

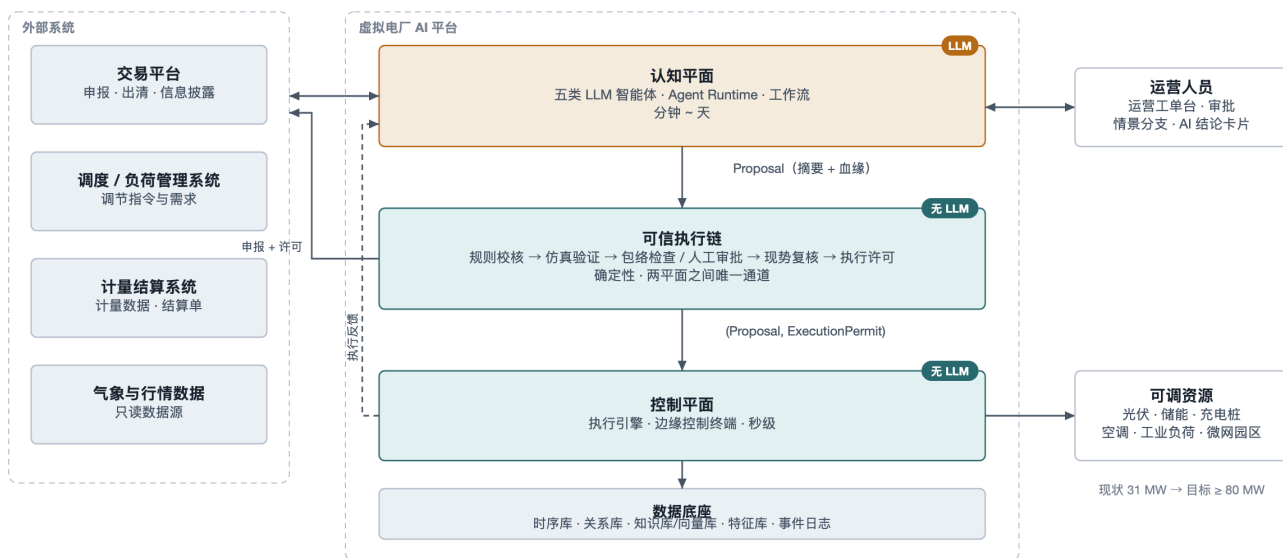
虚拟电厂 AI 平台 — 系统架构

基于光明电力大模型的虚拟电厂多时空协同智能运营平台

本平台是面向湖北虚拟电厂运营的 AI 协同决策与受控执行系统。五类 LLM 智能体生成市场申报、调度方案与负荷控制计划；任何对外效果产生之前，都必须经过一条确定性的可信执行链。三条设计承诺定义了整个架构：

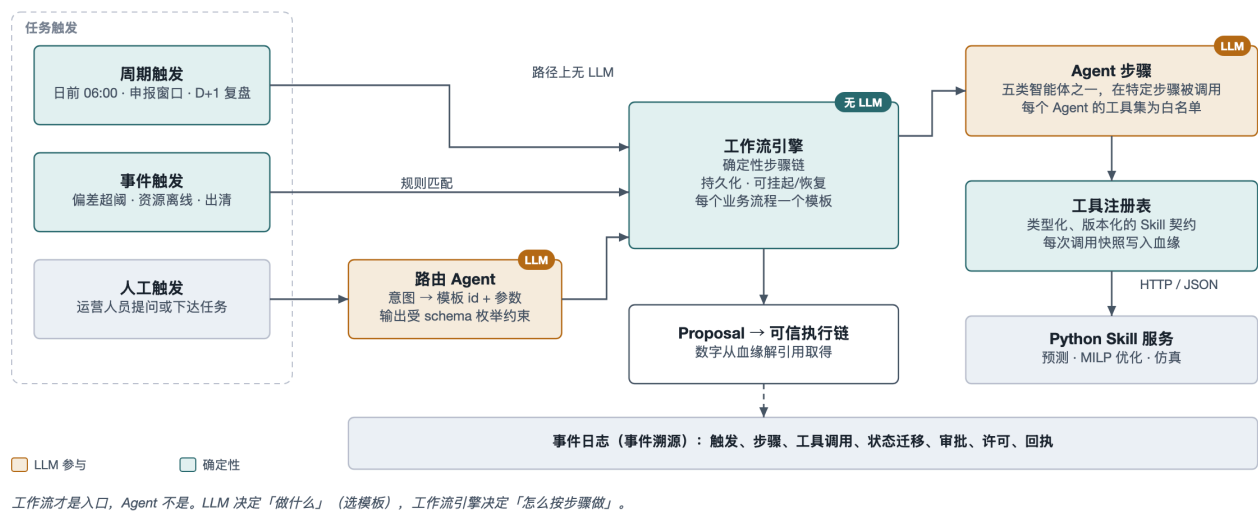
- **两平面分离，唯一通道。** 慢速的认知平面（LLM 智能体，分钟到天）与快速的控制平面（确定性执行，秒级）严格分离，可信执行链是两者之间唯一的通道。
- **LLM 不算数。** 申报价格、电量与控制设定值只能来自版本化的预测、优化与仿真工具；Proposal 中的每一个数值字段都是对某次已记录工具调用的引用。
- **包络自治。** 人工审批的对象是包络（带有效期的边界授权），而不是每一笔动作。包络内的动作经规则校核与仿真验证后自动放行，其余一律升级人工。包络从空集起步，仅凭证据放宽。

