，不只进程存活

例如：D1 额度异常先定位查询和轮询消耗，不无限重试；生成中切出去后断流，要区分生成、传输和展示恢复，避免单方自报已修就关闭。

本轮请求确认：先让第一阶段跑起来。

建议本轮批准以下三项，以 AI 自有样板验证价值，同时为三方联调准备条件。

1. 一期开工

启动 AI 自有样板的只读接入、质检校准与最小问题闭环；不等待 Bear / Ice 的交接。

2. 小批质检费用额度

批准一次有费用上限的小批试测；执行前由 Gary 确认单批上限，先测算真实成本，再决定日常频率与额度。

3. 授权 Gary 协调一次性交接

协调 Bear / Ice 提供测试包、只读身份和部署材料，为隔离环境及跨端验收准备条件。

审批建议，尚未执行。此处不包含生产发布或自动运维处置授权；后续按明确范围另行决策。

[需要审查具体做法，继续查看实施手册 ↓](#)

从全景图，找到对应的落地办法。

主线读完即可理解方案。以下用于审查权限、步骤、证据、失败处理和验收目标。

评审增补：让三方交付与边界都可检查

三方质检矩阵：不仅检查 AI 回答

现有接入以只读为起点。以下是计划检查范围，不代表已经全部实现；公开案例只描述现象，不预判责任。

交付对象	检查内容	证据与验收
后台采集 → 入库 → 卡片	标题、正文、语言、来源链接、时效、图片和编辑残留	同一文章来源与入库记录可关联；裸 URL 占标题、错误链接或残留文本样本能被检出
后台基础设施与契约	字段或结构变化、凭据变化影响、回调可达性	变更影响清单及消费者回归；配置读者和依赖服务不遗漏
App 展示与交互	断流恢复、分享卡、引导问题、登录态、页面状态与版本兼容	测试设备截图与日志关联同次请求；生成、传输、展示结果分别验证
AI 回答、翻译与其他出口	事实与数字、语义、时效、历史使用和契约	来源、输出及校验记录；已知错误能检出，普通回答不退化；未接 App 的能力不冒充已接入

quality_check 是已有实现与部署记录的接入先例，但不能据此推断所有业务都已送检；按业务类型核对覆盖。接口成功不等于内容合格。

权限与落地

使用受限接口、已有记录与测试包建立各自用例；证据不足登记缺口，所需一次性改动由 Gary 协调原负责人。完整跨端用例必须在隔离环境执行。

边界契约：请求、回调、字段变更分别回归

对象与方向

Bear → AI 是请求入口；AI → Bear 是结果或回调；App ↔ Bear 是客户端业务交互。具体入口和字段以现行代码契约为准，不把“四个回调入口”作为已核实事实。

执行步骤

先清点请求与响应、异步回调、鉴权、ID、字段语义、状态与错误处理；用脱敏真实示例建立版本化样本。发布前在测试环境验证正常、缺字段、类型错误、新旧版本、超时、重复回调和失败重试。外发全部指向测试接收端。

变更与配合

字段、数据库结构、凭据或回调地址变更，记录生产者、消费者、受影响服务与生效版本。Agent 自动整理可读取差异，模块负责人确认无法推断的兼容要求；Gary 协调跨端安排。

验收与失败

请求被正确接收，返回与回调满足对应版本预期；重复操作不产生重复业务结果。破坏性变更未协调、必要消费者未复验时标受阻。不能在生产注入负例以凑齐证据。

跨端调查：先定位问题层，再确认修复归属

任一方发现异常，先建立统一问题 ID，再按文章、会话、消息或任务 ID 获取同次证据。区分生成、传输、入库、展示和恢复；无法关联时记录缺口，不急于定责。

确定问题层后由对应负责人修复；跨端争议交 Gary 协调。复验要覆盖发生问题的完整链路，同一问题保留修复版本、复验证据和发布状态；不能只凭某一方“已修复”关闭。

验收目标

一条跨端问题能从发现走到证据定位、修复及独立复验。无法定位保持待确认；无权限不扩大读取范围。复验后的案例加入相关模块回归集。

备份矩阵：现状先核实，恢复能力用演练证明

对象	当前证据边界	建设与验收
向量库及索引	历史资料存在备份待办；当前周期、数据量及有效性待核实	确定一致性快照和可重建范围，备份到隔离存储；恢复后检索与数据核对通过
运行数据库与必要文件	不能从部署前手动备份推断已有定期保障	使用数据库支持的一致性备份，记录保留与访问策略；隔离恢复并验证代表业务
D1 等托管数据	托管不自动等于满足恢复要求；平台能力与适用范围待核实	核对可用恢复和导出机制、保留窗口；在独立目标恢复，记录损失窗口与耗时
Bear MySQL	获取现有策略和执行记录；不把未知标成无备份	书面确认完成盘点；取得隔离恢复演练证据后，才确认恢复能力

