



MODEL CAPACITY FUTURES EXCHANGE

期算白皮书

发布日期

2026年7月26日

版本

v3.1

目录

部分	内容	页码	部分	内容	页码
00	阅读说明	03	01	执行摘要	04
02	产品与合约	06	03	市场参与者与会员	08
04	交易机制	09	05	中央清算与全额担保	12
06	实物交割	15	07	价格、流动性与商业模式	17
08	市场诚信与运行风险	18	09	GLM-5.2 完整交易示例	19
10	运营门槛与成功标准	21	11	结论	22
12	附录：术语表	23			

阅读说明

本文说明 KAI 完整期货交易所的产品、市场、清算与交割制度

白皮书定位

KAI 是在香港运营、面向机构的模型服务容量期货交易所。“期算”是未来指定交割小时内的标准化模型服务容量期货合约。本文规定其产品、市场、清算、交割、风险、治理与运营标准。

本文建立产品、市场、清算、交割、风险与治理的一致制度基线。所有规则与参数通过正式评审、版本控制和生效日统一管理，技术实现与运营流程始终采用同一生效版本。

除非品种或场景另有说明，本文的市场与运营语境默认在香港，业务时区使用 [Asia/Hong_Kong](#)，计价币种使用 HKD；交割服务区域仍由各品种明确。

规范用语

- **必须**表示硬性规则。
- **应、原则上**表示除规则簿明确列出例外外均需遵守。
- **可以**表示规则允许但不要求的能力。
- 示例主体、价格、日期、模型与阈值用于解释机制；现实参数以正式规则与服务文件为准。

术语约定

术语	本文含义
期算	以未来指定小时的标准化模型服务容量为标的、通过交易所连续撮合、由 KAI 中央清算并最终实物交割的期货合约
一手	在整个交割小时内，每个连续 60 秒计量窗口持续可用的 1,000,000 TPM
多头	在交割关闸后取得模型服务容量使用权的一方
净空头	在交割关闸后承担模型服务容量交付义务的供应商头寸
全额担保	多头覆盖最坏成交价款与费用；净空头同时受到可交付容量和履约担保约束
中央对手方	成交后替代原交易双方，分别成为每个多头的卖方和每个空头的买方的 KAI 清算体系
交割关闸	停止交易并锁定最终多头权利、供应商净空头义务及适用规则的时点

执行摘要

KAI 让未来模型服务容量成为可交易、可清算和可实物交割的机构期货

一句话定义

KAI 是面向机构的模型服务容量期货交易所：用统一合约表达未来模型服务能力，用完整订单簿形成价格，用全额担保和中央清算控制信用风险，并以真实 Chat Completions 服务完成交割。

生成式人工智能进入生产系统后，企业关注点从平均 Token 单价扩展到确定时间、区域和模型版本下的容量保障。供应商则需要把不可储存的未来能力提前出售，改善资源排期与收入确定性。KAI 在双边合同、普通按量 API 和底层 GPU 预订之外，提供跨供应商标准、连续价格、持仓转让和统一交割。

期算把模型服务容量按小时标准化。每手合约代表：

$$1,000,000 \text{ TPM} \times 1 \text{ delivery hour}$$

每个 60 秒计量窗口的加权 TPM 为：

$$\text{TPM} = I + W \times O$$

I 为经确认计费的输入 Token， O 为实际生成并计费的输出 Token， W 为品种规则冻结的输出权重。输入 Token 按原数计入，输出 Token 按 W 倍计入；容量按窗口独立核定，并在交割小时边界结束，代表持续吞吐能力。

KAI 的制度核心是：

- 1. 完整交易所。** 所有有效买卖意图进入统一中央限价订单簿，按价格—时间优先成交，并支持完整订单类型、市场状态、行情、持仓和市场监察。
- 2. 机构会员制。** 企业客户通过交易会员接入；清算会员承担资金与交收责任；容量供应商另需取得交割参与者资格。
- 3. 无经济杠杆。** 多头订单预留最坏价款和费用；只有拥有已核准容量及履约担保的供应商可以形成净空头。
- 4. 中央清算。** 成交立即更替为 KAI 中央对手方义务，资金、持仓、未平仓量和交割责任在同一清算体系内管理。
- 5. 实物交割。** 最终多头在指定小时取得 KAI Chat Completions 容量；最终供应商净空头以真实推理服务履行合约。