1. 系统上下文



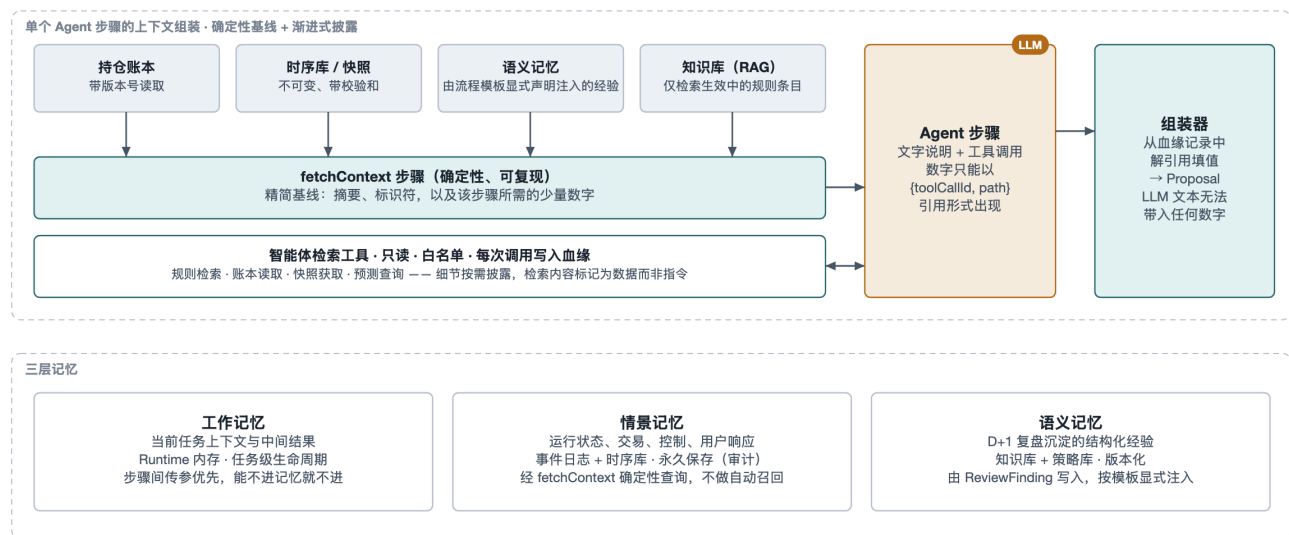
平台向上对接省级交易平台、调度/负荷管理系统与计量结算系统，向下聚合可调资源（现状 31 MW，目标不低于 80 MW）。运营人员通过运营工单台工作：每一项 AI 产出都挂在一张有明确目标、责任人和期限的工单之下。平台是既有 3060 平台之上的叠加层：复用主数据与权限体系，并把所有申报与控制动作收敛到唯一一条留痕通道，使任何动作都无法绕过可信执行链。

