ITEM #214 - A Next-Generation X-Features IR & Interaction Architecture: Structural Anchoring, External Events, and Controlled Evolution

Conversation：IR 层生成算法改进

20260106

Authors: Sizhe Tan & GPT-Obot

Abstract

This document formalizes a next-generation architectural skeleton for the X-Features IR subsystem within the Digital Brain Model – Chain of Thought (DBM-COT) framework.

Based on concrete evidence from an existing, production-capable implementation, we identify that the system has already evolved an implicit Intermediate Representation (IR) contract, but lacks an explicit structural boundary. This absence has led to complexity inflation, DataManager centralization, Excel-style dominance risks, and difficulty in integrating external anchored events such as earnings reports.

We propose a minimal, non-negotiable IR core consisting of **TSChannel**, **EventIR**, and **Tierwise Context Ladder**, along with a unified Operator and Interaction model. The design preserves existing capabilities while establishing long-term structural invariants, enabling safe evolution, external event interaction, and downstream DBM reasoning.

1. Background and Motivation

The existing X-Features IR system demonstrates strong practical capability:

- JSON-driven, replayable operation pipelines
- Multi-level event generation (numeric, structural, compositional)
- Rich domain coverage in financial time series
- High exploratory efficiency via Excel-like cut–calculate–paste workflows

However, these strengths coexist with increasing structural tension:

- The DataManager has become a de facto **source of truth, runtime, view, and serializer**
- Event context data is treated as auxiliary rather than first-class IR
- Spreadsheet-style tools tend to overtake the execution core
- External events (e.g., earnings) lack a principled integration point

The issue is **not implementation quality**, but **the absence of explicit IR contracts**.

---

2. Empirical Skeleton of the Existing System

Analysis of the existing codebase reveals three implicit but stable axes:

2.1 Execution Axis

JSON-described pipelines mapped to registered Java operations form a de facto DSL for computation graphs.

2.2 Event Generation Axis

Clear separation already exists between:

- Numeric sequence events (e.g., acceleration)
- Structural events (e.g., peak/valley)
- Composite / bundle events

These distinctions, however, remain confined to class hierarchies.

2.3 Fact Aggregation Axis

A centralized manager aggregates channels, sequences, events, and serialization responsibilities, creating unavoidable coupling.

---

3. Root Cause Analysis

The observed complexity is not caused by "too many features", but by a missing declaration of **what constitutes IR truth**.

Without a fixed IR contract:

- Auxiliary tools compete for authority
- Context remains non-actionable
- External events cannot be placed without distortion
- New features are forced into existing containers

---

4. The Minimal, Non-Negotiable IR Core

We define three invariant IR primitives:

4.1 TSChannel (Continuous Facts)

A TSChannel represents continuous, windowable, stream-capable numerical reality.
Raw and derived signals belong here.

4.2 EventIR (Discrete Facts)

An EventIR represents a discrete structural occurrence anchored in time.
It is **not** a numeric point and must not be interpolated.

4.3 Tierwise Context Ladder (Explanatory Structure)

Event context is not an untyped map, but a multi-tier, comparable, downstream-actionable structure.

IR = TSChannel ∪ EventIR (with Tierwise Context Ladder)
Everything else is tooling.

---

5. Operator Unification

Existing generators naturally map to a unified Operator family:

Existing Role	Unified Operator Concept
Sequence event generator	EventOperator

Existing Role	Unified Operator Concept
Bundle event generator	CompositionOperator
Numeric event logic	NumericEventOperator
Structural event logic	StructuralEventOperator

JSON should describe **Operator Graphs**, not execution details.

---

6. External Anchored Events and Interaction

6.1 Earnings and Similar Events

Earnings reports are:

- External
- Sparse
- Anchored at disclosure time
- Structurally summarized

They must **not** be converted into time series.

6.2 Interaction Operator

External events interact with market events through:

- Context enrichment
- Regime and influence window generation
- Attribution and explanation

They modify **interpretation**, not raw prices.

---

7. Proper Role of Excel / Spreadsheet-Style Tools

Spreadsheet-like computation is valuable but must be constrained.

Correct placement:

- **Scratchpad** for exploration
- **TraceSink** for visualization

Incorrect placement:

- Source of truth
- Execution authority
- Module exchange format

Excel is a workbench, not a warehouse.

