

Contents

认知计算协议（COP）：公开征评版 v1.0	2
COP.B01 理论定义、相变机制与理论溯源	2
一、核心命题	2
二、承接行为的三个必要条件	3
三、COP不处理的问题	3
四、COP在LMM十步中的位置	3
五、COP的通用性	3
六、关键概念定义	4
七、相变机制	4
八、理论溯源	5
九、本文档与其他文档的关系	6
COP.B02 • 五协议主文	7
总论：五协议的关系	7
P1 • 输入采样协议	7
P2 • 状态编码协议	8
P3 • 诊断判定协议	9
P4 • 转介与有限干预协议	10
P5 • 反馈学习校准协议	10
跨协议纪律	11
COP.B02-P1 • 输入采样协议详细版	11
一、P1 在诊断管线中的位置	11
二、采样帧完整规格	12
三、四种信号采集通道	12
四、采样偏倚声明规范	13
五、最低采样纪律	13
COP.B02-P2 • 状态编码协议详细版	13
一、P2 在诊断管线中的位置	13
二、编码流程：两步走	13
三、COP 核心编码	14
四、编码输出完整结构	14
五、编码不做什么	15
COP.B02-P3 • 诊断判定协议详细版	15
一、P3 在诊断管线中的位置	15
二、五种分流状态的完整定义	15
三、判定矩阵（完整版）	16
四、判定纪律	17
COP.B02-P4 • 转介与有限干预协议详细版	17
一、P4 在诊断管线中的位置	17
二、五种分流 → 动作映射	17
三、转介路由详细规则	18
四、COP 可以做与不可以做	19
COP.B02-P5 • 反馈学习校准协议详细版	19
一、P5 在诊断管线中的位置	19
二、校准数据采集时点	19
三、三类误判的详细校准	20
四、误判样本库规范	21

五、校准版本管理	21
六、校准纪律	21
COP.B03 • 诊断走通示例：从采样到校准	22
场景	22
P1：输入采样	22
P2：状态编码	23
P3：诊断判定	23
P4：转介与有限干预	24
P5：反馈校准（假设的后续）	24
走通总结	24
COP 诊断协议 v1	25
一、继承合同	25
二、输入合同	25
三、输出合同	26
四、冻结与转介规则	27
五、复核与争议处理	28
六、误判成本	28
七、复核一致性	29
八、版本回滚门槛	29
九、数据边界	30
十、版本与回滚	30
十一、最短使用顺序	31
十二、最短压缩句	31

认知计算协议（COP）：公开征评版 v1.0

Cognitive Operations Protocol: Public Review Edition

作者：Yi Fu（付毅，ODDFounder）

定位：复杂现实进入结构诊断后形成的诊断压缩+风险分流+转介协议层。

ewpage

COP.B01 | 理论定义、相变机制与理论溯源

承接协议理论 • 理论总门 版本：V1.1 | 状态：现行 | 日期：2026-04-16

一、核心命题

承接的本质，是把用户对自己处境的解释权，从旧框架迁移到新框架。解释权迁移完成，入泊自然发生。

推论一：承接不是等待，是触发。触发的对象不是用户的需求，而是用户的认知结构。

推论二：对被灯塔筛过的有效用户，若识认序列完整完成，则大概率发生主动消费。

推论三：付款是解释权迁移之后的自然结果，不是承接行为的设计目标。

二、承接行为的三个必要条件

COP适用于一切“承接行为”。承接行为的成立，需要同时满足以下三个条件：

条件	定义	缺失后果
激活主体 承接锚点 跨越动作	对方已存在属集张力，正在寻找解释或出路，需在上游（灯塔层）解决 新框架有具体的附着位置，用户能够发生定位且无处停留，入泊失败 主体主动迁移，非被推入、被诱导	解释或出路，需在上游（灯塔层）解决 用户能够发生定位且无处停留，入泊失败 入泊成立但信任基础薄弱，后续转化成本极高

判断原则：三个条件缺少任意一个，COP的机制均不能完整启动。条件缺失时，应向上追溯至相应层级修复，而非在COP层强

三、COP不处理的问题

COP的边界是清晰的。以下问题不属于COP的工作范围：

上游问题（灯塔层负责） - 如何吸引陌生流量 - 如何让目标用户第一次看到内容 - 如何过滤解法开放度极低的用户

下游问题（泊位内层负责） - 入泊后如何深化信任 - 产品交付质量 - 复购与口碑的形成机制

不可控变量 - 用户的解法开放度（由用户内部结构决定，COP只能有限影响） - 用户的行动窗口（时机问题，COP通过留存设

四、COP在LMM十步中的位置

灯塔层

| 内容吸引 → 激活有效用户 → 筛出属集紧张状态人群



COP层（承接协议）

| 入口触达 → 相变触发 → 解释权迁移 → 入泊完成



泊位内层

| 信任深化 → 识认序列完成 → 主动消费 → 大概率付款

COP是灯塔层和泊位内层之间的相变层。它不产生流量，也不负责最终转化，它负责的是让有效用户从外部进入内部的这个跨

五、COP的通用性

COP不绑定任何特定产品、内容形式或行业。以下场景均适用：

- 知识产品（文章、课程、付费报告）
- 咨询与教练服务
- 社群入口设计
- 品牌与用户关系的建立
- 任何存在“说服他人主动迁移”需求的场景

通用性的前提：三个必要条件均成立，且承接方能够提供一个真实的新框架（不能是空洞的口号或虚假的解决方案）。

六、关键概念定义

解释权（Interpretive Authority） 个体对自身处境的解释所依赖的框架体系。每个人在任何时刻都持有某种解释框架——即便这个框架是错误的、模糊的或令其痛苦的。COP的目标是以新框架替代旧框架，完成解释权的迁移。

属集张力（Attribute-Set Tension） 用户现有认知框架与其正在经历的现实之间持续存在的张力。张力越大，相变越容易触发。

入泊（Docking） 用户带着主动意愿从外部进入承接方系统的跨越动作。入泊的标志是用户的主动深入行为——不是被推进来，而是自己走进来。

相变（Phase Transition） 用户认知结构在某一刻发生的质变。不是渐进的说服积累，而是在特定触发条件下的突然重组。

七、相变机制

7.1 相变的本质

认知相变不是说服，不是积累，是用户属集结构在某一刻发生的质变。

物理意义上的相变不是慢慢接近，而是在临界点上的突然转换。认知相变的逻辑与此相同：用户不是被一点一点说服的，而是在重组之前，他是访客。

重组之后，他是入泊者。

在 **ASTO** 框架中，个体的认知状态由其当下持有的属集结构决定。正常情况下，属集结构稳定，外部信息进入后被现有结构消解。

7.2 相变的前提条件

触发相变需要三个前提条件同时成立：

1. **属集紧张状态** 用户的现有认知框架与其正在经历的现实之间，已经存在持续张力。
2. **失败的自我解释积累** 用户已经试图理解自己的处境，但反复失败。
3. **情境激活** 用户的属集张力在接触点上被重新唤起。

