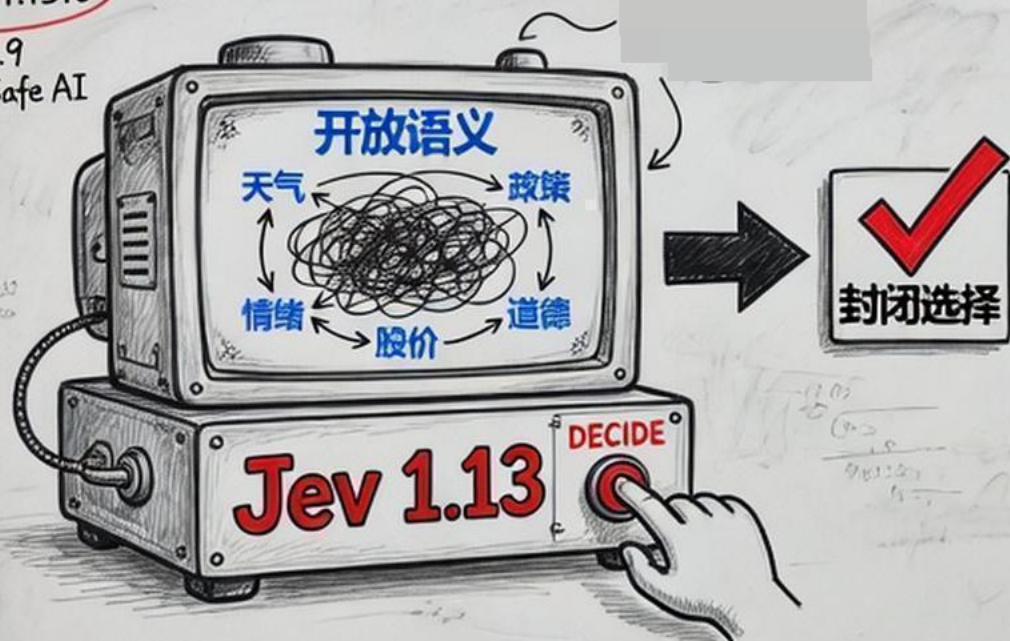


Jev深度研究报告

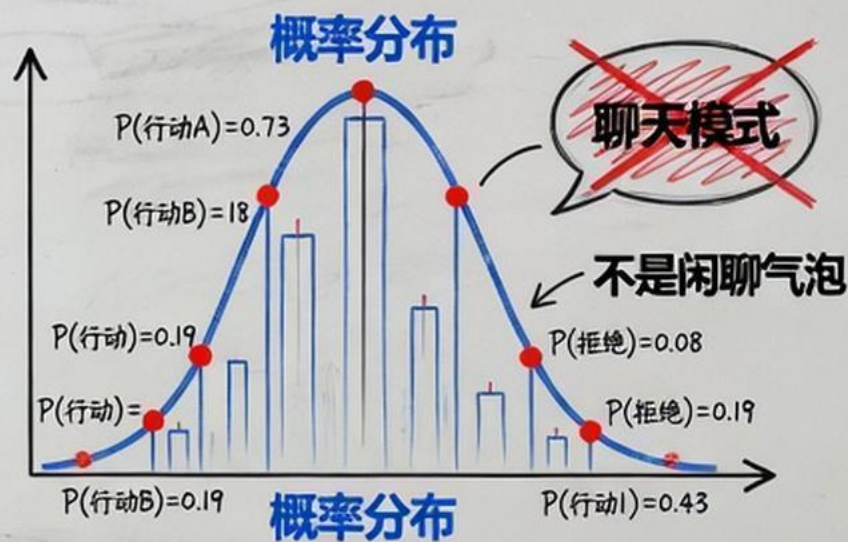
System One模型 • RLCD • 机器原生决策智能

jev-1.13.0

2026.9
TypeSafe AI



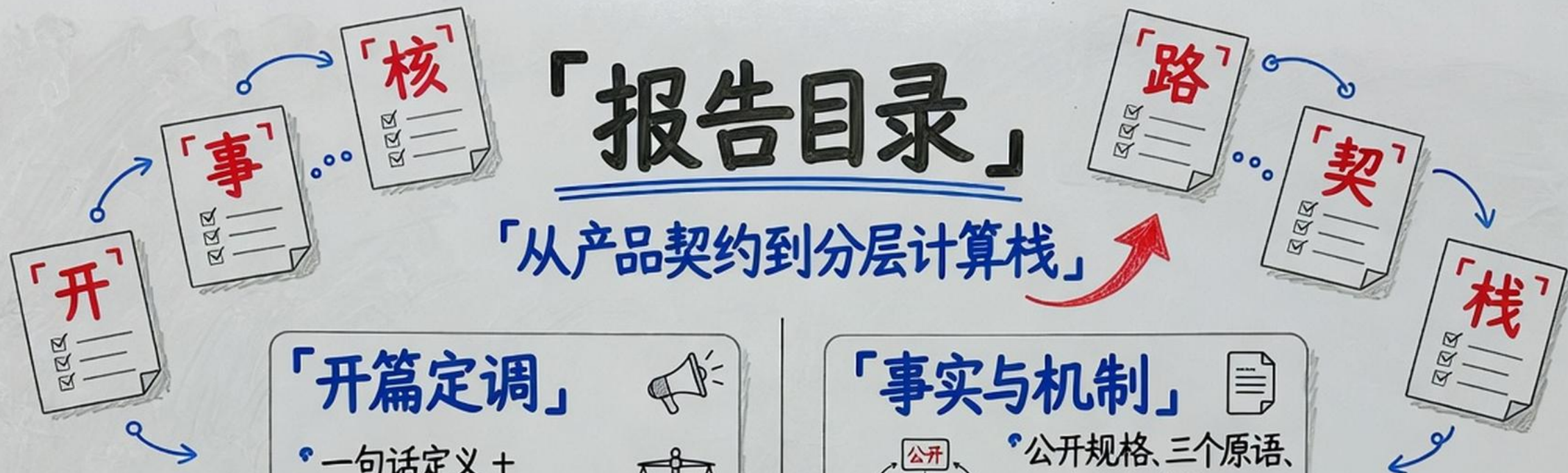
- 研究对象: TypeSafe AI 2026年9月发布的 Jev 1.13
- 当前公开版本: jev-1.13.0
- 承认接口革命 拒绝把未公布的训练细节写成科学革命



