GateRestart AI V3 合成磁迹数据字典

数据格式、字段语义与质量规则 | Data Contract v3
品牌叙事实验 · 合成测试环境
版本 v3 · 文档制作 2026-11-14 · 原始数据版本 `grv3-synthetic-20261114-r1`

> 本字典解释**已经交付且实际运行的数据软件**，不是拟议的磁力计硬件协议。配套下载是固定版原始 `synthetic_events_v3.csv`、`field_dictionary.json`、`cases_v3.json`、`data-manifest.json`、`report-schema.json`，全部保持源文件字节不变。

阅读目录

1. 数据集简介与不变契约
2. 15 列 CSV 字段清单
3. 单位、时标与行完整性
4. 缺失、饱和与时间状态
5. 12 案例场景及含义
6. JSON 报告结构分层
7. CSV/JSON 消费者接口规则
8. 版本迁移和校验步骤
9. 来源、版权与使用

1. 数据集简介与不变契约

V3 合成数据包有 12 个 `SIM-01` 到 `SIM-12` 情境，每案 601 个时点，共计 7,212 条数据行和一行表头。时间网格为 **-1.50 到 +4.50 秒，闭区间，步长 0.01 秒**；每案索引为 `0..600`。`unit` 恒为 `a.u.`（任意归一化演示单位）。这些数字是程序生成、用于比较质量规则的演示信号，不是校准过的磁场 nT/T 值。

契约项	固定值	读取提示
数据源	`synthetic_events_v3.csv`	原固定版文件，不另造曲线
案例数 / 总数据行	12 / 7,212	不包括标题行
每案点数	601	0 到 600，固定 0.01 s 间隔

契约项	固定值	读取提示
采样网格	-1.50…+4.50 s	两端均含，t=0 是模拟道闸事件
数值通道	bx、by、bz、reference	4 个独立演示字段，非实测 OPM
数值单位	a.u.	不得称为 nT
数据版本	`grv3-synthetic-20261114-r1`	每行标明，报告也包含同一版本
实验标志	`synthetic=TRUE`	CSV 字符串，不是现场真实性声明
编码/换行	UTF-8 无 BOM；原 CSV 为 CRLF	原文件按字节保留；不正规化

2. 15 列 CSV 字段清单

以下顺序与固定版 `field_dictionary.json.columns_in_order` 完全相同。数值空值在 CSV 中写成两个分隔符间的空字段，而不是数字 0。

顺序	字段名	类型 / 单位	解释
01	`scenario_id`	string	案例号，SIM-01 至 SIM-12
02	`sample_index`	integer	每案从 0 开始的样本序号
03	`time_s`	number / s	相对模拟门控事件的时间，含负数
04	`bx_au`	number 或空 / a.u.	目标 Bx 合成值；若饱和则为显示裁剪值
05	`by_au`	number 或空 / a.u.	目标 By 合成值
06	`bz_au`	number 或空 / a.u.	目标 Bz 合成值；饱和时不是可靠峰值
07	`reference_au`	number 或空 / a.u.	参考通道合成值；LOST 时为空
08	`bx_status`	enum	Bx 当前点的质量状态
09	`by_status`	enum	By 当前点的质量状态
10	`bz_status`	enum	Bz 当前点的质量状态
11	`reference_status`	enum	参考通道当前点的质量状态
12	`timebase_status`	enum	相对道闸零点的时钟映射状态
13	`unit`	string	固定 `a.u.`
14	`synthetic`	boolean_text	固定 `TRUE`，强调非实测
15	`data_version`	string	当前合成数据契约版本

举例说明一条真实存在于原文件的数据行：SIM-01 的 `sample_index=0` 对应 `time_s=-1.50`、`bx_au=0.012976`、`by_au=-0.023503`、`bz_au=0.005349`、`reference_au=-0.008546`，四个通道状态均为 `VALID`。这只是生成波形的一个采样点，不对应某辆实际汽车。

3. 单位、时标与行完整性

a.u. 是任意演示单位。 即使小数值看起来和纳特斯拉类似，也无法从数值大小推导磁场单位。将原 V2.1 `*_nT` 数据直接连到 V3 图表、把旧 0–3 秒时间窗与 V3 数据拼接，都会造成严重语义错误。旧版源文件保留在旧路径，新数据应在独立 v3 路径消费。

`time_s` 为相对值，不含日期、UTC、GPS 或系统时钟时间。t=0 是模拟道闸控制器事件标记，不是探头自动检测到起动机转动的时间。软件 `timebase_status` 的状态反映**模拟的时基条件**，不是实际设备的时钟校准报告。