若这三个前提缺一，

则应回到上游修复，

而不是在 **COP** 层强推。

7.3 三个结构变量

在任何承接场景中，触发相变的材料和方式都需要通过以下三个变量来确定：

变量	问题	典型取值
张力类型	用户的张力来自哪里	外部压力型 / 内部矛盾型 / 认知缺口型
旧框架形态	用户现在用什么框架解释自己	归因于自己 / 归因于他人 / 归因于环境
命名缺口	用户的困境被命名到什么程度	完全未命名 / 模糊命名 / 已被命名 / 过度命名

7.4 两阶段共鸣

相变通常经历两个阶段：

1. **认出自己（冷共鸣）** 用户判断“这在说我这个人，不是在说一类人”。
2. **被理解（热共鸣）** 用户感到“终于有人懂了我为什么会这样做”。

真正的相变发生在第二阶段，
不是第一阶段。

7.5 可观测信号

相变无法直接观测，
但可以通过主动搜索自身记忆、主动追问、主动提供更多细节、主动进入下一步等行为来判断。
核心判断原则是：

相变发生的标志，是用户开始主动在自己的记忆里搜索，并用新框架重新解释自己的过去。

7.6 不发生的三类原因

原因类型	表现	修复方向
前提缺失 触发器失效 解法开放度极低	用户本身没有属集张力，或上游修复精度不足 描述泛化、批判选择、防御性细节精度，移除判断 用户有识认，但不相信任何外部方COP 处理范围	向上游修复 提高细节精度 移除判断 扩大COP 处理范围

八、理论溯源

8.1 推导链总览

COP 不是独立产生的工具理论，
它从以下链条自然推出：

DM -> ASTO -> TAT -> LMM -> COP

每一层都是上一层在新维度的应用，
不是并列关系，
而是继承与延伸关系。

8.2 DM -> ASTO

DM 提供本体论基础：

差异是存在的根本形式。

ASTO 提供直接支撑：

个体通过持有的属集来感知和解释现实，
属集可以变迁，
相变就是属集变迁的激变形式。

因此：

- 用户的旧框架 = 当下属集结构
- 属集张力 = 属集结构与现实的张力
- 相变 = 属集结构突然重组
- 解释权迁移 = 新旧稳定态转换

8.3 ASTO -> TAT

TAT 把抽象张力具体化为组织责任处境。

例如：

- 责任向下穿透、权力向上截断
- 背锅与责任闭合失败
- 沉默的代价结构
- 贡献可见度低

这些都为 **COP** 提供了高价值原材料。

8.4 TAT -> LMM

LMM 负责：

- 激活用户的属集张力
- 过滤解法开放度极低的用户

也就是先把“船找到灯塔”这一步完成。

8.5 LMM -> COP

LMM 解决的是：

船能找到灯塔

COP 解决的是：

船能不能真正靠岸

因此 **COP** 的理论必要性在于：

有效用户到达接触点之后，仍然需要一套机制触发其主动入泊；这套机制不是说服，而是解释权迁移触发的认知相变。

8.6 理论创新点

相较于上游理论，**COP** 的独立贡献主要在于：

1. 将相变确立为承接的核心机制
2. 提出两阶段共鸣理论
3. 建立三变量框架
4. 定义入泊的可操作性边界

8.7 当前待解问题

当前版本尚未完整解决的问题包括：

1. 解法开放度的可影响性
2. 相变的可逆性
3. 多框架并存
4. 集体相变

这些问题继续由 **COP.B04** 作为活档案持续记录。

九、本文档与其他文档的关系

文档	与本文档的关系
COP. R01 《接触点操作手册》	本文档核心命题的操作化实现
COP. B0AI 《AI调用协议》	本文档的结构化压缩版，供AI直接调用
COP. B04 《验证与迭代记录》	本文档待解问题、修订与验证的持续活档案

本文档为COP理论体系的理论总门，建议所有读者首先阅读本文档。

ewpage

COP. B02 • 五协议主文

定位：COP 核心协议层——五份协议各自独立、互相衔接，共同构成”从复杂现实到可计算诊断”的完整管线。
 状态：现行母文件 作者：Yi Fu（付毅，ODDFounder）

总论：五协议的关系

五协议不是并列的五个功能模块，而是一条诊断管线：

输入采样 → 状态编码 → 诊断判定 → 转介与有限干预 → 反馈学习校准
 ↑ |

校准结果回流到采样策略

每一步把前一步的输出压得更”可计算”：

协议	输入	输出	核心问题
P1 输入采样	原始现实信号	结构化采样帧	我们看到了什么
P2 状态编码	采样帧	编码向量 + 结构标签	这属于什么类型的状态
P3 诊断判定	编码向量	分流状态 + 置信度	当前是否清楚、是否安全
P4 转介干预	分流状态	动作建议 + 转介路由	接下来该做什么
P5 反馈校准	真实结果	协议参数更新	我们上次判断对了吗

P1 • 输入采样协议

1.1 职责

把原始现实信号（自然语言描述、问卷回答、行为日志、指标面板）转换为结构化采样帧。

1.2 核心原则

- 不在此阶段做判断：采样只管”收进来”，不管”对不对”
- 信号源显式标注：每条采样记录必须标注来源、时间、采集方式
- 采样偏倚自报：采样策略必须说明自己更可能漏掉什么

1.3 采样帧结构

```
sampling_frame:
  frame_id: ""
  collected_at: ""
```

```

source_type: questionnaire | log | interview | sensor | document
source_desc: ""
signals:
  - signal_id: ""
    content: ""          # 原始内容, 不做改写
    modality: text | numeric | categorical | binary
    reliability: high | medium | low | unknown
sampling_bias_note: ""   # 本帧最可能漏掉什么信号

```

1.4 最低采样纪律

1. 每条信号必须保留原始措辞，不做同义改写
2. 不知道的信号写”未知”，不编造
3. 采集方式（用户主动/系统日志/第三方报告）必须标注
4. 采样偏倚声明不得跳过

P2 • 状态编码协议

2.1 职责

把采样帧中的原始信号映射到 COP 的编码空间，输出编码向量 + 结构标签。

2.2 核心原则

- 先继承 ASTO 的结构编码：在编码前先确认”当前结构处于什么状态/节律/位点”，而不是跳过结构直接从表面标签起
- 分类集中度与结构风险分开编码：不要混成一个分数
- 编码不是解释：编码是”这更接近 X”，不是”这是因为 Y”

2.3 编码输出结构

```

encoding_output:
  encoding_id: ""
  source_frame_id: ""

  # ASTO 前置编码（必填）
  asto_precoding:
    state_type: ""          # 五态之一
    stage: ""              # 六阶之一
    sequence_type: ""      # 七序之一
    boundary_status: clear | blurred | contested
    exception_flag: false

  # COP 核心编码
  clarity_vector:
    signal_completeness: 0-1  # 信号完整度
    signal_consistency: 0-1  # 信号一致度
    category_concentration: 0-1 # 分类集中度

  structural_risk_vector:
    responsibility_diffusion: low | medium | high
    power_asymmetry: low | medium | high
    irreversibility_potential: low | medium | high

