石蛤蟆 · 商务简介

产品名称：石蛤蟆（非结构化数据 AI 智能处理平台） **面向对象**：企事业单位（B 端） **合作模式**：项目型销售，一事一议、一项目一签单

网址：https://www.shihama.com.cn

一、一句话说清它是什么

企业内部 80% 的信息不是数据库里整齐的字段，而是**会议录音、合同扫描件、现场照片、监控视频、Excel 台账**——散落在每个人的电脑和手机里，找不着、读不完、用不上。

石蛤蟆做的就是一件事：把这些「沉睡的数据」吞进来，变成能用的成果。

三端协同：桌面端与移动端自动采集，管理端统一治理与创作，成果再分发到员工的日常场景。

二、为什么叫「石蛤蟆」

这个名字本身就是一句话——**吞进去的是数据，吐出来的是财宝。**

六百年的地标

山东曹县「石蛤蟆」是当地有六百多年历史的地标。据《曹县志》记载，元代曹县县城所在地为磐石镇，济水流经此地水流湍急。有一年济水冲来几块青石，其中一块形状略似蛙形，居民觉得灵异，便将其雕琢成石蛤蟆，放在岸边**镇禳河患**。明正统十一年（1446 年）曹县建城后，石蛤蟆被移到十字街口，这条街也因此得名「石蛤蟆街」。现存的石蛤蟆已是第四代。

两个深入人心的传说

- **镇水救民**：相传很久以前发大水，一只大蛤蟆从河里爬到街上，洪水涨它就涨，这条街始终安然无恙。水退后蛤蟆消失，街上留下一块形似蛤蟆的石头，大家便把它当作保护神。
- **扃金元宝**：老人们常说，石蛤蟆会在夜深人静时扃金元宝，留给每天起最早的人，懒人从来捡不着。所以石蛤蟆街周围几百年来日子过得红火的，都是勤快人家。

为什么这个名字寓意好

金蟾招财是流传千年的老传统，不少企业本身就会购买金蟾摆件求个好彩头。镇宅避灾、招财进宝，**无需多解释**：

传统寓意	石蛤蟆做的事
蟾能吞财	把企业散落的录音、扫描件、影像、台账**吞**进来
镇宅避灾	数据集中管理，权限到人、操作留痕，**防得住风险**
厨金元宝	识别、解析、治理后**产出**可用的成果，支撑获客与经营

进得来、留得住、能生钱——同一个道理，一听就懂。

三、三端能做什么

BS 管理端（12 个功能页面）

治理、检索、统计与成果制作的集中入口，也是唯一具备以下能力的终端：

- **成果制作**：选中素材 → 勾选成果类型 → 选提示词模板或自定义 → 提交任务队列。四类成果：识别文档、解析文档、网络推文、动漫短片
- **调阅权限**：四级控制在人——全部开放 / 有限分享（指定部门人员）/ 本人独享 / 未设置默认仅本人；叠加单位 → 部门 → 人员三级组织隔离
- **标签管理**：AI 自动标签 + 手工标签并存，形成企业自己的分类口径
- **统计分析**：九项汇总指标 + 按单位/部门/人员/类型多维统计，支持导出 Excel
- **日志审计**：下载日志与查询日志全留痕，满足内审与合规追溯

通用能力：大文件分片上传与 MD5 秒传、语音识别（自动区分发言人）、图像 OCR、全文检索（沿组织架构树导航）。

CS 桌面端（7 个功能页面）

Windows / MacOS / Linux 客户端，定位是**存量数据采集 + 结果查看**。

- **目录批量采集**：添加一个或多个文件夹，递归扫描 46 种文件格式，过滤后批量上传
- **三层去重闭环**：① 本地上传记忆（路径+大小+修改时间三元组，未变动即跳过）→ ② 服务端 MD5 秒传 → ③ 分片级断点续传（取消或失败的分片自动回队）
- 支持拖放或多选上传；处理结果查询（9 项条件）、详情查看、识别结果编辑、成果下载（PDF/Word/TXT）
- 上传成功后立即持久化记忆，进程意外退出也不会重复上传

关于自动化程度的如实说明：当前版本采用「手动触发扫描 + 手动启动上传」，尚未实现后台定时轮询或文件夹实时监视。其「重复扫描不重复上传」依靠上述三层去重机制实现。已上传队列后的**上传过程全自动、无需干预**，队列可能在数分钟、数小时、数天甚至数十天内持续跑完，中途可随时暂停、继续或重试。