每个场景内部应满足：`sample_index` 递增 0 至 600、`time_s` 等于 -1.5 + index × 0.01（按格式允差检查）、单位及版本一致、记录共 601 条、合成标志恒为 TRUE。如果缺行或重复索引，应在解析层报错而非通过插值补齐。

4. 缺失、饱和与时间状态

字段/状态	字面含义	渲染/计算要求
`VALID`	原记录点数值有效	可参与演示特征计算
`LOST`	原始值缺失	数值格为空，JS 中解析为 `null`；曲线断线
`SATURATED`	数字已被演示上下限截顶	可以展示裁剪形状，但不得作为真实幅值或峰值
`NOMINAL`	合成器按假设对齐时基	只是模拟条件，无硬件同步证明
`OFFSET_UNCERTAIN`	合成器设置时钟偏差未校准	不做精确时序归因

正确示例：若 `reference_status=LOST`，即使目标通道存在峰值，也必须将参考曲线中对应区间留空，并将需要依赖参考的判断置于弃判路径。错误示例：将空字符串执行 `Number('')` 得到零、将缺口填零以绘制连线，或者把 `SATURATED` 数值报告为 OPM 的真实磁峰强度。

SIM-12 特别用于证明缺口语义：参考通道有 **131 LOST 与 470 VALID**，缺失发生于 0.30–1.60 s；`decision.status=ABSTAIN`、`label=REFERENCE_LOST`。该事件不代表某种实体通信模块的真实失效率。