基础盘点一期启动；生产备份写入、删除和恢复另列权限。恢复目标由 Gary 与技术负责人确认，未达目标不算完成。备份失败入单，不能删除唯一副本。

告警与容量：少打扰，但不能静默失明

复用历史告警机制，先核对当前任务、渠道、阈值、接收人和恢复通知。轻微预警进入质量 / 运维汇总报告；明确故障与恢复发送通知，不恢复已停用的早报 workflow。

检查对象	拟定处理	验收办法
回调、断供、域名、批次异常	复用现有记录，核对实际运行；同类聚合并提供恢复通知	隔离异常与恢复均可发现，不重复轰炸
备份任务失败	故障通知并入单；恢复成功后更新状态	模拟失败可通知，成功不能仅凭任务启动判断
质检短暂失败、测试环境不健康	记录预警和受阻用例；不能显示通过	失败和漏跑在执行账本可见

检查对象	拟定处理	验收办法
质检连续漏跑、整体失联或已错过验收窗口	升级为故障通知；外部探针监测检查系统	禁用调度后能发现失联；严重度不被永久压为预警
容量接近规划阈值	汇总给 Gary 做升级决策；影响业务时按故障处理	观察窗口、阈值及通知可解释，模拟越界可验证

容量如何定阈值

采集并发、CPU/内存、配额消耗、查询延迟和任务积压，记录日期、版本、采样窗口与业务条件。历史并发拐点、余量倍数或服务器规格仅作参考线索，本页不作为当前值引用。

先在隔离环境测量，再由技术负责人提出阈值、持续时间、回落条件和升级动作，Gary 协调投入。未填齐不能启用自动处置；先修重复调用与查询浪费，再判断是否扩容。

阶段验收与投入：不以“必须找出新 Bug”为目标

一期检查实际执行、覆盖与证据质量，用已确认历史或预置错误证明发现能力；没有新 Bug 不等于失败，也不能为了周报凑问题。安全和备份缺口逐项处理，高风险未解决项不得宣称核销。

AI 自有样板可先开工，三方覆盖仍依赖权限和交接。二期条件包括测试环境、测试包、身份、数据与外发隔离、版本、验收预期和费用授权，环境到位并非唯一前提。

时间量级：一期 AI 自有样板约 3–5 周，在自有样本、标准、权限及试测额度确认后推进，不依赖三方交接；二期环境与其他验收条件齐备后约 4–8 周，环境准备等待时间单列。以上是方案估算，不是交期承诺。质检费用按 200 条 / 天、单条 US\$0.005–0.03 占位，约 US\$1–6 / 天，尚未实测；服务器、执行端、测试生成、搜索及额外复核重试另计。未批准额度时付费执行保持关闭。

公开页用于审阅，只有脱敏方案；内部证据和可利用细节留受限记录。方案页按授权公网开放以便审阅；这不等于批准体系实施、采购或生产处置。

接入与交付

接入方式：开放必要权限，复用现有记录

已有基础

已验证 Bear 会话 ID 对应 AI conversation_id，message_id 对应保存的回答 ID；新闻 id 可关联翻译审计与 request_id。仅抽查验证，非全链路通过。