```

```
encoding_confidence: 0-1
encoding_note: "" # 编码过程中的不确定性和解释
```

2.4 最低编码纪律

1. ASTO 前置编码不得跳过——必须先标五态/六阶/边界
2. 分类集中度与结构风险必须分开输出
3. 编码置信度低于 0.6 时，必须写 `encoding_note`

P3 • 诊断判定协议

3.1 职责

基于编码向量和风险向量，输出分流状态和处置建议级别。

3.2 五种分流状态

状态	含义	触发条件	默认动作
RESOLVE	可自动处理	信号完整 + 一致 + 分类集中 + 结构风险低	允许继续自动推进
MIXED	混合信号	信号不一致 或 分类不集中	标注混合区域，有限推进
FREEZE	冻结	结构风险高 或 触及红线	暂停自动推进，升级人工
UNKNOWN	信息不足	信号完整度低	要求补充信息
REFER	转介	超出 COP 可处理范围	转介到 TAT / ODD / 人工

3.3 判定规则（简化矩阵）

信号完整度	分类集中度	结构风险	→ 分流状态
>0.7	>0.7	low	RESOLVE
>0.5	>0.5	low-medium	MIXED
any	any	high	FREEZE
<0.4	any	any	UNKNOWN
any	any	out_of_scope	REFER

注：这是概念级矩阵，实际判定权重由 `COP. 诊断协议.v1.md` 和 `COP. 校准样本协议.v1.md` 详细定义。

3.4 诊断判定输出

```
diagnostic_output:
  diagnosis_id: ""
  source_encoding_id: ""

  triage_status: RESOLVE | MIXED | FREEZE | UNKNOWN | REFER
  confidence: 0-1
  primary_label: ""
  alternative_labels: []

  freeze_reason: "" # triage_status=FREEZE 时必填
  referral_target: "" # triage_status=REFER 时必填
  missing_info_hint: "" # triage_status=UNKNOWN 时建议补充的信息
```

diagnosis_note: ""

P4 • 转介与有限干预协议

4.1 职责

基于分流状态，生成下一步动作建议和转介路由。

4.2 五种分流→动作映射

分流状态	动作类型	具体动作
RESOLVE	放行	允许进入 RT6 完整六步流程，附带诊断摘要
MIXED	有限放行	进入 RT6，但标注混合信号区域供步骤 2 额外关注
FREEZE	冻结 + 升级	停止自动推进；路由到人工复审或 TAT 评审
UNKNOWN	补充采样	向用户或系统请求指定类型的额外信息
REFER	转介	生成转介摘要，路由到指定目标（TAT/ODD/ECET/人工）

4.3 有限干预原则

COP 可以做： - 基于规则的有限建议（不涉及高风险强制动作） - 转介路由选择（根据问题类型匹配目标层）
- 信息补充请求（明确请求什么类型的信息）

COP 不得做： - 强制性高风险动作授权 - 行为锁定的最终裁决 - 同意、申诉、补偿结构定义（这些必须回引 TAT）

4.4 转介路由

问题类型	转介目标	转介内容
触及责任门槛	TAT	诊断摘要 + 风险向量 + 触发信号
契约/门禁/证据问题	ODD	诊断摘要 + 涉及的产出物引用
治理边界争议	ECET	诊断摘要 + 治理约束分析
COP 自身无法判断	人工复审	完整诊断链 + 无法判断的原因

P5 • 反馈学习校准协议

5.1 职责

收集诊断后的真实结果，对比诊断结论与实际情况，校准协议参数。

5.2 校准数据采集

每次诊断后，在以下时点采集校准信号：

时点	采集内容
即时反馈	用户/系统对诊断的反应
RT6 步骤 2 反馈	扫描诊断是否与 COP 诊断一致
RT6 步骤 6 反馈	复诊时发现的问题是否与初始诊断匹配

时点	采集内容
事故回溯	FREEZE/REFER 后被证实为正确的冻结（避免了一次损害）或错误的冻结（阻碍了本可进行的处

5.3 三类误判的校准

误判类型	定义	校准动作
False Safe	判定为 RESOLVE，实则不应推进	提高该类型的分类集中度阈值或结构风险敏感度
Over Freeze	判定为 FREEZE，实则本可推进	下调该类型的冻结触发阈值
Wrong Type	分流方向正确但分类标签错误	调整该信号模式的编码映射

5.4 校准输出

```
calibration_output:
  calibration_id: ""
  source_diagnosis_id: ""

  actual_outcome: ""          # 真实发生了什么
  diagnosis_accurate: true | false | partial
  misclass_type: false_safe | over_freeze | wrong_type | none

  parameter_adjustments:
    - parameter: ""          # 哪个参数需要调整
      direction: up | down
      magnitude: minor | moderate | major
      reason: ""
```

跨协议纪律

- 1. 协议不可跳过：不得从采样直接跳到判定（跳过编码），也不得从判定直接跳到校准（跳过转介）
- 2. 不安全的输出必须 FREEZE 或 REFER：宁可多冻一次，不可错放一次
- 3. 校准必须闭环：每次误判必须在误判样本库中留档
- 4. AST0 前置编码不可省略：这是防止 COP 退化为”表面分类器”的关键机制

ewpage

COP.B02-P1 • 输入采样协议详细版

定位：COP 诊断管线第一协议——如何把原始现实信号转成结构化采样帧。本文件是 B02 五协议主文中 P1 的完整展开。 作者：Yi Fu（付毅，ODDFounder）

一、P1 在诊断管线中的位置

输入采样(P1) → 状态编码(P2) → 诊断判定(P3) → 转介干预(P4) → 反馈校准(P5)

P1 是管线的入口。P1 出错，后面的编码和判定全部建立在错误或偏倚的信号上。

二、采样帧完整规格

```
sampling_frame:
  frame_id: ""          # 唯一标识
  collected_at: ""       # ISO 8601
  source_type: questionnaire | log | interview | sensor | document | observation
  source_desc: ""        # 人类可读的来源描述

# 信号通道
channels:
  - channel_id: ""
    modality: text | numeric | categorical | binary | image | audio
    signals:
      - signal_id: ""
        raw_content: ""      # 原始内容，不做任何改写
        collected_at: ""
        reliability: high | medium | low | unknown
        reliability_basis: "" # 为什么判断为这个可靠度

# 采样偏倚声明（必填）
sampling_bias_note: ""
# 示例：本帧信号全部来自用户主动填写的问卷，
# 最可能漏掉：用户自己未意识到的结构性问题、
# 第三方视角的信息、非自愿参与者的信号。

# 采样完整性自评
completeness_self_assessment:
  signal_density: sparse | moderate | rich
  blind_spots: []      # 已知的盲区
  recommended_supplement: "" # 建议补充什么类型的信号
```