- 研究方法: **官方文档 + 评测核验**
- 技术逆向推断 + 系统架构比较
- 应用与研究路线设计

「报告目录」

「从产品契约到分层计算栈」



「开篇定调」

- 一句话定义 + 六项判断与证据强度



证据强度

「事实与机制」



- 公开规格、三个原语、可确认机制与未公开内容



「核验与边界」

- 官方倍率、零幻觉收缩、九类失败模式与结构不变量



九类失败模式



「路线与概念」



- 竞争比较、四层计算栈、以及五个团队



首创概念

@ 清新研究团队简介

- 领导学术研究团队近30人。指导大数据、AI、人形机器人等多个产业团队。
- 团队坚持：整体主义、实证主义、社会建构、进步主义。

沈阳：清华大学新闻学院/人工智能学院双聘教授、博导



六大研究方向：

视频号：@清新研究；公众号：@清新研究



1. AI大模型理论与哲学



2. AI文艺



3. AI应用



4. 新媒体与网络舆论

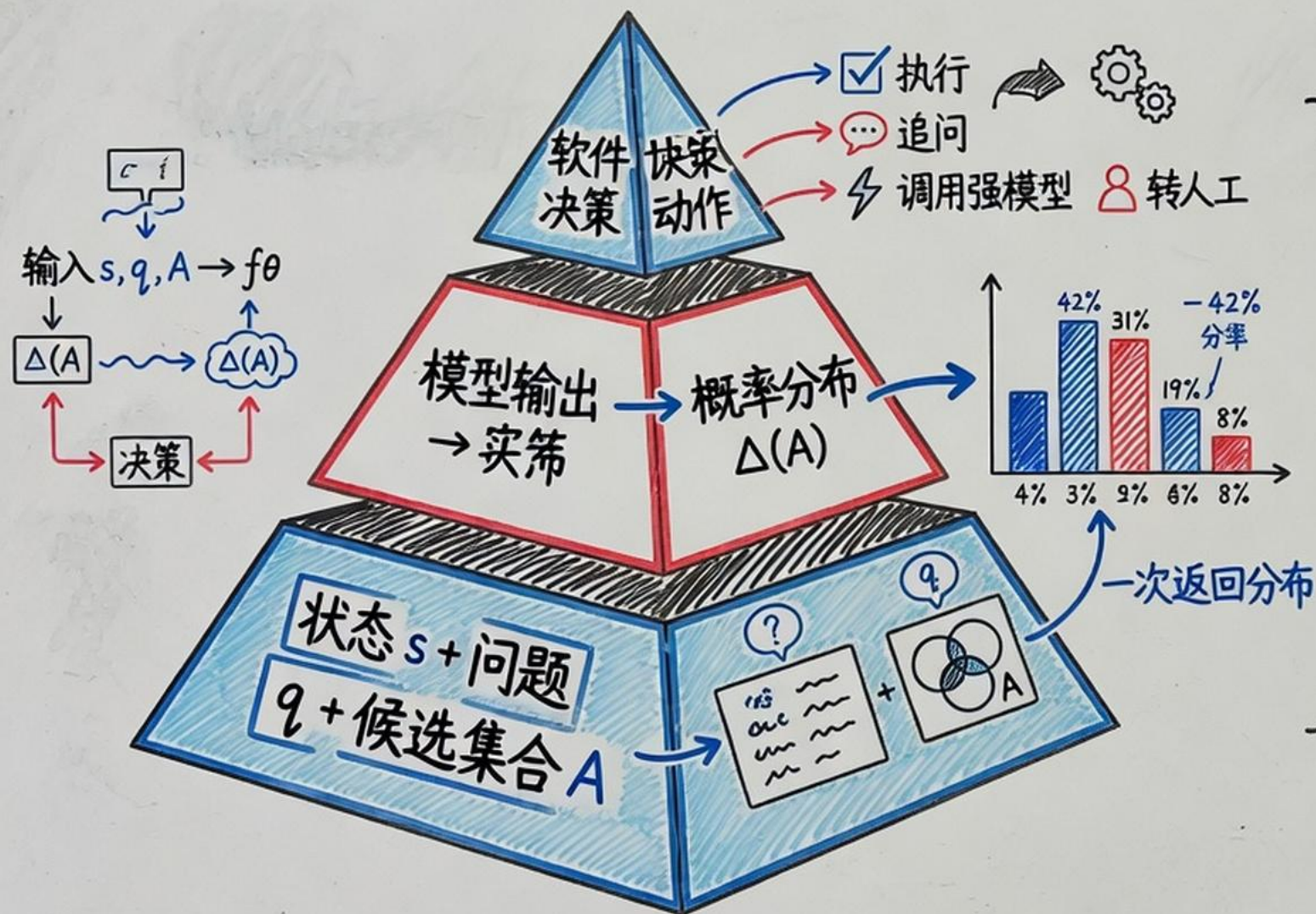


5. 大数据



6. XR应用

「一句话定义：条件决策算子」 $\rightarrow f_{\theta}(s, q, A) \rightarrow \Delta(A)$



- Choice: 取最大概率选项
- Score: 概率加权期望排序
- Noul: 命题为真的概率

软件依据



概率 + 风险阈值 + 业务规则

决定下一步



@ 清新研究团队 | 2026年9月

数据来源: <https://docs.typesafe.ai/primitives>

「研究方法与证据纪律」

先打开原页,再写结论 → 

从搜索  → 核验  → 收缩表述 

2026.9.18

研究启动

Jev早期访问开启



2027

核验阶段

对照TypeSafe模型页
+ 博客 + 原语页 + 评测站原文



2023

政策引用

仅引用中国政府网
与中央网信办框架



2024

收缩表述

未找到权威来源的
规模、份额、估值一律删除



2025

结论撰写

先打开原页
再写结论



研究日期: 2026年9月18日至19日