6. **价格风险与交付风险分离。** 平仓产生交易价差；履约担保覆盖替代采购、质量和集中度风险，服务价值始终锚定真实容量。

产品与合约

统一标的、交易单位、权利和期限，让每笔成交指向同一种可交付服务

标准合约参数

每个挂牌品种由版本化规则完整定义：

参数	规则
模型与版本	固定模型名称及可验证版本 digest；重大升级建立新品种
上下文上限	单次请求的统一总上下文上限
交易单位	一手为每个交割窗口持续可用的 1,000,000 TPM
交割小时	使用明确 IANA 时区的连续 60 分钟
交割区域	固定服务、网络及数据处理区域
报价	HKD/手，按固定 tick size 的十进制价格
数量	整数手，按固定 step size
超额处理	达到窗口上限后暂停新增请求；排队、降速或补购由品种规则明确
基本交付标准	可用性、系统错误率、首 Token 延迟、TPM 达成和证据要求
交易日历	开市、休市、停牌、最后交易和交割关闭时点
规则版本	订单、成交和交割全程绑定；每项事实按其绑定规则解释

经济含义

多头获得未来容量使用权，其范围由吞吐、可用性、延迟和服务标准界定。供应商净空头承担在指定小时持续提供合格服务的义务；GPU、模型权重和最终用户身份均位于合约标的之外。

每手在每个窗口的容量上限为：

$$C_{\max}(N) = N \times 1,000,000 \text{ TPM}$$

N 是持有手数， $C_{\max}(N)$ 是任一 60 秒窗口的容量上限。手数每增加一手，该窗口上限增加 1,000,000 TPM；未使用额度不累积到后续窗口，并在交割小时边界结束。交割小时结束前已由 KAI 接受的请求继续完成，新增合同内请求以小时结束时点为截止。

完整期货与全额担保

KAI 以完整期货市场基础设施配合全额担保资金制度，提供标准合约、中央订单簿、持仓、平仓、未平仓量、中央清算、交割和市场监管：

- 多头开仓前覆盖最坏可接受价款及最大费用。
- 普通买方和流动性参与者的卖出数量以其已有多头为上限。
- 只有具备交割资格的容量供应商可以在已核准、未占用容量及履约担保范围内建立净空头。
- 账户购买力等于已提交的可用资金，经济敞口保持 1:1 覆盖。
- 标记价格服务于风险、价格保护和履约担保；价差在反向成交时结算。

产品范围

维度	产品范围
市场准入	完成账户、权限和资金设置的会员与机构客户
产品	按小时、实物交割、全额担保的模型容量期货
空头	由具备交割资格的参与者以核准容量和履约担保支持
服务承诺	标准化容量和服务等级；输出内容责任按服务文件界定
治理	交易、供应和争议裁决采用清晰的利益冲突隔离

市场参与者与会员

用会员、清算和交割资格把交易权、资金责任与服务责任分开

KAI 市场基础设施

KAI 是交易、清算与交割的唯一运营组织：

组织	核心职责
KAI	品种准入、会员管理、订单簿、行情、市场监察、CCP 合约更替、抵押品、账本、违约管理和最终交割

交易、清算与交割不再对应独立组织。KAI 在同一组织内通过职责分离、访问权限、独立账本科目、风险委员会和利益冲突控制保持职能边界；市场数据只服务于授权职能，清算资源专用于 CCP 风险与履约。

会员与资格

资格	能力	责任
交易会员	为自身或机构客户管理订单	授权、订单控制、记录与交易行为
清算会员	为自身及关联交易会员清算	资金、抵押品、限额、违约与对账
交割参与者	登记容量并形成净空头	Provider 接入、担保、计量与补救

做市和数据使用是叠加在相应会员或用户上的权限。一个机构可以按规则持有多项资格，并分别接受审批、监控和撤销；交割参与者通过交易会员账户管理订单，Provider 控制面专注于容量、健康、计量和争议。客户订单始终归属于交易会员、清算会员、客户账户和控制组，技术子账户沿用同一控制关系。