所需权限与配合

Bear 提供受限只读身份及测试部署材料；Ice 提供测试包和运行说明。凭据独立保管，不写报告。

执行步骤

由我们建设三个读取能力：列出新增记录、按 ID 读取证据、查询任务状态。优先已有接口，其次受限查询；使用水位、分页与去重，避免全表扫描。

输出证据

字段映射、权限范围、来源记录 ID、缺失证据清单；内部实施基线持续维护并接入 init。

验收目标

选定问答与新闻能由业务记录追溯到 AI 证据；权限不能写生产；缺失字段被明确标记。

失败处理

没有必要权限或资料时标记受阻。App 无法切环境、后台缺少部署材料时，登记一次性交接任务，不能把“已给权限”当成可用。

扩展条件

初期我们测试、诊断、报障和复验；开放仓库后可提交修复 PR，由原负责人审核。CI 权限齐备后再接入合并保护。

开发记录、独立验收与版本绑定

已有基础

复用现有测试、回测和验收框架；开发自测与独立验收分别留证。

所需权限与配合

候选代码或构建产物、需求和原型、必要测试权限。Gary 确认产品预期，模块负责人确认技术影响。

执行步骤

Agent 整理目标、差异、影响面、用例和回退方式。独立验收使用独立任务上下文和受限身份，亲自执行。报告绑定 App 构建及摘要、后台版本、AI 版本、配置与用例版本。发布前再次核对。

输出证据

开发记录、逐条验收结果、版本清单、执行日志；测试通过、发布成功和线上验证分开记录。

验收目标

必需项全通过才 GO；明确失败为 FAIL，未执行或缺证据为 BLOCKED。空报告、过期版本、缺失用例和伪造自测不能通过。

失败处理

失败返回对应负责人；版本变化后重评影响范围并复验。初期不把文档审批误称为已实施的 CI 强制闸。

扩展条件

先用人工执行的版本核对清单；仓库及 CI 权限开放后加入合并保护和构建验证。

Buglist：发现、归并、修复与复验

已有基础

Bear 已有反馈处理，AI 有错误与调用记录；统一问题库为拟建能力，不改原有反馈语义。

所需权限与配合

我们维护独立质量工作台；初期只读接入来源，按 ID 链接回证据。

执行步骤

事件与主问题分开存储。相同问题归并但保留实例。状态为待分析→待确认 / 已确认→修复中→待复验→待发布→线上验证关闭；复发重新打开。Bear、Ice 按原方式修复，我们复验。

输出证据

列表只放 ID、模块、摘要、影响、次数、状态、负责人和下一步；展开可见来源、版本、分析、修复与复验证据。

验收目标

一个真实问题走完完整闭环；缺证据或版本不符不能关闭。误报和重复项记录依据；代码修复、部署、业务恢复分别可核查。

失败处理

无人认领或跨模块争议交 Gary 协调；证据不足留待确认。Agent 不得擅自改卡、删数据或发布内容。

扩展条件

仓库权限开放后提交修复 PR；长期保留同一问题与版本关联，新增模块只补接入工具。

内容质检与问题处理

回答与历史上下文质检

已有基础

已有会话、消息、模型调用及部分判断证据；本次验证了记录关联，未证明每个出口证据齐全。

所需权限与配合

只读获取问题、必要前文、检索或工具来源、最终回答和版本；新的付费复核须有预算。

执行步骤

异常、负反馈和新版本产出优先，同时抽取正常回答。先检查证据完整性与数字等确定性规则，再判断是否答到重点、事实受支持、历史使用合理、是否遗漏明确有益补充。每项问题写出片段、标准和来源。