三、四种信号采集通道

3.1 问卷通道

最常用，但最容易引入偏倚。

纪律： - 每条问题必须区分”事实题”和”感受题” - 必答题不超过 9 题（超过则用户疲劳 → 信号质量下降） - 至少 2 道开放题（提供非结构化信号）

3.2 日志通道

来自系统日志、行为记录。

纪律： - 每条日志必须标注是”主动记录”还是”被动记录” - 日志的缺失不等于事件的缺失——“什么都没记录”本身是一条需要标注的信号

3.3 访谈通道

来自对话、面谈。

纪律： - 原始措辞保留——不把”我觉得很累”改写为”用户报告疲劳” - 访谈者的提问本身也是信号来源——标注访谈者问了什么问题

3.4 文档通道

来自已有文档、报告、历史记录。

纪律：- 文档的写作目的和受众必须标注——一份给投资人看的报告和一份内部事故报告，其信号可靠性不同
- 二手引用必须标注——“据 XX 报告称”不能写成“事实是”

四、采样偏倚声明规范

每个采样帧必须写偏倚声明。以下是最常见的偏倚类型：

偏倚类型	示例声明
自选偏倚	“本帧信号仅来自主动参与的用户，不包含沉默者或退出者的信号”
通道偏倚	“本帧仅包含文字信号，不包含语音、表情、行为等非文字信号”
时点偏倚	“本帧采集于事件发生后 2 周，可能包含回忆偏倚”
权力偏倚	“本帧中的下属视角信号缺失——只有上级提供了描述”
工具偏倚	“问卷的问题框架可能暗示了某种回答方向”

五、最低采样纪律

1. 不编造：不知道的信号写”未知”，不推测补全
2. 不改写：原始措辞保留，不”优化”用户的表达
3. 标注来源：每一条信号必须能追溯到来源
4. 声明偏倚：不假装采样是客观完整的
5. 不在此阶段做判断：采样只管收，不管对不对

ewpage

COP.B02-P2 • 状态编码协议详细版

定位：COP 诊断管线第二协议——如何把采样帧映射到编码空间。本文件是 B02 五协议主文中 P2 的完整展开。作者：Yi Fu（付毅，ODDFounder）

一、P2 在诊断管线中的位置

输入采样(P1) → 状态编码(P2) → 诊断判定(P3) → 转介干预(P4) → 反馈校准(P5)

P2 是从”原始信号”到”可计算状态”的关键转换。不跳过 P2 直接做诊断——否则诊断就是基于表面标签的猜测。

二、编码流程：两步走

第一步：AST0 前置编码（必填，不可跳过）

在进入 COP 核心编码之前，必须先继承 AST0 的结构编码：

```

asto_precoding:
  state_type: 自在 | 共识 | 编码 | 物化 | 定向
  stage: 潜伏 | 显性 | 扩散 | 峰值 | 衰减 | 残留
  sequence_position: 识别 | 定位 | 分析 | 设计 | 执行 | 验证 | 迭代
  boundary_status: clear | blurred | contested
  exception_flag: false | true
  exception_detail: "" # exception_flag=true 时必填

```

ASTO 前置编码的快速判断法

五态判断： - “大家隐约觉得不对但说不清” → 自在 - “大家都知道有问题但没写成规则” → 共识 - “已经有明确的规则/流程/SOP” → 编码 - “规则已经被系统/工具强制执行” → 物化 - “系统能自己判断什么时候该改规则” → 定向

六阶判断： - “才刚开始出现苗头” → 潜伏 - “已经能稳定观察到” → 显性 - “开始影响其他部分” → 扩散 - “现在是最严重的时候” → 峰值 - “在好转/收敛” → 衰减 - “基本平息但留了尾巴” → 残留

边界判断： - “所有人都同意这个边界在哪” → clear - “有人说不清这个边界” → blurred - “有人在主动挑战这个边界” → contested

三、COP 核心编码

清晰度向量 (clarity_vector)

维度	含义	评分指南
signal_completeness (0-1)	信号够不够做判断	0.8+: 各通道信号齐全; 0.4-0.7: 有部分通道缺失; <0.4: 关键通道空白
signal_consistency (0-1)	不同来源的信号是否一致	0.8+: 多源交叉验证一致; 0.4-0.7: 部分矛盾; <0.4: 严重矛盾
category_concentration (0-1)	信号是否集中在单一类型	0.8+: 信号明确指向某一类型; 0.4-0.7: 混合信号; <0.4: 分散到无法归类

结构风险向量 (structural_risk_vector)

维度	含义	评分指南
responsibility_diffusion	责任扩散程度	high: 没人能说出谁在负责; medium: 有人负责但授权不清晰; low: 责任主体明确
power_asymmetry	权力不对称程度	high: 承压点与决策点完全分离; medium: 有一定不对称; low: 权责基本匹配
irreversibility_potential	不可逆潜力	high: 下一步动作可能有不可逆后果; medium: 可逆但有成本; low: 低风险动作

四、编码输出完整结构

```

encoding_output:
  encoding_id: ""

```

```

source_frame_id: ""

asto_precoding: { ... } # 见上

clarity_vector:
  signal_completeness: 0.0-1.0
  signal_consistency: 0.0-1.0
  category_concentration: 0.0-1.0

structural_risk_vector:
  responsibility_diffusion: low | medium | high
  power_asymmetry: low | medium | high
  irreversibility_potential: low | medium | high

encoding_confidence: 0.0-1.0
encoding_note: ""

# 低置信度时的必填项
low_confidence_reasons: [] # encoding_confidence < 0.6 时填写

```

五、编码不做什么

- 不跳过 ASTO 前置编码直接从表面标签起步
- 不把”这个人看起来很焦虑”编码为”火局”——需要走判局流程
- 不把分类集中度与结构风险混成一个分数——必须分开输出
- 不把编码当作解释——“这更接近 X”不等于”这是因为 Y”

ewpage

COP.B02-P3 • 诊断判定协议详细版

定位：COP 诊断管线第三协议——基于编码向量做分流判定。本文件是 B02 五协议主文中 P3 的完整展开。作者：Yi Fu（付毅，ODDFounder）

一、P3 在诊断管线中的位置

输入采样(P1) → 状态编码(P2) → 诊断判定(P3) → 转介干预(P4) → 反馈校准(P5)

P3 是诊断管线的判定核心。它回答：当前是否清楚、是否安全、是否该继续自动推进。

二、五种分流状态的完整定义

RESOLVE —— 可自动处理

含义：信号完整且一致，分类明确，结构风险低。系统可以继续自动推进。

准入条件（全部满足）： - signal_completeness > 0.7 - signal_consistency > 0.7 - category_concentration > 0.7 - responsibility_diffusion = low - irreversibility_potential = low

输出示例：

```
triage_status: RESOLVE
confidence: 0.85
primary_label: "个体技能缺口-可训练"
alternative_labels: ["中期职业倦怠-轻度"]
```

MIXED —— 混合信号

含义：信号存在矛盾或分类不集中。可以有限推进，但必须标注混合区域。

触发条件（满足任一）： - signal_consistency 在 0.4-0.7 之间 - category_concentration 在 0.4-0.7 之间 - 两个以上的 primary_label 候选得分接近