机构 B2B

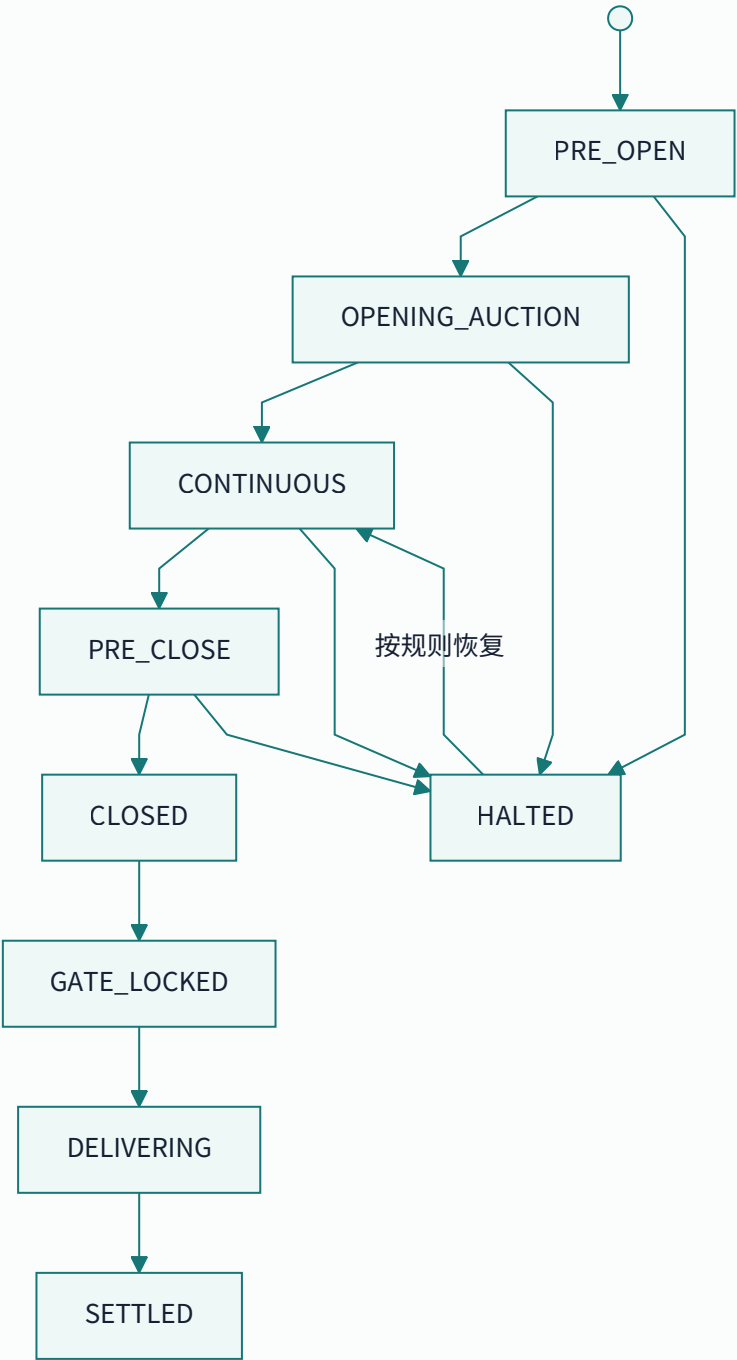
市场服务面向具备真实容量需求、资金能力和对应操作团队的机构。企业客户完成 KAI 账户、权限、资金与技术设置；供应商另须证明模型服务来源、区域、可持续 TPM、并发、延迟、运营连续性和资金能力。技术子账户统一聚合至控制组与交割持仓。

交易机制

完整订单簿把所有有效买卖意图转换为唯一、确定和可审计的成交顺序

市场状态

每个品种按版本化交易日历运行：



HALTED 可以从交易阶段进入，并经明确规则恢复。已经成交的合约在停牌期间持续有效；未成交订单按停牌原因和品种规则保留、撤销或冻结。

开盘集合竞价依次选择最大成交量、最小未成交量、最接近参考价的价格，并使用固定规则打破并列。连续交易按价格—时间优先。收市价格和参考价格由合格成交、集合竞价或规则规定的替代来源自动形成。