---

8. Safe Evolution Path

Phase 0

Introduce EventIR and Tierwise Context Ladder as overlays.
Behavior remains unchanged.

Phase 1

Explicit Operator categorization.
Split DataManager responsibilities.

Phase 2

Add Interaction Operators and fully externalize spreadsheet tooling.

---

9. Long-Term Structural Test

For any future feature, ask:

Does it generate TSChannel, or generate/enrich EventIR?

If neither, it does not belong to the IR core.

Conclusion

This document does not propose a rewrite, but a **constitutional declaration**.

By fixing the IR contract, the X-Features system transitions from a powerful indicator engine into a structurally grounded intelligence substrate, aligned with the long-term vision of DBM-COT.

ITEM #214 - 新一代 X-Features IR 与 Interaction 架构:结构定锚、外生事件与可控演化

摘要

本文正式固化 DBM-COT 体系中 X-Features IR 子系统的新一代骨架设计。

基于一个已经可运行、可回放的真实工程系统，我们发现：系统事实上已经自然演化出一套 IR（中间表示），但该 IR 从未被显式声明为"不可侵犯的结构合同"。这一缺失导致了复杂度膨胀、DataManager 过度集中、Excel 思维反客为主，以及外生事件（如财报）无法被正确建模。

本文提出一个**最小且不可协商的 IR 核心**：
TSChannel、EventIR 与分层 Context Ladder，并在此基础上统一 Operator / Interaction 模型，在不否定既有代码价值的前提下，为未来长期演进确立结构宪法。

1. 背景与问题动机

现有 X-Features 系统具备显著优势：

- JSON 驱动、可回放的运算管线
- 多层级事件生成能力（数值 / 结构 / 组合）
- 对金融时间序列的高覆盖建模
- 类 Excel 的高效探索能力

但与此同时，也暴露出结构性张力：

- DataManager 同时承担事实源、运行态、视图与序列化职责
- Event 的 Context 仅作为参考信息存在
- 表格工具极易夺取主线地位
- 财报等外生事件缺乏稳定落点

问题不在于"代码写得不好"，而在于 **IR 合同从未被钉死。**

---

2. 现有系统的真实骨架（基于代码证据）

2.1 执行主线

JSON 在系统中已实际承担 **Pipeline 描述语言（DSL）** 的角色。

2.2 事件生成主线

系统已经明确区分：

- 数值型事件
- 结构型事件
- 组合事件
  但这些区分仅存在于类层级，而非 IR 合同层。

2.3 事实汇聚主线

集中式 DataManager 成为复杂度膨胀的物理根源。

---

3. 根因判断

系统复杂度的根因并非功能过多，而是：

IR 的"事实边界"从未被明确声明。

---

4. 不可侵犯的 IR 核心定义

4.1 TSChannel（连续事实）

表示连续、可窗口化、可流式的数值世界。

4.2 EventIR（离散事实）

表示一次已经发生的结构性事实，锚定时间但不连续。

4.3 Tierwise Context Ladder（解释结构）

事件上下文必须是分层、可比较、可下游推理的结构。

IR = TSChannel ∪ EventIR（携带 Context Ladder）

---

5. Operator 的统一与重命名

既有 Generator 自然映射为统一 Operator 家族，
无需推倒重写，只需显式收敛。

---

6. 外生事件与 Interaction

财报不是序列，不是 pattern，不应插值。

它们应被建模为 **External Anchored Event**，

通过 Interaction Operator 影响事件解释与权重，而非价格本身。

---

7. Excel / bigspreadsheet 的正确位置

- 可作为 Scratchpad 与 TraceSink
- 不得成为事实源或执行核心

一句话定锚：

Excel 是工作台，不是仓库。

---

8. 安全迁移路径

- Phase 0：结构覆盖，不改行为
- Phase 1：职责拆分，合同收敛
- Phase 2：Interaction 与工具外置

---

9. 长期判断标准

新功能是否生成 TSChannel，
或生成 / 增强 EventIR ？

否则，不属于 IR 核心。

结语

这不是一次重构，而是一次**立宪**。

通过明确 IR 合同，X-Features IR 将从指标系统，
升级为 DBM 体系中的**结构智能基石**。