输出示例：

```
triage_status: MIXED
confidence: 0.55
primary_label: "组织结构模糊-水局倾向"
alternative_labels: ["个体过载-木局倾向", "两种模式混合"]
mixed_areas: ["用户描述同时包含'没人管'和'越做越多'两种信号"]
limited_action_allowed: true
```

FREEZE —— 冻结

含义：结构风险高或触及红线。暂停自动推进，升级人工。

触发条件（满足任一）： - irreversibility_potential = high - responsibility_diffusion = high AND power_asymmetry = high - 触及 ASTO 例外标记 - 信号包含以下关键词模式：法律风险、人身安全、自杀、暴力、严重违规

输出示例：

```
triage_status: FREEZE
confidence: 0.90
freeze_reason: "结构风险为高不可逆 + 权力严重不对称"
freeze_trigger: "用户描述中包含'被要求承担违法后果'的模式"
escalation_target: HUMAN_REVIEW
```

UNKNOWN —— 信息不足

含义：信号不够做判断。需要补充信息。

触发条件： - signal_completeness < 0.4

输出示例：

```
triage_status: UNKNOWN
confidence: 0.30
missing_info_hint: "缺少以下类型的信号：(1) 持续时间 (2) 第三方视角 (3) 具体实例"
suggested_supplement: "建议补充 2-3 个最近发生的事件描述"
```

REFER —— 转介

含义：问题超出 COP 可处理范围。路由到指定目标。

触发条件（满足任一）： - 问题类型属于工程执行 → 转 ODD - 问题类型属于责任门槛裁决 → 转 TAT - 问题类型属于治理边界争议 → 转 ECET - COP 自身置信度 < 0.3 且无法通过补充信息改善

三、判定矩阵（完整版）

complete	consist	concentr	resp_diff	power_asym	irrev	→ 分流
>0.7	>0.7	>0.7	low	low	low	RESOLVE
>0.5	>0.5	>0.5	low-med	low-med	low-med	MIXED
any	any	any	high	high	any	FREEZE
any	any	any	any	any	high	FREEZE
<0.4	any	any	any	any	any	UNKNOWN
any	any	any	any	any	out_of_scope	REFER

注：当多个条件同时触发（如 UNKNOWN + FREEZE），优先取安全等级更高的（FREEZE > REFER > UNKNOWN > MIXED > RESOLVE）。

四、判定纪律

1. 宁可多冻一次，不可错放一次。边缘情况默认 FREEZE。
2. 低置信度不得输出 RESOLVE。confidence < 0.6 时，最乐观输出为 MIXED。
3. FREEZE 必须写 freeze_reason。不能只标注 FREEZE 不给理由。
4. REFER 必须写 referral_target。不能只转介不说明转给谁。

ewpage

COP.B02-P4 • 转介与有限干预协议详细版

定位：COP 诊断管线第四协议——基于分流状态生成下一步动作与转介路由。本文件是 B02 五协议主文中 P4 的完整展开。作者：Yi Fu（付毅，ODDFounder）

一、P4 在诊断管线中的位置

输入采样 (P1) → 状态编码 (P2) → 诊断判定 (P3) → 转介干预 (P4) → 反馈校准 (P5)

P3 判定”什么状态”，P4 决定”接下来做什么”。

二、五种分流 → 动作映射

RESOLVE → 放行

```
action:
  type: PROCEED
  target: RT6
  payload:
    diagnostic_summary: ""      # 诊断摘要
    confidence: 0.85
    recommended_starting_step: 1 # RT6 六步中的建议起点
    attention_areas: []         # 需要注意但不足以阻止的区域
```

MIXED → 有限放行

```
action:
  type: PROCEED_WITH_CAUTION
  target: RT6 | LMM
  payload:
```

```

diagnostic_summary: ""
confidence: 0.55
mixed_areas: [] # 需要额外关注的混合信号区域
step_2_extra_attention: "" # RT6 步骤 2（扫描诊断）需额外关注的问题
review_required_after_step: 2 # 在哪一步之后必须回看

```

FREEZE → 冻结+升级

```

action:
  type: FREEZE_AND_ESCALATE
  target: HUMAN_REVIEW | TAT_REVIEW
  payload:
    freeze_reason: ""
    freeze_trigger_signals: [] # 触发冻结的具体信号
    diagnostic_chain: "" # 完整诊断链（采样→编码→判定）
    recommended_question_for_human: "" # 建议人类回答的核心问题
    minimum_wait: "" # 在人类回复前至少等待多久

```

UNKNOWN → 补充采样

```

action:
  type: SUPPLEMENT
  target: ORIGINAL_CHANNEL # 回到原采样通道
  payload:
    missing_info_hint: ""
    suggested_questions: [] # 建议追问的具体问题
    suggested_channels: [] # 建议从哪些通道补充
    max_supplement_rounds: 2 # 最多补充几轮

```

REFER → 转介

```

action:
  type: REFER
  target: TAT | ODD | ECET | HUMAN
  payload:
    referral_reason: ""
    diagnostic_summary: ""
    triage_chain: "" # 完整分流链
    urgency: routine | elevated | urgent

```

三、转介路由详细规则

转 TAT

触发条件（满足任一）： - 涉及责任门槛（谁承接高影响后果） - 用户明确申诉 - FREEZE 原因涉及权力不对称导致的不可逆
- 需要同意/补偿/熔断结构判断

转介包内容：诊断摘要 + 风险向量 + 触发信号 + COP 自身置信度

转 ODD

触发条件： - 问题涉及契约/验证/门禁/证据/封存 - 产出物质量争议 - 审计追溯需求

转介包内容：涉及的 artifact_id + 契约链 + 门禁日志

转 ECET

触发条件： - 治理边界争议（多头管理、影子主权） - 元治理条件问题（“谁来治理治理者”） - 跨文明比较需求

转人工

触发条件： - COP 自身判定置信度 < 0.3 - 三个转介目标都不适用 - 人类明确要求人工接手

四、COP 可以做与不可以做

COP 可以做

- 基于规则的建议（不涉及高风险强制动作）
- 转介路由选择
- 信息补充请求
- 有限动作建议（标注置信度和边界）

COP 不可以做

- 强制性高风险动作授权
- 行为锁定的最终裁决
- 同意、申诉、补偿结构定义 → 回引 TAT
- 绕过 ODD 直接修改契约/门禁/封存
- 在 FREEZE 状态下仍输出 PROCEED

ewpage

COP.B02-P5 • 反馈学习校准协议详细版

定位：COP 诊断管线第五协议——收集诊断后的真实结果，对比诊断结论与实际情况，校准协议参数。本文件是 B02 五协议主文中 P5 的完整展开。 作者：Yi Fu（付毅，ODDFounder）

一、P5 在诊断管线中的位置

输入采样 (P1) → 状态编码 (P2) → 诊断判定 (P3) → 转介干预 (P4) → 反馈校准 (P5)