中央限价订单簿

- 买盘价格从高到低，卖盘价格从低到高。
- 同一价格档位按交易所接受序列先后成交。
- 市场数据公布最佳买卖价、聚合深度、成交和连续序列，客户身份保持机密。
- 每次被接受的订单修改都保留原订单 ID、取得新的时间优先序列；终态保持稳定。
- 已完成成交作为不可变事实保留，撤单、停牌和关闸只影响后续状态。
- 优先级完全由交易所接受序列和规则确定，撮合优化保持同一结果。

订单类型

类型	市场行为
LIMIT	按指定价格或更优价格成交，剩余部分按有效期处理
MARKET	在价格保护带和可用流动性内成交，价格边界始终有效
STOP / TAKE_PROFIT	触发后成为限价订单
STOP_MARKET / TAKE_PROFIT_MARKET	触发后成为受保护的市价订单
TRAILING_STOP_MARKET	根据已记录参考价和回撤比例更新触发阈值

有效期	市场行为
GTC	持续至成交、撤销、到期、停牌规则清理或关闸
IOC	立即成交可得部分，剩余撤销
FOK	必须完整成交，否则无任何成交
GTX	只做挂单；若会立即成交则拒绝
GTD	持续至指定且已记录的到期时间

品种规则可以限制可用组合，但交易核心、API 和测试必须实现完整集合。触发源取自已进入权威事件序列的价格，重放沿用同一历史事实。

交易前约束

订单进入订单簿前按固定顺序验证：

1. 交易会员、清算会员、客户和权限均有效。
2. 品种处于允许该操作的市场状态。
3. 价格、数量、订单类型和有效期符合当前规则版本。
4. 多头订单已预留最坏价款和费用。
5. 卖出后普通账户净持仓保持在零或以上。
6. 供应商新增净空头有足够容量 grant、履约担保和集中度额度。
7. 账户、会员、控制组和关联组未突破持仓或频率限制。
8. 自成交防护、价格带和临近交割限制全部通过。

未通过检查的订单在入簿前返回明确结果，持仓、资金和优先级保持原状。

持仓、平仓与未平仓量

交易完成后，账户持仓按品种分别记录多头、供应商净空头、加权平均开仓价、已实现价差和冻结数量。

- 买入先平已有净空头，再增加多头。
- 卖出先减少已有多头；只有交割参与者可以在容量和担保允许时继续增加净空头。
- 平仓价差按成交价与该方向加权平均开仓价计算，并通过清算账本结算。
- 市场未平仓量只在新多头与新空头同时增加时上升，在双方平仓时下降。
- 关闸后的最终净持仓决定交割权利、义务和实际 Provider 路由。

市场保护

品种规则必须定义静态或动态价格带、最大市价单影响、异常波动停牌、开盘参考价、错误交易处理和恢复程序。原成交作为审计事实持续保留，调整通过追加的取消、冲正或补偿流程完成，并遵循交易所规则和公平通知。

中央清算与全额担保

每笔成交在同一时点完成合约更替、资金占用、持仓更新和交割责任记录

CCP 合约更替

撮合成交后，KAI 分别成为多头的卖方和空头的买方，原双边信用关系转化为双方对 KAI 的资金、担保、交割和违约责任。

一个成交必须原子产生：

1. 买卖订单的成交与剩余数量。
2. 双方持仓、平均开仓价和市场未平仓量变化。
3. 多头价款、供应商履约担保和费用的占用或释放。
4. CCP 应收应付、客户资金及会员清算分录。
5. 行情、会员私有回报和市场审计事件。