2. Agent Runtime



工作流才是入口，Agent 不是。每个业务流程（日前申报、复盘、日内纠偏）都是一条确定性、可持久化的步骤链，Agent 只在特定步骤被调用，并且只能使用白名单内的工具：产出数字的计算类 Skill，以及按需获取细节的只读检索工具。周期触发与事件触发直接实例化流程模板，全程不经过 LLM；只有人工请求经过路由 Agent，而路由 Agent 的输出被 schema 约束为「已注册模板 id + 参数」。LLM 决定「做什么」，工作流引擎决定「怎么做」。每一次触发、步骤、工具调用、状态迁移、审批与许可都写入事件溯源日志。

3. 上下文管理与记忆



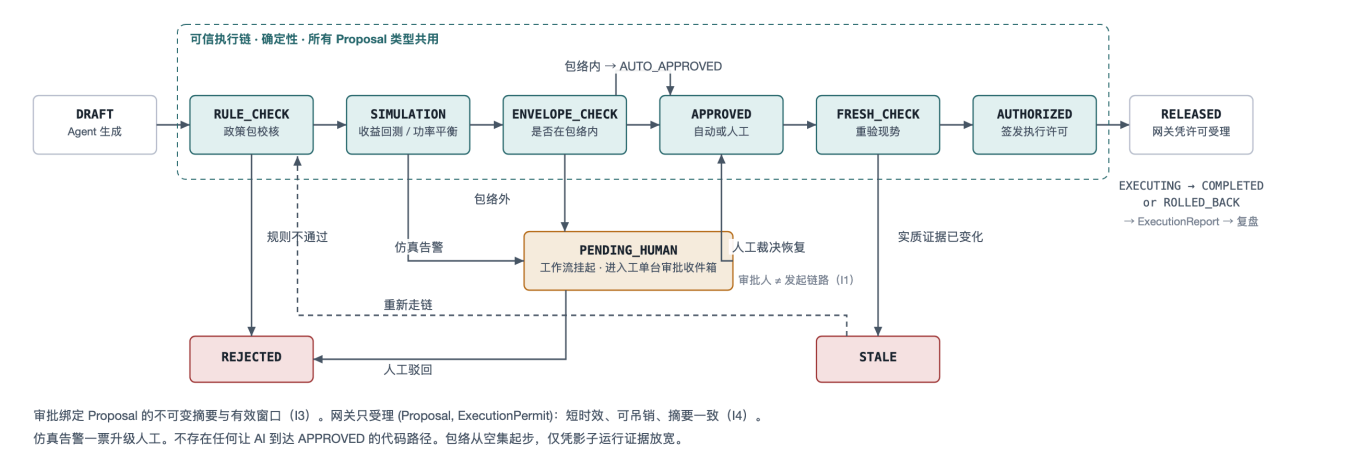
运营事实的权威存储是领域数据库，而非智能体记忆（不变式 18）。生产决策不依赖任何自动记忆召回。

每个 Agent 步骤的上下文由上游的确定性步骤组装：读取带版本号账本、不可变快照、由模板显式声明的语义记忆，以及生效中的知识条目。Agent 不选择自己的基线上下文，因此每次决策都可从复现的起点出发。Agent 输出中的数字只能以对工具调用血缘的引用形式出现，由组装器解引用填值——这意味着 LLM 文本不可能把任何数字带进 Proposal。

上下文按**渐进式披露**构建。基线是精简的：摘要、标识符，以及该步骤所需的少量数字，按保守的上下文窗口设计。更多细节通过**智能体检索工具**按需披露：只读、白名单化的注册表工具，如规则检索、账本读取、快照获取、预测查询，Agent 在需要时可以调用。每次检索连同输入输出记录到血缘，因此「Agent 当时看到了什么」可以完整重放；提示注入防线把检索到的内容一律当作数据，而非指令。同一原则也贯穿运营界面：AI 结论卡片先给出结论，展开后才显示背后的工具输出与数据来源。

三层记忆各有不同的生命周期。工作记忆只存活于单个任务；情景记忆就是事件日志与时序库，永久保存，通过查询而非召回使用；语义记忆存放 D+1 复盘写回的结构化经验，版本化，并由各流程模板显式注入。运营事实的权威存储永远是领域数据库，而不是智能体记忆。