输出证据

按 message_id 关联的质检结果、检查版本、来源锚点、缺证据原因和耗时费用；不以总分代替依据。

验收目标

已确认错误案例能被识别并正确举证；无关历史不被要求强行使用；无来源时不得编造结论。普通回答、隐私和不同用户隔离纳入回归。

失败处理

来源缺失记无法核实；模型失败记执行失败；疑似问题进入复核，不自动宣称事实错误。敏感来源删除后不能从问题库恢复原文。

扩展条件

先覆盖对话和历史上下文。低流量时评估全量异步质检，费用与隐私条件不满足则显式抽样；增长后按风险和版本分层。

新闻、摘要与多语言内容质检

已有基础

信息流可见原文来源、正文和语言；选定文章关联到了多语言翻译审计。未找到该样本的独立 quality_check_items 记录，覆盖原因待核对。

所需权限与配合

读取原文、译文或摘要、语言与任务 ID；不使用后台补单、修改和发布权限。

执行步骤

先确认业务类型和质检覆盖。逐条核对原文与输出，执行金额、单位、主体和日期检查，再由 AI 检查省略、夸大、因果和译意。不同语言保留独立结果。

输出证据

文章 ID、处理批次、语言、原文证据、问题片段、规则命中与质检结论。

验收目标

指定金额错、主体错、译意变化案例可检出；正常摘要允许合理省略；生成成功或翻译 ok 不替代质量通过。

失败处理

原文不可得记无法核实；独立质检记录不存在时记录覆盖缺口，不能直接推断没有任何校验。

扩展条件

先接一个证据完整的新闻出口，再接巨鲸、InsightBar 等；每个新出口单独验证，不能继承其他出口的通过状态。

质检员校准、调度与费用保护

已有基础

已有固定案例可复用；新质检模型和规则本身需要验收。

所需权限与配合

脱敏案例、人工确认标签、获批小批测试费用；正式调度运行在独立服务身份下。

执行步骤

将正确、错误和证据不足样本分为调试集与留出集。按固定版本跑留出验收；调度按批次水位去重，加锁、超时、有限重试与心跳。记录实际模型账单，先小批测算再启用正式频率。

输出证据

逐类误报漏判、样本量、执行批次、覆盖率、待检量、漏跑、实际费用及估算差异。

验收目标

指定严重错误必须检出且证据正确；正常留出样本不得误判为确认严重问题；缺证据样本保持无法核实。模型或规则升级后重跑。费用与执行条件缺失时不得自动启动。

失败处理

未跑、执行失败、超预算均不得显示通过；重试仍消耗预算。检查器离线由独立探针发现，日志文本不作为 Agent 执行命令。

扩展条件

随着积压、成本和质量指标调整频率与抽样；只读采集和付费复核分别限额，不让检查系统拖累生产。

测试环境与 App

测试环境建设与隔离验收

已有基础

尚未核实独立远程测试环境。App 已有 APK 入口，但当前版本、可安装性及连接环境未确认。

所需权限与配合

Bear 提供可部署产物或代码、依赖与配置说明；Ice 提供指向测试环境的包；Gary 协调采购与权限。

执行步骤

先形成服务与资源清单，再采购部署。独立进程、数据库、Worker/D1/KV 等涉及资源、测试身份和数据。用脱敏固定案例；外发默认关闭或进入测试接收端。限定资源与模型费用，准备重置脚本。

输出证据

环境拓扑、三方版本清单、凭据权限检查、网络与回调目标清单、数据重置记录和隔离验证日志。

验收目标

测试请求及后台任务不能写生产；测试环境不持有生产写凭据；数据可重置、版本可识别、资源受限。检查跨环境错误配置会被拒绝。

失败处理

隔离无法证明则 BLOCKED，禁止用生产环境补跑破坏性场景。缺依赖由对应负责人一次性交接，不绕过边界。

扩展条件

后续需要并行版本时增加临时测试实例；不一开始要求团队迁移生产部署方式。

App 执行端与跨端证据

已有基础

已有后台版本页和问答 ID，但尚未安装运行 App，不把后台成功当成体验通过。

所需权限与配合

测试包、构建标识、测试身份、必要设备权限；客户端日志能按任务关联。

执行步骤

服务器调度独立设备或模拟器执行端。安装指定包，运行固定操作，保存截图、必要日志和关联 ID；核对 UI、后台、AI 最终内容与状态。先验证执行端能力再选择设备，不假定普通云服务器支持模拟器。