成交与清算在同一个原子决定中成立；清算约束通过后，成交事实才正式生成。

多头全额担保

多头订单的最低可用资金为：

$$C_{\text{long}} \geq P_{\text{worst}} \times Q \times U + F_{\text{max}}$$

C_{long} 是可用于买单的资金， P_{worst} 是限价或市价保护带确定的最坏成交价， Q 是手数， U 是每手乘数， F_{max} 是最大费用。系统先按最坏成交情形计算全部价款，再加上最大费用；可用资金至少达到该合计，订单才可入簿。部分成交只占用实际成交价格款并保留剩余订单所需资金。

卖出已有多头时，相应权利冻结；成交后释放原持仓成本并结算已实现价差。普通账户的可卖数量等于其未冻结多头上限。

供应商履约担保

供应商净空头必须同时由可交付容量和履约担保支持。最低担保为：

$$IM_s = \max(M_{\text{min}}, E_{\text{replacement}} + A_{\text{stress}} + A_{\text{concentration}})$$

IM_s 是供应商 s 的最低履约担保，取两项中的较大值：固定底线 M_{min} ，或替代采购敞口 $E_{\text{replacement}}$ 、压力附加 A_{stress} 与集中度附加 $A_{\text{concentration}}$ 之和。因此担保不会低于固定底线，并会随替代采购或附加风险上升。

其中，替代采购敞口按下式计算：

$$E_{\text{replacement}} = \max(0, P_{\text{mark}} - P_{\text{entry}}) \times Q_{\text{open}} \times U$$

P_{mark} 是标记价， P_{entry} 是平均开仓价， Q_{open} 是未平净空头手数， U 是每手乘数。 $\max(0, \cdot)$ 表示只计入标记价高于开仓价时的正向替代价差，再按未平手数和乘数放大为总敞口。

标记价变化可以导致追加担保、限制新增空头或启动违约处置。正常头寸的价差在反向成交时结算；供应商买入平空时，同时释放对应容量与担保。

会员清算与客户资金

企业客户通过交易会员下单，由清算会员向 KAI 承担清算责任。客户资金与 KAI 自有资金、会员自有资金和 CCP 风险资源分账管理。

KAI 通过清算会员接收资金与抵押品。交易会员和清算会员保存客户级余额、持仓、订单及控制组映射，使 KAI 能按账户和控制组聚合监控；客户资金与会员、CCP 和 KAI 自有资源分账管理。

不可变双重记账

所有资金事实通过不可变分录表达。每个账本交易在资产维度借贷相等，余额由分录投影产生。成交价款、平仓价差、手续费、履约担保、交割结算、违约扣款和赔付均使用独立业务原因。

代表性账户包括：

账户组	含义
清算会员现金	会员向 CCP 提交的 HKD
客户资金负债	会员或清算体系对客户资金的负债
多头价款占用	已成交但尚未最终交割的全额价款
供应商待结算款	最终交付完成后应付供应商的服务价款
履约担保负债	受限制的供应商担保资源
CCP 风险资源	KAI 自有风险资本及共同违约基金
费用收入	已满足确认条件的交易、清算或交割费用

违约瀑布

发生资金或交付违约时，按规则依次使用：

1. 违约客户或供应商的可用余额、被冻结价款和履约担保。
2. 对其负责的违约清算会员担保及其他可用资源。

3. KAI 预先承诺的自有风险资源。
4. 共同违约基金。
5. 规则定义的追加分摊、保险或外部信用增级。

服务补救优先寻找合格替代容量；其他情形按规则执行退款、直接损失补偿和责任上限。CCP 的责任范围由清算规则、交割规则和会员协议共同界定。

实物交割

最终持仓在关闸后转换为真实 Chat Completions 权利和 Provider 服务义务

容量登记

供应商首先提交版本化能力声明，包括模型 digest、区域、上下文、可持续 TPM、并发、延迟、维护计划和证据。KAI 核准后形成 `CapacityGrant`。

容量维度允许新增净空头的最大手数为：

$$C_{\text{free}} = \max(0, C_{\text{approved}} - C_{\text{reserved for orders}} - C_{\text{committed short}} - C_{\text{locked delivery}})$$

C_{approved} 是已核准容量；系统依次扣除已为挂单预留、已支持成交净空头及已锁定交割的容量。外层 $\max(0, \cdot)$ 把结果下限固定为零，防止任何状态组合产生负的可用容量。

$$N_{\text{shortable}} = \left\lfloor \frac{C_{\text{free}}}{1,000,000 \text{ TPM}} \right\rfloor$$