MS 移动端（石蛤蟆 App）

App 端，定位是**增量采集与成果分发**。

- **后台自动上传**：Android（WorkManager）、iOS（BGProcessingTask）、鸿蒙（长时任务）三平台原生实现，注册一次后由操作系统调度，**员工零操作**
- **五重校验才启用**：设备标识 → 登录凭证 → 服务端配置 → 三重开关（向导完成/协议授权/总开关）→ 监控目录；每次后台唤醒都重新校验，不信任缓存
- **不影响员工**：仅 WiFi 上传、深夜时段优先（22:00–07:00）、三档限速、单文件大小上限
- **隐私合规**：不使用 IMEI，改用首次安装生成的随机 UUID；授权可随时撤销
- **内容分发**：成果以视频、推文、文档形式呈现，支持浏览、点赞、收藏、关注、两级评论、分享到微信

四、合作方式：项目制交付

以项目为单位报价与签约，单项目金额按范围与复杂度核定。

交付流程

阶段	内容
① **需求对接**	梳理业务场景与数据现状，明确本项目要达成的目标
② **标准功能落地**	部署三端并完成既有功能配置，接入单位组织架构与账号体系
③ **增量定制**	针对本项目特有流程与报表专项开发，纳入交付范围

报价构成

- **包含**：三端部署、标准功能交付、组织架构对接、使用培训
- **不含**：增量定制开发、第三方模型调用费用（如涉及）

增量定制承诺

在现有功能基础上，客户提出的新需求可专项开发并纳入本项目交付。

- 定制内容单独报价、单独确认工期，**不影响标准功能的先期上线**
- 定制成果归客户使用，后续版本升级时一并纳入维护范围
- 我们不要求客户去适应软件，而是让软件适配客户的业务流程

目前已实现的交付成果

后台服务端自动任务，把现有能力串成一条全自动流水线：

- 对已上传的非结构化资料进行**自动分拣、自动识别、自动解析、自动处理**
- 基于这些素材**自动生成企业产品推广短视频**
- **自动生成企业推广的微信公众号文章**
- 支持客户指定的**任意成果类型**，形成完整的内容生产流水线

五、能呈现哪些成果

成果	**具体形态**
可检索的文字库	录音、扫描件、视频里的内容变成文字，进入统一检索库
规范化的识别文档	语音结果按发言人分段整理，可导出 TXT / PDF / Word
结构化的解析文档	合同、单据、台账的关键信息抽取成结构化字段
可复用的数据标签	AI 标签 + 人工标签，形成企业自己的数据分类口径
多模态成果文件	网络推文、动漫短片等衍生成果，可下载、可浏览、可分享
完整的行为台账	上传、查询、下载全留痕，数据家底与使用情况一屏掌握

六、对企事业单位的价值

采集侧：让数据自己汇集起来

数据汇聚最大的阻力不是技术，而是「让员工多做一步」。移动端后台自动上传 + 桌面端目录扫描，把这件事变成**员工不需要额外付出**的动作。

效率：把「人找数据」变成「机器先读」

场景	传统做法	使用石蛤蟆之后
听会议录音	1 小时录音约需 3-4 小时整理	自动转写并按发言人分段，分钟级出稿
翻找历史资料	挨个文件夹翻，或问同事	全文检索关键词直达，秒级定位到段
整理扫描件	人工录入合同、单据字段	OCR + 解析自动抽取关键信息
统计数据家底	各部门手工报数，口径不一	按单位/部门/人员/类型自动汇总并可导出

管理：看得见、管得住、说得清

- **看得见** —— 文件总量、存储占用、类型分布一屏掌握
- **管得住** —— 每个文件独立调阅权限，未授权者检索不到、下载不了；组织级隔离
- **说得清** —— 谁上传、谁查询、谁下载全程留痕

战略：为后续 AI 应用铺好底座

企业做 AI 应用，卡点通常不在模型，而在**没有可用的数据**。石蛤蟆解决的是「数据从哪来、能不能用」这个前置问题——数据只处理一次，成果可被所有 AI 应用与全部终端复用。

成本与风险

维度	收益
人力成本	减少人工听录、录入、整理、找资料的重复劳动
存储成本	三层去重与秒传，同一文件全单位只存一份
合规风险	权限 + 留痕 + 不使用设备唯一硬件标识

维度	收益
决策风险	基于真实数据家底做决策，而非部门口头报数

七、部署与实施

- **部署形态**：支持私有化 / 内网部署，数据不出单位；BS 端浏览器即可使用；桌面端与移动端可选配
- **对接能力**：复用单位 / 部门 / 人员公共表；标准 HTTP 接口；成果文件可回调业务系统

阶段	目标	内容
第一阶段	试点跑通	选 1 个部门，部署 BS + 移动端，打通自动上传与识别
第二阶段	治理上线	启用权限、标签、日志；桌面端接入存量目录
第三阶段	成果应用	成果制作与检索复用；移动端内容分发
第四阶段	全面铺开	多单位接入 + 对接客户自有 AI 应用