危机预演平台

从派生到自研的升级说明

三种时间尺度 · 一套证据 · 全部数字可重跑

版本	0.5.0
日期	2026-11-23
代码规模	7008 行 Python / 19 个测试文件 / 170 项测试
上游依赖	0 行上游代码，运行期零第三方依赖

一、一页看懂

这份文档说明三件事：**原来的版本为什么必须改、现在有什么能力、为什么它比一份报告值钱**。所有数字都来自仓库里的实跑结果，可以用一条命令复算。

起点很具体：早期的危机预演，是给一个别人的开源引擎打补丁。核心模拟能力不在自己手里，82.2% 的回测能重算、不能重跑，许可证是 AGPL，最主流的商业买断动机被法律堵死。

现在：**零上游代码、自研 8×8×72 动力学模型、三种预演、170 项测试、运行期零依赖**。同一件事从“证明我做对了”变成“你可以自己跑一遍”。

关键变化	之前（0.3）	现在（0.5）
上游派生代码	555 行补丁 + 依赖外部引擎	**0**
核心模型	在别人仓库里，用 API key 驱动	自研 8 群体 × 8 渠道 × 72 小时
回测数字	能重算，**不能重跑**	`crh dynamics` 一条命令重跑
交付形态	CLI + 一份报告	CLI / API / Web UI / 工作坊 / DOCX / PPTX
许可证	AGPL（派生）	**MIT，可商业授权**
测试	90 项	**170 项全通过**

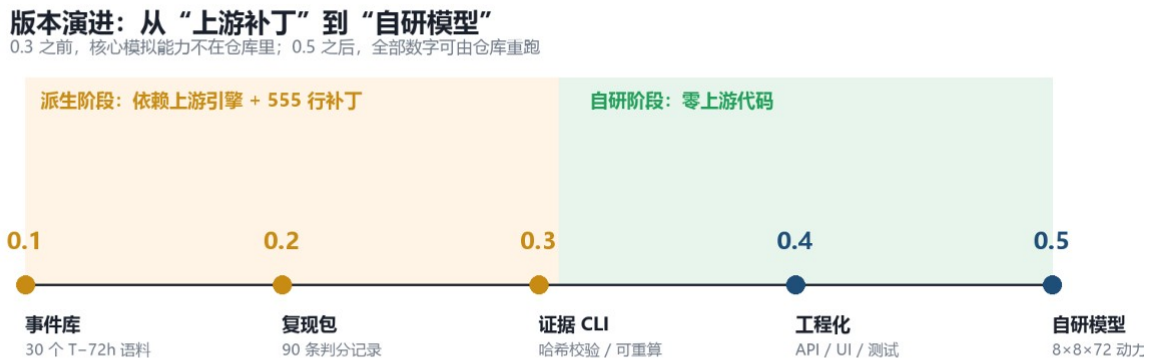


图 1　版本演进：0.3 之前依赖上游，0.4 起全部自研

二、为什么必须从派生走向自研

2.1 派生版本的三个硬伤

- **引擎不在手里。** 仓库只有 555 行补丁，买家两天就能复现，任何报价都会被“我自己做”击穿。
- **证明依赖旧管道。** 90 条判分记录由已不发布的引擎产生，能重算、不能重跑——尽调第一问就答不上来。
- **AGPL 锁死商业授权。** 派生作品的著作权不完整，不能给买家闭源授权，最主流的买断动机直接消失。