↑

|

校准结果回流到采样策略

P5 是诊断管线的闭合环。没有 P5，COP 就是一个只会做判断但从学习判断是否准确的系统。

二、校准数据采集时点

每次诊断后，在以下时点自动或手动采集校准信号：

时点	采集方式	采集内容
即时 RT6 步骤 2	自动 对比	用户对诊断摘要的反应（接受/质疑/补充） COP 诊断标签 vs RT6 扫描诊断结论是否一致

时点	采集方式	采集内容
RT6 步骤 6	对比	复诊发现的问题是否与初始 COP 诊断匹配
事故回溯	手动	FREEZE/REFER 是否正确—— 是阻止了一次损害还是阻碍了本可进行的处理

三、三类误判的详细校准

3.1 False Safe（假安全）

定义：COP 判定为 RESOLVE/MIXED 允许推进，但实际不应推进。

发现方式：- RT6 步骤 6 复盘发现：初始诊断低估了风险 - 用户报告：COP 说”可处理”但实际发生了损害
- 事故回溯：本该 FREEZE 的被放行了

校准动作：

```
misclass: false_safe
calibration:
  - parameter: category_concentration_threshold_for_RESOLVE
    direction: up # 提高阈值
    magnitude: moderate
    reason: "上一批次中同类信号模式出现 2 例假安全"
  - parameter: irreversibility_sensitivity_for_this_pattern
    direction: up
    magnitude: minor
```

3.2 Over Freeze（过度冻结）

定义：COP 判定为 FREEZE，但人工复审后发现本可推进。

发现方式：- 人工复审结论：“此例本可 MIXED 有限推进” - FREEZE 被 override 且 override 理由被后续证实为正确

校准动作：

```
misclass: over_freeze
calibration:
  - parameter: freeze_trigger_threshold_for_this_pattern
    direction: down
    magnitude: moderate
    reason: "同类模式近 5 例中 3 例被人工判定为过度冻结"
```

3.3 Wrong Type（类型错误）

定义：分流方向正确但分类标签错误（如本该是 MIXED 但被标为 RESOLVE，或主标签标错）。

发现方式：- RT6 步骤 2 的扫描诊断给出了不同的结论 - 编码置信度与最终结果不匹配

校准动作：

```
misclass: wrong_type
calibration:
  - parameter: encoding_mapping_for_this_signal_pattern
    direction: remap
    magnitude: major
    reason: "该信号组合在最近 10 例中 6 例被修正为不同的主标签"
```

四、误判样本库规范

每次误判必须入库。样本库最小字段：

```
misclass_entry:
  entry_id: ""
  source_diagnosis_id: ""
  misclass_type: false_safe | over_freeze | wrong_type
  actual_outcome: ""          # 真实结果

# 严重度
severity:
  false_safe: low | medium | high | critical
  over_freeze: low | medium | high
  wrong_type: low | medium | high

# 根因分析
root_cause_guess: ""          # 初步判断的根因
# 示例: "信号中缺少持续时间信息导致高估了信号完整性"

# 校准结果
parameter_adjustments: []
review_status: pending | reviewed | applied
```

五、校准版本管理

每次协议参数调整产生一个新版本：

```
protocol_version:
  version_id: "v1.2.0"
  previous_version: "v1.1.0"
  changes:
    - parameter: ""
      old_value: ""
      new_value: ""
      reason: ""
      based_on_misclass_entries: [] # 基于哪些误判样本
  release_date: ""
  rollback_conditions: ""          # 什么情况下回滚这个版本
```

六、校准纪律

1. 不得基于单例调整阈值。至少 3 例同类型误判才能触发参数调整。
2. 严重度 critical 的 false_safe 单例即可触发紧急复审。不等待积累到 3 例。
3. 每次校准必须记录基于哪些样本。可追溯、可回滚。
4. 校准不得在事故当天的情绪驱动下进行。至少间隔 48 小时冷静期。
5. 校准版本保留回滚条件。如果新阈值导致反向误判增多，应回滚。

ewpage

COP.B03 • 诊断走通示例：从采样到校准

定位：COP 五协议诊断管线的端到端走通示例——展示一例完整信号如何经过 P1→P2→P3→P4→P5。

作者：Yi Fu（付毅，ODDFounder）

场景

一位 42 岁的中层管理者，在一家 500 人制造企业负责生产调度。最近一年持续感到“被夹在中间”——上级给目标但不给资源，下级抱怨目标不合理但自己无法向上推动。他在考虑辞职，但不确定是“真的该走”还是“只是太累”。

P1：输入采样

采样帧

sampling_frame:

```
frame_id: "MFG-MGR-2026-001"
collected_at: "2026-04-28"
source_type: questionnaire
source_desc: "用户主动填写的9题诊断问卷+2道开放题"
```

channels:

- channel_id: "self-report"
modality: text
signals:
 - signal_id: "S1"
raw_content: "我被要求对生产线的效率负责，但我没有权限调整排班表"
reliability: high
reliability_basis: "这是对具体工作流程的描述，不是感受表达"
 - signal_id: "S2"
raw_content: "每次我跟上面要人，就说让我自己想办法"
reliability: medium
reliability_basis: "这包含感受成分——'每次'可能不是字面的每次"
 - signal_id: "S3"
raw_content: "连续三个季度考核拿了C——不是因为效率没达标，是因为上级说我执行力度不够"
reliability: high
reliability_basis: "可核实的事实陈述"

sampling_bias_note: "本帧信号全部来自用户单方面描述。最可能漏掉：上级视角的叙述、同级别管理者的比较信息、组织

completeness_self_assessment:

```
signal_density: moderate
blind_spots: ["上级视角缺失", "组织架构正式文件未获取"]
recommended_supplement: "建议补充一份组织架构图和该岗位的正式职责说明"
```

P2: 状态编码

ASTO 前置编码

```
asto_precoding:
  state_type: 共识      # 大家都知道有问题，但没写成规则
  stage: 显性          # 问题已稳定可观察（连续三个C考核）
  sequence_position: 定位 # 在"这是什么问题"的阶段
  boundary_status: blurred # 职责边界模糊——"对效率负责但不控制排班"
  exception_flag: false
```

COP 核心编码

```
encoding_output:
  encoding_id: "MFG-MGR-2026-001-ENC"
  source_frame_id: "MFG-MGR-2026-001"

  clarity_vector:
    signal_completeness: 0.6 # 三个信号明确但全部来自单方面
    signal_consistency: 0.9 # 三个信号高度一致
    category_concentration: 0.8 # 明确指向"权责割裂"类型

  structural_risk_vector:
    responsibility_diffusion: medium # 有明确责任人但授权不清晰
    power_asymmetry: high # 有责无权
    irreversibility_potential: medium # 辞职是重要的但非不可逆

  encoding_confidence: 0.75
  encoding_note: "信号高度一致指向金局（权责割裂）。置信度中等偏离因为缺少第三方视角验证。"
```

P3: 诊断判定

判定矩阵计算

维度	值	阈值	结果
completeness	0.6	>0.5	通过
consistency	0.9	>0.5	通过
concentration	0.8	>0.5	通过
power_asymmetry	high	—	关注
irreversibility	medium	—	关注