$N_{\text{shortable}}$ 是容量维度允许卖空的完整手数。系统把可用容量除以每手所需的 1,000,000 TPM 并向下取整，因此不足一手的余量不能支持新增净空头。实际新增净空头再取容量、履约担保和持仓限额三者支持手数的最小值。同一能力对应唯一登记与承诺，成交后进入受约束状态；新增、减少、到期和冻结都通过版本化事实记录。

关闸与最终分配

关闸按以下顺序执行：

1. 停止该品种新订单并取消规则要求清理的剩余订单。
2. 锁定最终多头、供应商净空头、会员及控制组。
3. 冻结模型、区域、计量、基本交付标准和价款规则版本。
4. 按供应商净空头与合格容量生成交付池。
5. 将多头使用权绑定至账户和 KAI Chat Completions 交割凭证。
6. 按确定性规则建立 Provider 优先级与备用安排。

KAI 统一管理 Provider 身份与路由。供应商接收完成服务所需的最小请求字段，买方名称、账户余额、交易价格和其他供应商商业数据保持隔离。

Chat Completions 请求与路由

交割小时内，最终多头只通过 `POST /v1/chat/completions` 请求模型服务。KAI 不以 Responses、Assistants、Realtime、legacy Completions 或自定义推理接口交割。KAI 网关验证多头权利、窗口容量、上下文上限和规则版本；路由决策先作为已提交事实记录，再以相同 Chat Completions 契约向 Provider 发送请求。

Provider 接受前明确失败可以改路由；接受结果未知时按 KAI 请求 ID 查询。Provider 接受后保持单一路由，尤其在已经输出 Token 时沿原请求恢复。取消采用尽力而为语义，已经产生的合格使用量继续按规则计量。

窗口计量

每个连续、不重叠的 60 秒窗口独立计算：

$$TPM = I + W \times O$$

I 与 O 分别是窗口内经确认计费的输入和输出 Token， W 是该品种固定的输出权重。系统先把输出用量乘以权重，再与输入用量相加，所得 TPM 才与该窗口的容量上限比较。

每个请求至少保存接受时间、窗口、模型版本、输入输出 Token、缓存与思考项、首 Token 延迟、结束状态、Provider 报告、平台观测和证据哈希。原始观测永久保留，规范化和更正以新版本追加。

每个窗口使用独立容量预算。聚合使用达到上限时暂停新增请求；排队、降速或实时补购按合约已启用的机制执行。

对账、争议与结算

交割结束后依次：

1. 停止接受新的合同内请求。
2. 等待已接受请求完成或达到规则时限。
3. 汇总每个窗口与 Provider 的 TPM、错误率和延迟。
4. 比较 Provider、KAI 和可验证客户证据。
5. 形成通过、不足、争议或违约结果。
6. 结算供应商服务价款、费用、担保释放和违约分录。

争议通过追加版本处理，原始计量和初始对账持续保留。错误模型、错误区域、未披露限流、容量不足或严重质量偏差按规则触发替代交付和违约处置。

价格、流动性与商业模式

价格由真实订单、清算约束和最终交割能力共同形成

价格形成

期算价格综合反映模型服务成本、时段稀缺、区域约束、基本交付标准、替代能力、履约担保和流动性。供应商卖单表达未来交付的最低可接受价格；买单表达企业锁定能力的最高愿付价格；连续成交形成可审计期限曲线。

优秀流动性意味着：

- 企业可以在合理价差内建立或减少真实多头。
- 供应商可以出售真实容量并管理净空头。
- 临近交割仍有合格替代能力。
- 深度来自独立参与者而非关联账户循环成交。

市场成功以真实交付、独立成交和公平价格为基础，成交量与这些质量指标共同评价。

多小时组合

每个交割小时都是独立合约。品种启用“条带订单”时，连续四小时、工作时段或整日需求由多个小时合约组成，并在所有子合约同时满足价格、资金和容量条件时整体成交；未启用条带订单的品种分别管理各小时合约。