升级不是“重写一遍”，而是把这三条逐个拆掉：模型自研（引擎回到手里）、动力学可重跑（证明闭环）、代码全原创 + MIT（授权自由）。

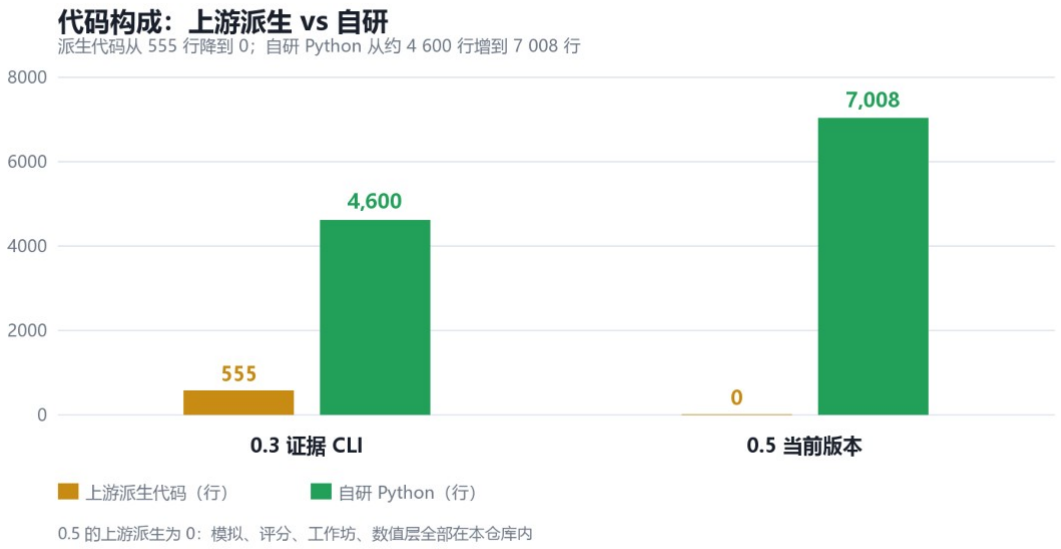


图 2　代码构成：上游派生 555 → 0，自研代码 4 600 → 7 008 行

2.2 升级前后对照

维度	0.3 证据 CLI	0.5 当前版本
模拟	外部引擎 + LLM key	自研动力学，确定性，<1 秒
复现	重算统计，不能重跑	重跑整条链路，逐字节一致
服务	无	HTTP API / Web UI / Prometheus
交付	HTML + Markdown	+ DOCX / PPTX / 工作坊控制台
数值	无	区间 + 覆盖度 + 发布门禁
商业	AGPL，难以买断	MIT，可买断、可授权、可 OEM

三、现在有什么：八项能力

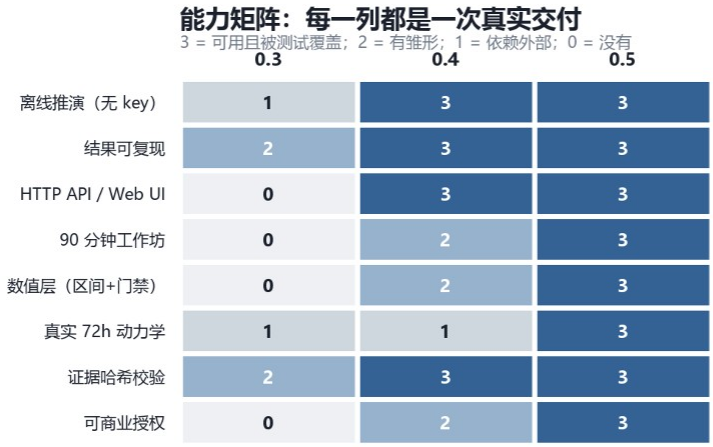


图 3　能力矩阵：3 = 可用且被测试覆盖，0 = 没有

能力	具体是什么	怎么验证
语料与证据	30 个中文危机事件、90 条判分记录、50 个文件哈希	`crh evidence verify` 退出 0
动力学引擎	8 群体 × 8 渠道 × 72 小时，参数在 JSON 里	`crh dynamics --sensitivity`
风险本体引擎	17 个风险原型，规则可解释、可改	`crh rehearse --event 1`
三种预演	72 小时危机 / 90 分钟工作坊 / 90 天产品	`crh workshop` / `crh blueprint`
数值层	区间、方向概率、样本外覆盖、发布门禁	`crh forecast --csv ...`
Oracle 协议	只收分布，未验证不放行	`crh oracle` 退出 3
服务面	CLI / API / UI / 指标 / 容器	`crh serve`
交付件	HTML / Markdown / DOCX / PPTX / JSON	`--formats docx,pptx`

四、三种预演，三种时间尺度

预演	时间尺度	输入	输出	卖給谁
危机预演	72 小时	T−72h 场景包	三路径轨迹 + 决策卡	企业 CRO / 公关
工作坊	90 分钟	决策卡 + 4 个注入	决策日志 + 会后 Readout	高管团队（一场 3–5 万）
产品预演	90 天	一份 brief	先做哪个 bot、技能路线、失败预演	做 AI 智能体的人