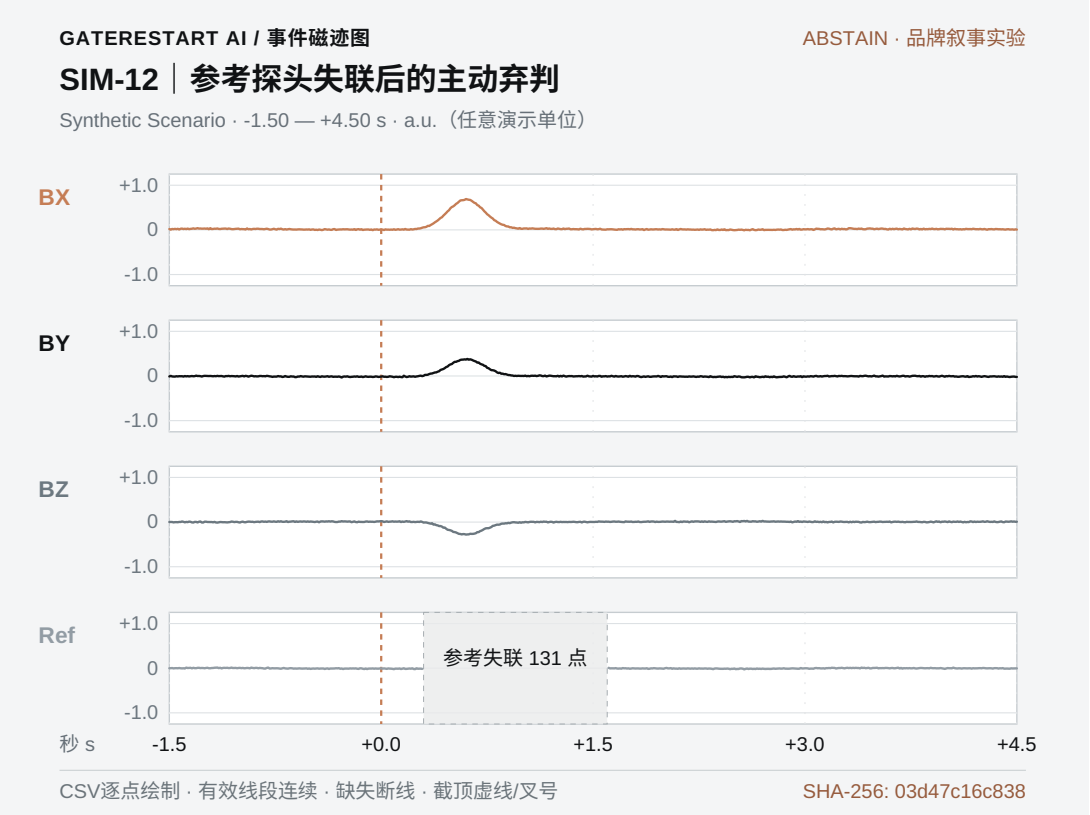


图 1｜SIM-12：原始 V3 SVG，依据同一 CSV 全样本生成。缺失段可视化必须断开。

5. 12 案例场景及含义

下表标签属于既有情境数据，描述生成器预设的教学问题，不能作为从波形反演所得的真实场源事实。

案例	情境名称	阅读重点
SIM-01	单次基准重启磁脉冲	合成单峰展示，而非健康起动机标准
SIM-02	额外脉冲形态	多峰形态与规则边界
SIM-03	道闸电机的共同磁扰动	两通道共扰，不把时间相关当归因
SIM-04	相邻车辆重启的事件重叠	多来源无法从单站唯一划分
SIM-05	抬杆却没有发动机重启	避免仅凭放行时间声称有起动
SIM-06	重启后的延迟磁脉冲	时间窗及延时假设
SIM-07	坡道起步中的车体运动磁扰动	目标以外的运动磁变化
SIM-08	钢结构导致的背景场缓慢漂移	比较窗口前后基线
SIM-09	OPM 观测通道超量程假设	裁剪值不可用于峰幅解释
SIM-10	金属维护推车经过探头	场源多样性与拒绝归因
SIM-11	道闸时间戳与磁迹不同步	记录偏移不等于精确因果证据

案例	情境名称	阅读重点
SIM-12	参考探头失联后的主动弃判	原始缺失保持空值并弃判

上述名称仅为读者定位用中文简述；编程时以原始 `cases_v3.json` 的字段、`SIM-XX.json` 的具体报告结果与 `decision.label` 为权威。特别固定：**SIM-03 = 闸机电机干扰；SIM-04 = 邻车叠加**，不可交换。

6. JSON 报告结构分层

原始报告采用 JSON Schema Draft 2020-12，标识 `report_schema_version=3.0.0` 和 `analysis_version=gr-edge-synthetic-3.0.0`。报告保留以下一级字段，不允许在现有文件上悄悄增删、重命名或以生成器假设覆盖数值特征。

一级对象	作用	主要判读原则
`report_id`、`scenario`	报告编号、场景 ID/标题/路径	场景文本是虚构背景
`synthetic`、`synthetic_label`、`unit`	合成身份及量纲	只能解释为 a.u. 演示
`provenance`	CSV、casebook、manifest 哈希及种子	支持数据来源追溯，不证明实测
`selection`	分析窗、源采样数、实际选用数	与窗口裁切契约一致
`assumptions_from_synthetic_generator`	注入的场源类别、时基、参考缺口	**生成器预置的条件**，不是算法识别
`features_computed_from_csv`	覆盖率、基线、峰值、差分、相关等	**从数值和状态读取/计算**，不是实测诊断
`decision`	RENDER_EVENT/ABSTAIN 及原因	规则选择结果，没有真实诊断置信概率
`narrative`	情境、结果与学习目的	需保留其“合成”语境
`cannot_infer`	明确不可推出的现实结论	不能删除后再发布为性能报告

决策相关字段 `is_trained_diagnostic_model=false`、`real_vehicle_health=null`、`diagnostic_confidence=null` 明确表达了原分析器的范围。与公司故事有关的产品和团队资料不应被注入 `features_computed_from_csv`，避免破坏可复核性。

7. CSV/JSON 消费者接口规则

1. **分组。** 按 `scenario_id` 分组，逐组要求 601 行和连续索引；不利用未标注的文件顺序推断案例身份。
2. **解析。** 数值字段仅在非空且状态允许时转换；空值返回 `null`，不把空字符串变成零。
3. **绘图。** Bx/By/Bz/reference 分开上图，所有 LOST 处断线，SATURATED 段可用特殊线形；图注明确 a.u.。
4. **计算。** 仅使用两侧同时 VALID 的采样对执行目标/参考差分；记录被排除点数。
5. **来源。** 场源类型仅在 `assumptions_from_synthetic_generator` 解释；若 UI 显示“闸机电机/邻车”，应标明是合成情境名称。
6. **报告。** 通过 `report-schema.json` 验证，报告不可抹去 `cannot_infer` 或将弃判变成“异常故障”。
7. **版本。** 对任何导出保留 `data_version`，并记录对应原始 CSV 的 SHA-256；无版本数据应隔离。