→ 判定: MIXED（信号虽一致但来自单方面，权力不对称高）

```
diagnostic_output:
  triage_status: MIXED
  confidence: 0.75
  primary_label: "金局-权责割裂（中层管理者典型模式）"
  alternative_labels: ["火局-责任引爆（三个C考核=起火信号）"]

  mixed_areas: ["信号全部来自用户方，缺少上级视角和组织结构文件"]
  limited_action_allowed: true
```

P4: 转介与有限干预

```
action:
  type: PROCEED_WITH_CAUTION
  target: RT6

payload:
  diagnostic_summary: "用户处于典型中层管理者权责割裂状态（金局），伴有火局早期信号（连续考核C=正在被点燃）。"
  confidence: 0.75
  mixed_areas: ["缺少第三方视角验证"]
  step_2_extra_attention: "RT6步骤2建议加入'组织架构对比'——用户的正式职责 vs 实际被问责的内容"
  review_required_after_step: 2

# RT6不启动的备选路径
suggested_escalation_if_declines: "如果用户在步骤3说不进入处理——暂时不启动RT6，但建议一个月后复诊"
```

P5: 反馈校准（假设的后续）

假定结果：四个月后

用户在 RT6 的支持下做了一个决定：不是辞职，而是向上级做了一次”责任边界对话”——要求明确”我对什么有决定权”

校准评估

```
calibration:
  diagnosis_accurate: true
  actual_outcome: "用户通过结构对话缓解了权责割裂——验证了COP的诊断方向正确"

# COP没有误判——金局+火局早期信号是正确的
# 但有一个可以改进的地方:
improvement_note: "在P2编码中，signal_completeness被标记为0.6因为只有用户视角。但本案例中用户视角最终被证明是"
```

如果诊断错了

替代场景：如果用户实际上不是权责割裂，而是”自己的执行确实有问题，上级的批评是合理的”（即诊断过度归因于结构问题）——这就是一个 False Safe 误判（COP让用户相信是结构问题，但实际上用户自己的执行也有责任）。

在这种情况下，校准动作是：提高 power_asymmetry 从 medium 到 high 的判断标准——需要至少两个独立来源确认”有责无责” high。

走通总结

这个示例展示了 COP 诊断管线最重要的纪律：

1. P1 采样时不编造——缺少上级视角就标注”缺少”，不推测上级在想什么
2. P2 编码前先做 ASTO 前置——五态/六阶/边界一个不跳
3. P3 判定时宁可保守——即使信号一致，因为来源单一，只给 MIXED 不给 RESOLVE
4. P4 转介时标注混合区域——让 RT6 在步骤 2 做额外验证
5. P5 校准时区分”诊断对了”和”可以微调阈值”——不因为一次正确就过度自信

ewpage

COP 诊断协议 v1

定位：本文是 **COP Phase 1** 的正式协议主文。它不替代 README.md，而负责把“输入什么、输出什么、何时冻结、何时”

状态：v1 协议化草案 / 可公开征评

一、继承合同

COP 诊断协议默认继承：

1. **DM** 提供最小结构承诺与研究纲领边界。
2. **ASTO** 提供桥接语法与状态语言。
3. **ECET** 提供治理边界与文明尺度约束。
4. **TAT** 提供责任门槛、冻结、申诉、补偿与责任承接规则。
5. **ODD** 提供工程审计、版本留痕、回滚与执行验证接口。
6. **LMM** 提供前线显影、接引与现实承接路径。

本文只新增：

- 输入采样规则
- 诊断输出合同
- 冻结与转介规则
- 诊断级数据边界

本文不处理：

- 治理合法性的终局裁决
- 高风险动作的独立授权
- 工程系统的完整实现细节
- 市场前台的话术与产品包装

1.1 配套工具文

本文默认与以下三份工具文并置使用：

- COP. 校准样本协议. v1. md
- COP. 误判样本库规范. v1. md
- COP. 版本发布纪律. v1. md

分工如下：

1. 本文定义主协议
 2. **校准样本协议** 定义什么样的样本有资格进入学习与校准
 3. **误判样本库规范** 定义误判如何记录、复盘与归因
 4. **版本发布纪律** 定义协议、权重、阈值如何发布、冻结与回滚
-

二、输入合同

2.0 当前场景包声明

v1 当前使用：

- **A-D** 四类型
- **Q1-Q9** 九题问卷

作为首个场景化输入包。

这不等于 COP 的全部分类学，

而只是当前 Phase 1 的最小可部署采样协议。

2.1 当前 v1 的最小输入

Phase 1 当前要求：

1. Q1-Q9 九题答案全部有效
2. 每题只能取 A / B / C / D 之一
3. 任何缺失、重复、多选、无效值都不得被静默吞掉

2.2 缺失值规则

v1 的最低纪律是：

1. 缺题 若任一题缺失，默认不进入正式分类，直接返回 `trriage_status = UNKNOWN`。
2. 无效值 若出现非 A-D 值，直接返回 `trriage_status = UNKNOWN`，并要求重新提交。
3. 低信号输入 即使九题齐全，若总激活信号过低，也允许返回 `UNKNOWN`，而不是强行给出主类型。

最短判断：

1. 缺失值不是噪声，而是协议状态。
2. v1 宁可说不知道，也不伪装成知道。
3. A-D + 9题 是当前包，不是整层本体。

2.3 场景包扩展字段

若当前场景包处理的是老板线、组织链路或高影响责任分流，

建议在最小输入之外追加：

- `case_scope`
- `decision_owner`
- `execution_owner`
- `consequence_owner`
- `loss_signal`
- `evidence_index`

这些字段不替代分类输入，

而是为后续的 `primary_break_order` / `trriage_status` / `refer_to` 提供可复核背景。

三、输出合同

每次正式输出至少包含以下字段：

字段	含义	最低纪律
<code>primary_type</code>	当前最强候选类型	FREEZE / UNKNOWN 时允许为空
<code>secondary_type</code>	次强候选类型	只有存在实质竞争时才保留
<code>primary_break_order</code>	当前主断裂行动位点	场景包使用七序时建议必填
<code>secondary_break_order</code>	当前次断裂行动位点	存在竞争断裂时保留
<code>score</code>	四类原始得分	不能被前台隐藏到完全不可复核
<code>evidence_gap</code>	仍缺的关键证据	高风险时不得省略
<code>classification_confidence</code>	分类集中度	不得冒充真实风险

字段	含义	最低纪律
<code>structural_risk</code>	结构风险等级	不得由分类集中度直接顶替
<code>triage_status</code>	当前分流状态	必须使用固定枚举
<code>tags</code>	命中的关键结构标签	至少保留高风险标签
<code>refer_to</code>	当前应转介到哪里	高风险时不得缺失