指数与市场数据

在积累足够独立成交和交割数据后，KAI 可以发布模型容量基准指数、峰谷指数、区域价差和交付标准价差。指数规则必须公开样本资格、异常处理、关联交易排除、修订和治理程序。

实时深度、成交和私有订单回报具有不同访问级别。延迟行情带有清晰标签；再分发数据保留时间、来源和使用条件。

收入结构

KAI 的收入可来自交易与清算，交割、风险与会员服务，以及历史、深度和指数数据。收费依据分别是撮合与 CCP 管理、Provider 接入与持续监督、数据生产与服务；收入与真实采购、清算和交割质量共同增长，并持续维护准入标准和资金安全。

市场诚信与运行风险

交易所必须同时保护公平顺序、资金安全、交付能力和系统连续性

市场监察

KAI 需要实时和事后识别：

- 自成交、关联账户对敲、虚假流动性和分层挂单。
- 临近关闸操纵、异常撤单、价格带试探和挤仓。
- 供应商重复登记、超卖、隐瞒维护或跨渠道重复承诺。
- 会员绕过控制组、持仓限额或客户授权，或利用未公开订单、交割故障和客户需求进行关联交易。

调查保留订单全生命周期、会员、控制组、规则版本和精确接受序列。警告、限额、暂停、取消资格和违约处置都须有规则依据及审计记录。

持仓与集中度

品种规则分别设置客户、控制组、交易会员、清算会员、供应商和关联组限额。接近阈值时触发预警，超过硬限额的新开仓拒绝。大额持仓和供应商集中度按规则记录。

临近关闸时限额可以收紧。普通账户的最低净持仓为零；供应商净空头上限取核准能力与履约担保支持手数的较小值。

电子交易控制

会员为算法交易、直接市场接入和自动路由负责，并实施订单限额、价格检查、重复防护、紧急停机、测试、版本审批和人员监督。KAI 按连接限制速率、在途订单、消息和订阅，并将慢客户端与核心撮合隔离。

运行连续性

交易、清算、资金、交割和市场监察数据须有恢复目标、备份、灾备演练和人工接管。重大故障期间优先保护：

1. 每笔成交均有完整证明。
2. 已成交事实持久、唯一。
3. 客户资金和会员抵押品可重算。
4. Provider 已接受请求保持单次执行。
5. 市场状态、恢复时间和客户行动得到及时通知。

GLM-5.2 完整交易示例

一笔机构期算如何从订单簿进入 CCP、平仓、关闸和实物交割

示例边界

以下模型、主体、价格、日期和参数均为概念示例；现实参数以届时生效的产品与服务文件为准。

合约

参数	示例
品种	GLM52-HKG-20260918-20
模型	GLM-5.2，以规则登记 digest 为准
交割小时	2026-09-18 20:00–21:00，Asia/Hong_Kong
区域	香港
一手	每个窗口 1,000,000 TPM
输出权重	$W = 3$
tick size	1 HKD/手
交割关闸	2026-09-18 19:55

挂单与成交

东海智算和北辰云均为交割参与者：

- 东海智算已核准 40 手，挂出 6 手卖单，780 HKD/手。
- 北辰云已核准 30 手，挂出 4 手卖单，785 HKD/手。
- 星河电商经交易会员提交 10 手限价买单，最高 785 HKD/手。

订单簿先按 780 成交 6 手，再按 785 成交 4 手。星河电商总成交价款为：

$$6 \times 780 + 4 \times 785 = 7,820 \text{ HKD}$$

KAI 原子锁定星河电商的全额价款和费用，记录两家供应商的净空头、容量和履约担保，并成为双方中央对手方。

平仓与转让

活动前，星河电商只需 9 手，于是以 800 HKD/手卖出 1 手。由于它已有 10 手多头，该卖单将多头减至 9 手，净持仓保持为正。成交后：

- 星河电商持有 9 手最终多头候选头寸。
- 买入该手的另一机构取得 1 手多头。
- 星河电商按其平均开仓价与 800 HKD 成交价确认已实现价差。
- 两家供应商的合计净空头仍为 10 手；空头总交付责任持续与市场未平仓量一致。