输出证据

包摘要、App/后台/AI 版本、操作记录、截图、消息 ID、测试数据与最终断言。

验收目标

新版本可重复安装和运行；执行端离线能识别；同次回答在三个环节可关联。任何一步缺证据均不算全链路通过。

失败处理

离线或操作不稳定记录执行失败与具体步骤，停止无限重试；截图脱敏，不导出正式用户信息。

扩展条件

先 Android 核心流程，再按实际支持平台增加设备矩阵与并发执行端。

运维、安全与成长

服务器稳定性

已有基础

已有服务管理、健康检查和排水记录，现状逐项盘点。

所需权限与配合

服务状态、资源与必要日志只读权限；演练仅在测试环境。

执行步骤

建立服务、依赖、版本、配置读者清单；采集存活、磁盘、内存和任务积压；测试停服和资源压力。

输出证据

服务清单、监测曲线、故障时间线和恢复记录。

验收目标

异常可发现并定位；恢复后代表业务通过，不能只看进程存活。

失败处理

采集失败记未知；禁止无条件重启，未授权处置只报告。

扩展条件

资源瓶颈或多实例需求明确后再评估扩容和编排。

网络稳定性

已有基础

复用已有域名监控与抖动抑制。

所需权限与配合

域名探针及网络错误摘要；独立于被监测服务的外部检查。

执行步骤

覆盖 DNS、TLS、连接错误和延迟；交叉判断单点抖动与服务不可达；监测调度器本身。

输出证据

探针结果、异常持续时间、去重与恢复通知。

验收目标

隔离模拟网络失败可识别，恢复通知正确；调度器失联不会静默。

失败处理

状态不足标未知，禁止把单点超时直接作为重启依据。

扩展条件

按用户地区和可用性目标增加探针位置。

数据接口稳定性

已有基础

已有回调告警、断供哨兵和任务状态。

所需权限与配合

接口结果、回调及入库时间的受限读取。

执行步骤

检查超时、错误率、回调失败、数据新鲜度和任务积压；关联上下游版本。

输出证据

请求及任务 ID、断供区间、依赖状态与调查结果。

验收目标

测试“HTTP 成功但没有新数据”仍能触发异常；重复告警归并。

失败处理

没有数据到达时区分无业务与异常；证据不足留待确认，不盲目补单。

扩展条件

新增数据源逐项注册预期频率和失败语义。

Bug 监控

已有基础

已有 Admin 错误与用户反馈入口。

所需权限与配合

受限读取日志和反馈；写入独立问题库。

执行步骤

接入异常、测试、内容质检和用户反馈，归并主问题并保留实例；按版本与模块分派。

输出证据

问题 ID、影响、实例、负责人、修复与复验记录。

验收目标

同类不反复建单，复发可重开；缺复验证据不能关闭。

失败处理

采集漏跑单独告警；跨端责任不明交 Gary 协调。

扩展条件

按新增模块补接入，不新建互不关联的问题列表。

防攻击

已有基础

已有历史鉴权和限流整改记录，不直接认定当前漏洞。

所需权限与配合

入口与权限配置只读检查；负例在隔离环境执行。

执行步骤

盘点公网入口、鉴权、管理权限和限流；验证未授权、越权及滥用边界；生产修复另获授权。

输出证据

入口清单、风险依据、负例和正常请求复验。

验收目标

未授权访问被拒绝，正常业务仍可用；历史风险逐条核销。

失败处理

高危发现及时升级，未经授权不在线上利用或扫描施压。

扩展条件

按实际风险增加入口防护，工程规则负责拦截。

防数据抓取

已有基础

已有部分访问限制，完整覆盖待核实。

所需权限与配合

读取访问规则和脱敏访问统计；测试账号。

执行步骤

划分开关与受保护内容；测试批量读取、分页上限、身份配额和异常访问；限制由代码或网关执行。

输出证据

数据边界、访问模式、规则与测试结果。

验收目标

受保护数据不可越权读，超限请求受控；正常浏览不误伤。

失败处理

公开内容不承诺绝对无法抓取；疑似异常先调查，不自动大范围封禁。

扩展条件

访问规模或滥用增加时细化策略并回归正常流量。

系统漏洞

已有基础

复用历史审计清单，修复及误判状态需按证据区分。

所需权限与配合

依赖清单、必要配置与权限只读信息；修复由模块负责人审核。

执行步骤

检查依赖漏洞、敏感信息暴露和权限；按影响排序；修复后重跑漏洞检查及业务回归。

输出证据

风险分类、版本、证据、修复及复验结论；细节仅内部保存。

验收目标

每项有处置状态和依据；旧报告不冒充当前漏洞。

失败处理

无法核实保持待确认；严重风险升级处理，不等扩展阶段。

扩展条件

依赖或配置变更触发检查，扩大覆盖须核算成本。

备份与恢复