4.1 危机预演：同一冲击，三条不同的 72 小时

输入同样的初始冲击 0.55，选择不同，结果完全不同。这是“预演”与“预测”的区别：它不比谁猜得准，它把三条路的代价摆到同一张桌上。

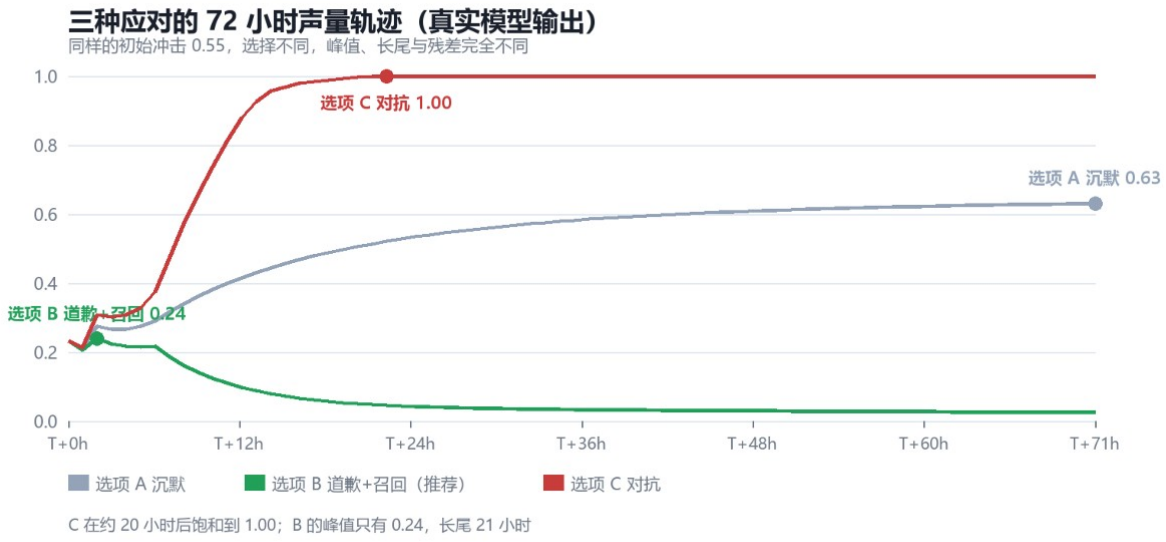


图 4　72 小时声量轨迹：C 在约 20 小时饱和，B 的峰值只有 0.24

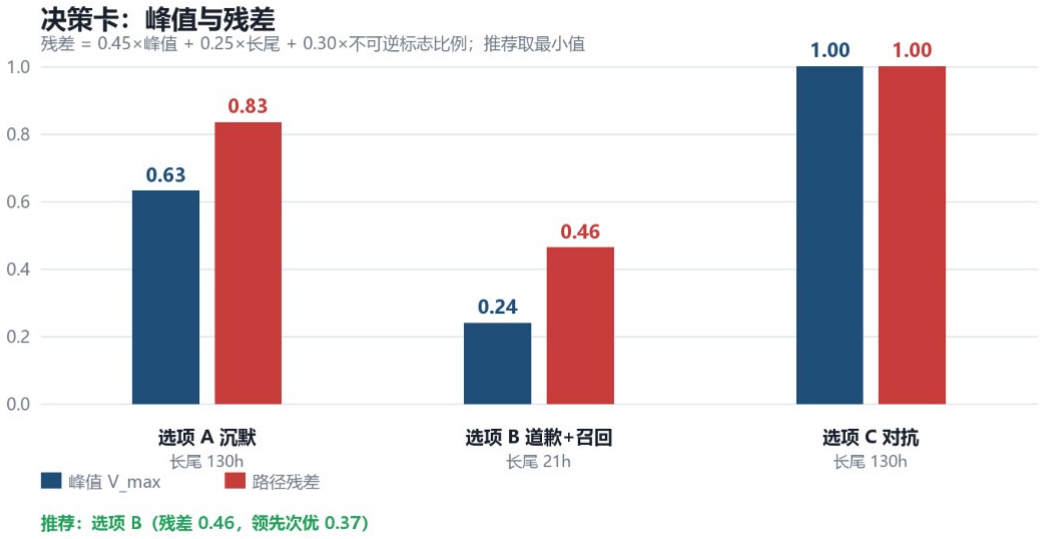


图 5　决策卡：残差排序给出推荐，且领先幅度写明

```
决策点：是否在 6 小时内公开道歉并启动实质整改/召回?
├ 选项 A 沉默        峰值 0.63 / 长尾 130h / 残差 0.834 / 可撤回
├ 选项 B 道歉+召回   峰值 0.24 / 长尾  21h / 残差 0.462 / 不可撤回   ★推荐
└ 选项 C 对抗        峰值 1.00 / 长尾 130h / 残差 1.000 / 不可撤回
触发条件：24h 内出现监管问询 → 立即转 B
反悔条件：道歉后 48h 内二次爆料 → 升级 B+（第三方审计）
```