如果北辰云之后买入 1 手平空，其净空头由 4 手降至 3 手，相应容量和履约担保释放。系统按成交后的正净持仓总和重算未平仓量；该笔交易是否使其下降，取决于卖方是在减少已有多头还是建立新的供应商净空头。

关闸与交割

19:55 关闸后，KAI 锁定最终多头和供应商净空头，建立交付池并向账户发放 KAI Chat Completions 容量。请求先记录路由决定，再发送给 Provider。每个窗口独立核算 TPM，窗口结束时容量同步到期。

如果北辰云发生故障，KAI 先查询已接受请求状态并保持单次执行；仍处于路由前的请求可以切换到备用供应商。额外成本先从北辰云资源及其清算责任链中承担，再按违约瀑布处理。交割完成后，KAI 结算供应商服务价款、费用、担保释放和任何违约分录。

运营门槛与成功标准

每项市场能力都必须在真实资金、流动性、交割和运行证据完备后启用

能力启用门槛

KAI 只启用已经满足对应完成标准的能力：

能力	启用条件
单品种交易	规则、会员、订单簿、CCP、全额担保和 Provider 交割完成端到端认证
条带与做市	子合约原子性、资金与容量约束、报价义务和异常处置通过验证
备用容量与指数	独立成交样本、替代成功率、数据质量和治理程序达到门槛
新模型或区域	每个新增品种独立通过流动性、清算、运营和交割评估

任何能力都不得以产品扩展为理由削弱订单类型、清算、交割、监察或客户资金保护。

成功指标

维度	指标
真实需求	机构重复采购率、交割容量用于真实业务的比例
流动性	有效价差、深度、成交影响、平仓成功率
清算	资金足额率、未解决差异、会员违约与恢复时间
交付	足额交付率、首 Token 延迟、错误率、替代成功率
集中度	单一供应商、会员和控制组持仓占比
市场诚信	自成交、异常撤单、操纵调查和规则违规
运行	可用性、恢复目标、账本重算和事件重放一致性
商业	每手收入、交割成本、客户获取成本和留存

结论

KAI 的价值来自交易、清算与交割共同形成的完整市场基础设施

期算把未来模型服务能力转换为标准化、可连续交易并最终实物交割的机构期货。完整订单簿形成价格，交易会员组织客户接入，KAI 通过全额担保和中央对手方控制信用风险，交割体系把最终持仓转换为真实模型服务。

这一设计坚持四条底线：

1. 价格必须来自公平、有序、可审计的真实订单。
2. 多头资金和供应商净空头能力必须在成交前得到证明。
3. 最终价值必须落到合格模型服务交割。
4. 交易、清算与交割约束必须在上线前通过完整验证。

最终判断

KAI 是一座完整、全额担保、以实物模型服务为锚的机构期货交易所。

附录：术语表

供产品、会员、清算、交割和技术团队使用的统一词汇

术语	解释
TPM	每分钟加权 Token 负载， $TPM = I + W \times O$
期算	未来指定小时的标准化模型服务容量期货合约
中央限价订单簿	依价格—时间优先组织订单和成交的权威簿
开仓	增加多头或供应商净空头
平仓	通过反向成交减少已有头寸
未平仓量	尚未平仓且等待最终交割的合约总量
平均开仓价	当前同方向持仓的加权平均成交价
标记价	用于风险和价格保护的版本化参考价；结算遵循成交与交割规则
全额担保	多头覆盖最坏价款，净空头覆盖容量及履约担保
履约担保	支持供应商替代采购、质量和集中度风险的受限资源
交易会员	代表自身或客户向交易所提交订单的会员机构
清算会员	对 KAI 承担资金、抵押品和交收责任的会员机构
交割参与者	可以登记容量并建立供应商净空头的机构
控制组	在持仓、频率和自成交控制中统一聚合多个账户或技术子账户的标识
CCP	成交后成为所有多头卖方及所有空头买方的中央对手方
交割关闸	停止交易并锁定最终权利、义务和规则版本的时点
实物交割	通过实际模型推理服务履行期算合约