已有基础

已有数据库和服务资产，备份有效性待盘点。

所需权限与配合

备份状态只读权限；隔离恢复目标；备份执行另列授权。

执行步骤

清点数据库、文件及必要配置；检查备份结果与访问控制；在测试环境恢复，核对数据和业务。

输出证据

备份对象、执行时间、恢复日志、实际数据损失窗口和恢复时长。

验收目标

隔离恢复后数据及代表业务通过；Gary 与技术负责人据实确认恢复目标，未达到目标不算完成。

失败处理

备份失败或恢复失败立即入单；不得删除唯一副本或向生产恢复演练数据。

扩展条件

随数据规模和业务恢复要求调整频率、保留策略与异地保存。

容量与费用

已有基础

D1 历史额度事故与已有模型账单是检查线索。

所需权限与配合

资源及费用统计的受限读取。

执行步骤

统计生产和检查系统自身消耗；定位高耗查询、重复轮询、重试与模型调用；先修浪费再评估升级。

输出证据

消耗趋势、费用来源、实际覆盖、待检积压和优化前后对照。

验收目标

模拟额度拒绝可控退出；预算不足明确降覆盖或停止，不隐藏失败。

失败处理

账单滞后标估算；无预算不启动付费任务，成本失控停止新增任务。

扩展条件

压力持续或检查积压增加时提交量化扩容方案。

受控处置与 K8s 演进

已有基础

现有服务命令、管理接口和脚本可作为接入基础；不因 Agent 接入要求生产迁移。

所需权限与配合

自动调查使用只读身份；每项生产执行单独定义对象、动作、时限和授权。

执行步骤

先选一项常见故障，固定前提、命令、重试次数、超时、恢复标准与停止条件。隔离演练成功后才申请白名单。工具限制执行能力，不能靠提示词代替权限。

输出证据

操作前后状态、授权、执行记录、业务恢复验证和失败升级；总开关可停止新动作。

验收目标

前提不满足、超时、证据过期、恢复失败均停止；不能循环重启或扩大权限。备份删除、密钥轮换、迁移等保持单独决策。

失败处理

异常停止并通知负责人，代码问题转 Buglist；不自动执行日志或页面里的命令。

扩展条件

多实例、滚动升级、扩缩容或维护负担明确时比较现有管理与容器/K8s，先试点再扩大；用户数本身不作为迁移理由。

投入与排期：先列依赖，再核算费用

已有基础

审批参考量级：一期 AI 自有样板约 3–5 周，不依赖 Bear / Ice 交接，以自有样本、标准、权限和试测额度确认作为开工前提；二期环境及其他验收条件齐备后约 4–8 周，准备等待时间单列。均为方案估算，非交期承诺。质检费用按 200 条 / 天、每条 US\$0.005–0.03 估算为 US\$1–6 / 天，单价未实测；先小批收准，不含服务器、执行端、测试生成、搜索及额外复核重试。

所需权限与配合

Gary 协调预算；各模块确认一次性交接内容与可交付时间。

执行步骤

分别列测试服务器/存储/执行端、日常质检与复核、测试生成与搜索三类预算。获批小批测算单条成本、重试和耗时，再定每日上限与频率。任务按可独立交付单元估时，依赖等待单列。

输出证据

资源清单、报价与有效期、实测账单、计划覆盖、每批/每日上限、负责人和阶段排期。

验收目标

付费执行前有额度；环境开通前有资源清单和采购批准；未估项明确待估，不按零成本处理。

失败处理

未获额度保持付费执行关闭；可继续只读盘点和本地规则验证。不因预算不足静默漏检。

扩展条件

每月复核实际费用与收益；重复查询或调用先优化，再提出扩容。

持续回顾与升级条件

已有基础

复用统一问题、版本与执行记录，不新增多套管理台账。

所需权限与配合

质量工作台的汇总读取；Gary 与技术负责人参与需要决策的回顾。

执行步骤

每月检查覆盖率、待检积压、误报漏判、问题复发、发现和恢复耗时、成本及人工介入。新模块登记证据与用例；规则升级重跑固定集。

输出证据

月度趋势、下一阶段候选任务、需要决策事项；更新内部实施基线及 init 索引。

验收目标

指标能追到执行记录，未跑不计通过；每次扩大范围有验收与回退方式。

失败处理

样本少不宣称高准确率；无数据的指标标未知。优先解决复发与高影响缺口。

扩展条件

检查积压增加时优化抽样或执行资源；反复人工操作先固化流程；明确编排需求后才评估平台迁移。

阅读方法：主页面讲清全貌，手册对应每块能力。本文不承诺杜绝错误，也不把接口成功、生成成功、单方自测或历史记录当作全链路通过。

InsigtX · v5 体验稿 · 2026-10-20

线上审阅材料；体系建设、采购与生产操作仍需按范围另行批准。