代码块 1　决策卡的结构：选项、代价、触发与反悔条件、责任人

4.2 工作坊：把报告变成一场会

工作坊控制台是一个自包含 HTML：90 分钟倒计时、4 个按时投放的注入、决策与异议日志（存在本机，刷新不丢）、一键导出 Markdown、可打印成 PDF。推荐结果按残差推导，不是写死的；会前自检会先把这份包的毛病列出来。

4.3 产品预演：给想做智能体的人

输入一份 brief（团队、预算、数据、买家），输出：先做哪个 bot、先建哪些技能（带依赖与上线门禁）、会怎么失败（按残余风险排序）、90 天计划、首单价与所需客户数。

样例 brief	先做	MVP 技能	最大残余风险	首单价
中小制造：售后+备件	选型与报价 agent（0.88）	6 项 / 13 人周	做出来了没人用（5.00）	12 000 元/年
给企业接单的小团队	销售线索跟进（0.75）	8 项	做出来了没人用（5.00）	175 000 元/年

五、证据：这些数字为什么可信

主结果是 30 个事件 × 3 跑的事件均值 82.2%（95% CI 78.5–85.9）。这个数字以前是“自证”，现在它可以被重跑，而且旁边站了一个对照组。

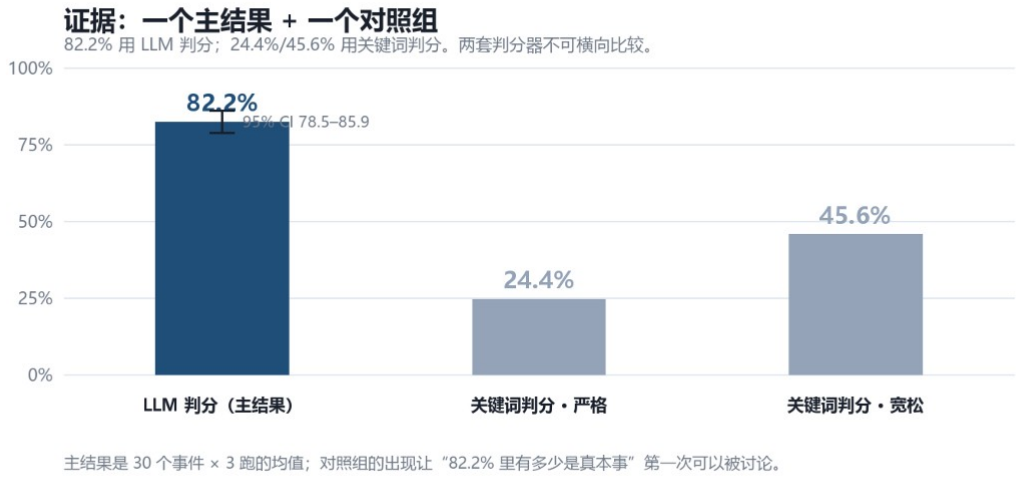


图 6　主结果与对照组：两套判分器不可横向比较

指标	数值	口径（引用时必须带上）
事件均值命中率	82.2%	LLM 判分，30 事件 × 3 跑，分级评分
95% 置信区间	78.5–85.9%	t 分布，自由度 29
跨跑一致	5 / 30	三次运行得分完全相同
离线基线	24.4% / 45.6%	关键词判分，严格 / 宽松，与 82.2% 不可比
稳定性 STA	未通过自设门禁	已公开，不隐藏

对照组的意义：单臂实验里，没人能回答“82.2% 里有多少是真本事、多少是判分器宽松”。把 24.4% 摆出来，这个问题第一次有了讨论的起点。

六、数值层：为什么只给区间，不给点位

实测：NVDA 次日收盘的朴素预测平均误差是 3.50 美元（2.02%），BTC 是 1 387 美元（1.65%）；动量模型比随机游走更差。误差由波动率决定，不由模型决定。