4. 可信执行链与状态机

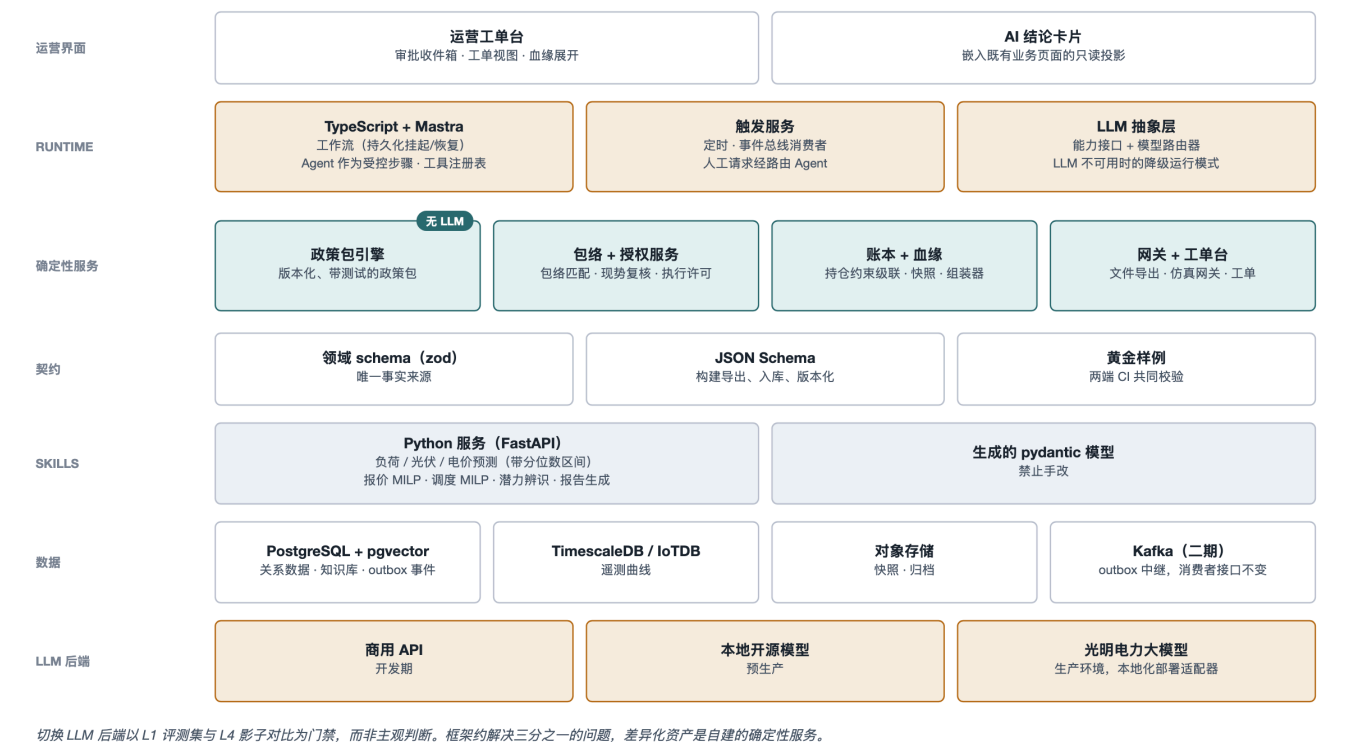


一切拟执行动作都是一个 **Proposal**：带类型化载荷、完整血缘与不可变摘要。所有 Proposal 类型共用同一条生命周期 workflow：

环节	内容
规则校核	确定性政策包：合规性、约束、账本一致性、血缘完整性、职责分离
仿真验证	申报类做收益回测，控制类做功率平衡；仿真告警一票升级人工
包络检查	落在已批准包络内：自动放行并完整留痕；否则 workflow 挂起，进入工单台审批收件箱
人工审批	由身份与发起链路分离的审批人恢复 workflow；审批绑定摘要与有效窗口
现势复核	下发前即时重验：资源在线、约束仍满足、证据未变化；否则置为 STALE 重新走链
许可与放行	签发 ExecutionPermit ：短时效、可吊销、摘要一致；网关只受理 (Proposal, ExecutionPermit) 对