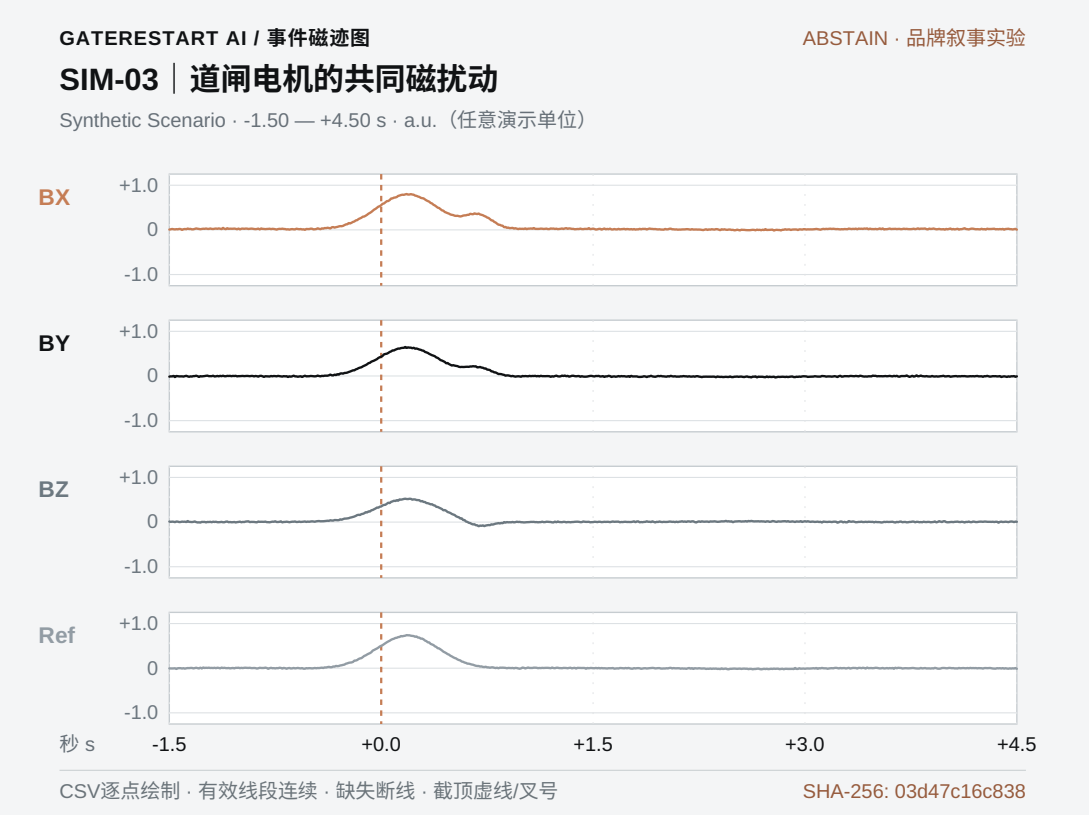


图 2｜SIM-03 的源数据渲染支持共同变化观察，但“闸机电机”仅为注入源情境。

8. 版本迁移和校验步骤

原始 CSV SHA-256： `03d47c16c8384da444303ac0ab82180cc7b6d504643276133a8fbfd86ec6abd9`。随包的 `data-manifest.json` 还列出 casebook、字典、生成器和各案例状态计数。在终端中可用 `sha256sum` 或 Python `hashlib.sha256` 对下载文件复核。ZIP 附带的 `VERIFY.md` 给出资料侧检查项目；它不是本地 Windows 或实际浏览器的验收结论。

旧版五案例 `*_nT` 列不能直接迁移为 `*_au`；若未来确有校准后真实磁测数据，应建立新 schema 与校准来源，而不是篡改这份合成数据。`SIM-01..SIM-05` 在旧 V2.1 和新 V3 可能具有同名编号，但数据定义、采样窗与版本不同，应以版本和根路径共同定位。

9. 来源、版权与使用

唯一数据权威： 用户提供的 `gaterestart-v3-data-fixed(1).zip`，其 `field_dictionary.json`、`report-schema.json`、CSV、casebook 和报告按字节保留。本字典未执行新的波形生成，也未重写任何分析器。叙事品牌身份与产品名取自独立 `brand-bible-v3.json`。

科学背景只限 `science-sources.json` 指向的四项实际公开来源：[S1] https://quspin.com/technology/；[S2] https://quspin.com/products-qzfm/；[S3] https://quspin.com/applications/；[S4] https://doi.org/10.3390/s20164563。它们说明通用原理/特定厂商与实验条件，**不对该 CSV 的物理真实性作背书**。

使用说明： CSV 可按现有 schema 复核、绘图、制作演示报告；引用需保留 `synthetic=TRUE`、`a.u.` 和版本。不允许将数据当作实车磁测或发出设备性能承诺。第三方文献仍依其许可；文档制作日期 2026-11-14 不代表网站已上线。