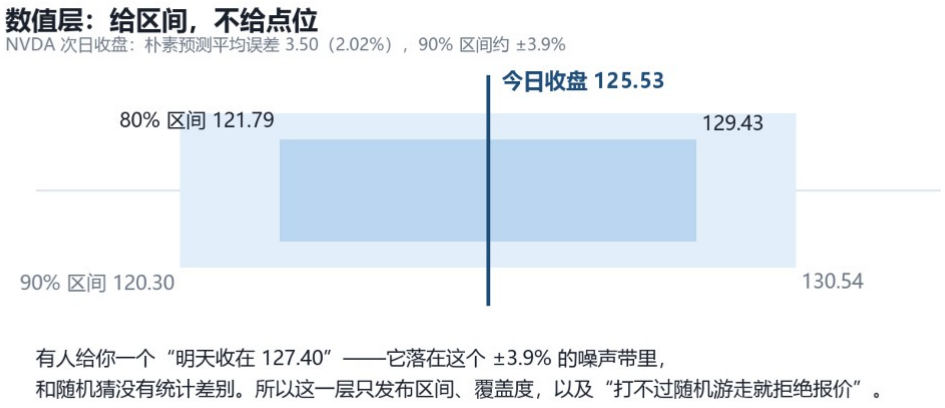


图 7　区间与噪声：所谓“精确点位”落在噪声带里

所以这一层只发布三样东西：区间、覆盖度（样本外测）、以及“打不过随机游走就拒绝报价”的门禁。退出码 3 就是“这个数不能对外引用”。

七、为什么值钱

同一套能力，卖法不同，单价差 50 倍。差别不在功能，而在你卖的是“文件”还是“决策”。



图 8　四种卖法的价格中位数（条件写在正文）

形态	买家	价格	交付
预演档案	个人 / 小团队	2 000 元 / 份	HTML + Markdown

形态	买家	价格	交付
90 分钟工作坊	公关公司、企业 CRO	3–5 万元 / 场	决策卡 + 会议包 + Readout
年度授权	企业、集团风控	5–10 万元 / 年	离线部署 + 语料更新 + 支持
OEM / 买断	集成商、有项目的公司	8–15 万元	源码 + 语料 + 过渡支持

为什么工作坊能卖 3–5 万而报告只能卖 2 000：报告是一次性交付，买家验证一次就结束；工作坊绑定的是流程与台账，每次事件都要重估，天然复购。

八、怎么用：六条命令

```
crh demo --out demo           # 三份自包含档案（离线，无 key）
crh dynamics --title X --sensitivity   # 自研模型：决策卡 + 参数敏感度
crh workshop from-event --event 6 --formats docx,pptx --out ws/
crh blueprint --brief brief.json --out bp/
crh forecast --csv data/market/nvda.csv   # 数值层：区间 + 门禁
crh serve --port 8080          # API + Web UI + /metrics
```

代码块 2　主要命令

一天跑通的路径：**上午**用 `crh workshop from-event` 生成一份真实事件的工作坊，**下午**用 `crh blueprint` 判断下一个该做什么，**晚上**把两份 DOCX/PPTX 发出去。

九、局限与下一步

- **判分器未人工校准。** 82.2% 的判分器是 LLM，没有人工 κ；这是下一步（2–3 周）。
- **语料是中文市场。** 30 个事件全部来自 2018–2026 年的中国公开新闻，英文语料是 0。
- **动力学参数是作者判断，不是拟合结果。** 参数在 JSON 里可改，但没有用真实数据校准过。
- **没有付费客户。** 上面所有价格都是可比锚点，不是成交记录。
- **Office 导出是可选依赖。** 主安装包保持零依赖，DOCX/PPTX 需要 `[office]` extra。

下一步三件事，按性价比排序：**第一笔付费案例**（把成本法变成收益法）、**证据闭环**（三臂盲评 + κ，解开授权与买断）、**英文语料**（海外才有的谈）。

附录 A　规模与构成

项	值
Python 源码	27 个文件 / 7008 行
测试	19 个文件 / 1772 行 / 170 项全通过
语料与目录	51 个数据文件（其中 50 个纳入哈希清单）
文档	8 篇工程文档 + 本报告
运行期依赖	0（标准库）
可选依赖	python-docx / python-pptx（仅导出）

— 完 —