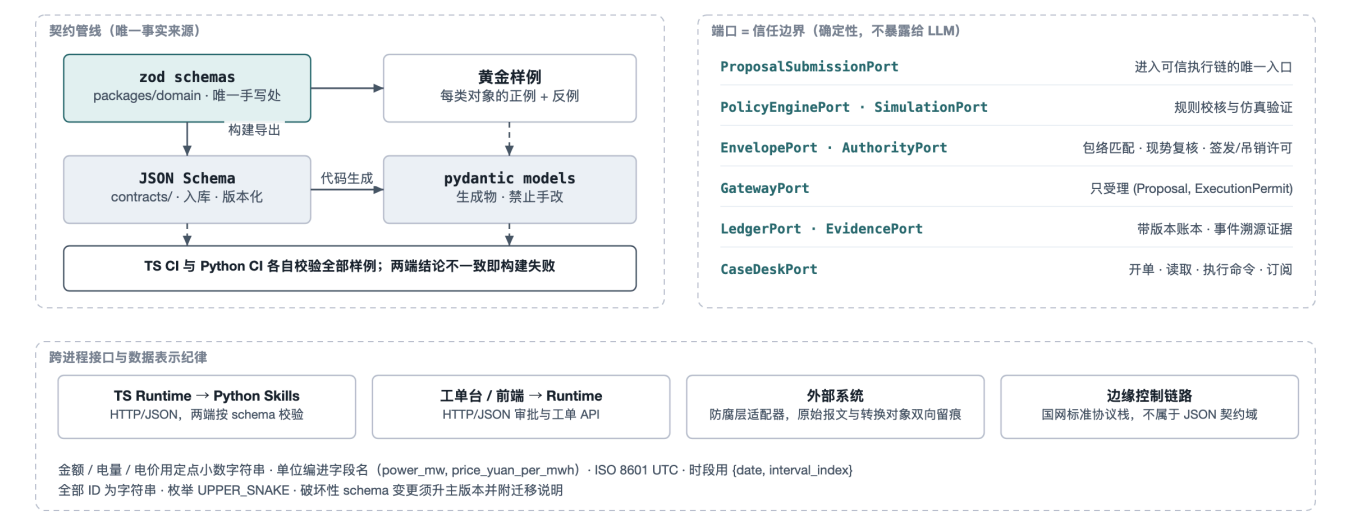
挂起中的审批可跨进程重启存活。许可吊销是最快的回退手段：网关立即拒收后续指令，边缘终端退回上一批准计划或安全曲线。

5. 技术栈



Runtime、智能体、工作流、领域 schema 与确定性服务均为 TypeScript + Mastra, 看重的是类型化工具契约与持久化的挂起/恢复能力; 专业模型 Skill 是 Python HTTP 服务。LLM 调用经由面向能力的抽象层与模型路由器, 后端从开发期的商用 API、预生产的本地开源模型, 切换到生产环境的光明电力大模型适配器, 业务代码零改动。Runtime 还提供 LLM 不可用时的降级模式: 预测、优化、校核、审批、申报与控制全部无需 LLM 即可运行。框架大约解决三分之一的问题, 差异化资产是自建的确定性服务。

6. 接口与契约



手写的 zod schema 是唯一事实来源。构建时导出 JSON Schema 到入库的 contracts 目录，再据此生成 pydantic 模型，禁止手改。黄金样例（含正例与反例）由 TypeScript 与 Python 两端 CI 各自校验，任何结论不一致即构建失败。平台内部，每条信任边界都是一个 LLM 不可调用的端口：Proposal 提交、政策引擎、仿真、包络、授权、网关、账本、证据与工单台。数据表示纪律也是契约的一部分：金额与电量用定点小数字符串、单位编进字段名、UTC 时间戳、时段用日期加序号、全部 ID 为字符串、枚举全大写。

7. 评测驱动的开发过程



由于每个决策都能从留存证据完整重放，审计留痕本身就是评测数据集。四层评测对象用于定位问题：L1 LLM 任务，L2 专业模型 Skill，L3 可信执行链正确性（每条不变式至少一个自动化测试），L4 通过影子运行与历史回测衡量的端到端决策质量。评测就是变更门禁：提示词变更须通过 L1 任务集与 L4 回测；后端切换须通过 L1 全集与 L4 影子对比；Skill 升版必须同在一次变更中更新其已提交的 L2 基线；政策包升版须规则测试全绿；包络放宽仅凭 L4 影子或在线数据审批。每条复盘结论都会成为候选评测用例，失败案例持续喂养下一版基线。