3.1 固定分流状态

`triage_status` 只允许取：

- RESOLVE
- MIXED
- FREEZE
- UNKNOWN
- REFER

3.2 行动断裂字段纪律

若场景包启用了 `ASTO` 的行动切片，

则：

1. `primary_break_order` 用于说明当前主断裂更可能落在哪一序
2. `secondary_break_order` 用于说明当前次断裂或竞争断裂
3. 这两个字段只承担解释与分流功能，不构成责任终裁
4. 若证据不足以支持该判断，应优先填充 `evidence_gap`，并允许进入 `UNKNOWN` / `FREEZE`

3.3 状态优先级

`UNKNOWN` > `FREEZE` > `REFER` > `MIXED` > `RESOLVE`

原因很简单：

1. 先判断能不能回答
2. 再判断要不要暂停自动化
3. 再判断要不要升级到更高位接口
4. 最后才讨论是不是混合态和常规解答

四、冻结与转介规则

4.1 何时必须 FREEZE

满足任一条即冻结：

1. 多条硬规则同时触发
2. `classification_confidence` 很低而 `structural_risk` = `HIGH`
3. 输出结果将直接碰到高风险动作、责任门槛或工程阈值，但当前证据不足

4.2 何时必须 REFER

满足任一条即默认转介：

1. `structural_risk` = `HIGH`
2. 命中高强度阈值缺失、外部约束过强、结构循环依赖等标签
3. 当前动作建议一旦执行，可能造成高不可逆损害、责任外溢或大规模组织后果

4.3 refer_to 的最低映射

条件	默认转介
任意 FREEZE / UNKNOWN / REFER	HUMAN_REVIEW
structural_risk = HIGH	TAT_REVIEW
命中阈值、外部约束、工程留痕问题	ODD_AUDIT
涉及文明、元治理或跨制度裁决	ECET_ESCALATION

五、复核与争议处理

5.1 复核纪律

1. 不得只给结论不给原始分布 复核至少要能看到四类得分与命中标签。
2. 不得让前台包装覆盖协议状态 前台页面再漂亮，也不能把 FREEZE / UNKNOWN / REFER 隐掉。
3. 不得把诊断结果直接当成执行许可 高风险输出必须经过人工或更高位接口。

5.2 最小争议处理

若用户或复核者对输出有异议，至少允许：

1. 重新提交原始答案
2. 查看命中标签与分流状态
3. 升级到人工复核
4. 在高风险情形下升级到 TAT / ODD

六、误判成本

COP 当前至少区分四类误判成本：

6.1 False Safe

本应 FREEZE / REFER 的高风险状态，被误写成 RESOLVE / MIXED。

这是最高优先级错误，因为它会把真正该升级的情形放行。

6.2 False Unknown

明明已有足够信号，却被长期压成 UNKNOWN。

它会降低协议可用性，但通常仍低于 False Safe 的危险度。

6.3 Wrong Type

高风险标签与分流方向大体正确，但主类型或副类型误判。

它会影响动作建议与后续学习质量。

6.4 Over Refer / Over Freeze

本可低风险处理的情形被过度升级。

它会抬高人工复核成本、延迟与摩擦，但通常仍优先于把高风险放漏。

6.5 当前成本优先级

False Safe > Wrong Type > Over Refer / Over Freeze > False Unknown

最短判断:

1. COP 最怕的不是多转几次人工, 而是把该转的高风险状态当成低风险放过去。
 2. 学习规则必须先服从误判成本排序, 不能只看“总体准确率”。
-

七、复核一致性

7.1 哪些结果必须进人工复核样本

以下输出默认进入复核样本池:

- 所有 FREEZE
- 所有 REFER
- 所有 structural_risk = HIGH
- 所有用户提出争议的结果

这些样本进入复核与学习流程时, 应进一步遵守 COP. 校准样本协议. v1. md。

7.2 最低双人复核规则

对于高风险复核样本, 至少记录三项一致性:

1. triage_status agreement 两名复核者对分流状态是否一致。
2. primary_type agreement 在非 UNKNOWN 条件下, 两名复核者对主类型是否一致。
3. critical_tag overlap 关键高风险标签是否有足够重叠。

7.3 当前最低门槛

在滚动复核窗口内, 若出现以下任一情况, 应暂停自动学习并进入规则审查:

- triage_status agreement < 0.85
- primary_type agreement < 0.70
- 高风险标签重叠长期低于一半

7.4 分歧处理

若两名复核者结论冲突:

1. 先保留更保守的分流状态
2. 进入第三方复核
3. 记录冲突来源, 用于后续题目或权重修订

误判、冲突与根因标签的登记格式, 统一回引 COP. 误判样本库规范. v1. md。

八、版本回滚门槛

8.1 必须冻结新版本的情形

满足任一条即停止继续放量:

1. 在滚动 50 个已复核高风险样本中, False Safe >= 2
2. 任一版本造成重大不可逆损害且协议明显低估风险
3. triage_status agreement 连续两个窗口低于 0.85

8.2 必须回滚到前一稳定版本的情形

满足任一条即建议回滚：

1. 新版本相对稳定版本明显放大 **False Safe**
2. **REFER + FREEZE** 比例显著上升，但没有带来高风险漏判下降
3. 新版本导致人工复核负担急升且无清晰收益

8.3 回滚后的最低动作

1. 恢复上一稳定权重
2. 冻结自动学习
3. 抽查争议样本
4. 重新评估题目、阈值与标签定义

具体的版本编号、发布门禁与冻结流程，统一回引 COP. 版本发布纪律.v1.md。

九、数据边界

9.1 最小化

v1 默认只收：

- 题目答案
- 必要的会话标识
- 协议输出

在未单独声明前，不额外收集与诊断无关的身份资料。

9.2 去标识化

用于后续学习的数据，默认应与直接身份标识分离。

9.3 保留周期

v1 若未发布单独数据政策，默认不得无限期保留原始反馈。

9.4 删除与退出

若前台产品允许账户或会话级追踪，就必须允许删除和退出，不得只有采集入口没有撤回出口。

9.5 训练边界

Phase 1 当前只允许记录，不允许在未补齐标签质量控制、版本回滚和误判成本规则前自动改写权重。

十、版本与回滚

1. **权重矩阵必须外置** 不能把权重写死在不可追溯代码里。
2. **版本号必须可追踪** 每次协议、权重、分流规则变化都应能定位。
3. **高风险变更必须可回滚** 若某次更新显著放大误判、压低复核一致性或造成过度转介，应可回退到前一稳定版本。
4. **校准、误判、发布三件事必须分文管理** 不得把“样本进入条件”“误判归因记录”“版本发布纪律”混成一张模糊施工

十一、最短使用顺序

若只想快速把 COP Phase 1 用起来，建议顺序：

1. README.md
2. COP.P1-V2-加权判定引擎.md
3. COP.P1-V3-工程化接口设计.md
4. 本文

若只想问一句“它什么时候不该继续自动回答”，直接先看本文第四节。

十二、最短压缩句

1. COP 诊断协议 v1 的核心，不是更敢判，而是更敢停。
2. 只要进入高风险现实，协议必须把“冻结、转介、复核”写得和“分类、动作”一样清楚。