Ulev处于早期访问, 样本和限额能快速变化

System One: 面向软件的快决策 ⚡

理解自然语言, 但不生成文本

- 官方文档写明: System One模型用于做出软件可以直接使用的快速、结构化决策。
- 它像大语言模型一样理解自然语言输入, 但返回类型化决策和概率, 而不是生成文本。
- Jev是TypeSafe的旗舰模型, 也是该公司公开的第一个System One模型。名称借用卡尼曼的快思考。



输出: 结构化决策 + 概率

带伞决策



带伞?	概率
✓ 是	87%
否	13%

87% 13%

直接给软件用

六项判断与证据强度

能确认的确认，不能确认的标成不能确认



证据强度: 高

Jev不同于聊天大语言模型

它不生成自由文本，直接输出封闭类型和概率。证据强度高。



封闭类型
+
概率

证据强度: 高



输出方式差异

不生成自由文本，直接输出封闭类型和概率。证据强度高。



证据强度: 中

~~完全不是语言模型?~~

它是否完全不是语言模型

不能确认。更准确的表述是非自回归生成的语言决策模型。证据强度中。



✓ 语法
✓ 类型
✗ 193.6倍
✗ 444.6倍

证据强度: 中

RLCD
不可复现

所谓零幻觉

只在语法和类型层面可成立。193.6倍与444.6倍不普适。RLCD尚非可复现算法。



高中中
高祥中中低

判断卡片示例
(右上角红蓝标
记证据强度)

六张判断卡片

证据强度



核心结论：改的是产品契约

不是发明了更会聊天的通用模型



改写



Jev最重要的创新：

- 把模型的产品契约从生成字符串改造成对开放语义进行封闭决策
- 用自然语言理解获得零样本迁移，用预定义的Choice、Score和Noul换取类型安全、并行计算、低延迟和可门控的不确定性

自然语言理解

零样本迁移

零样本迁移

类型安全 ✓
并行计算 ✓
低延迟 ✓
可门控的不确定性 ✓

从公开证据看，它已构成有意义的新产品路线，
但尚不足以证明一种完整的新基础模型范式。



第一章

公开事实 与产品边界

- 先把Jev当成可核对的托管接口来读，
而不是当成已经公开的新神经网络家族。
- 边界写清楚，后面的倍率、幻觉和架构
推断才不会互相打架。

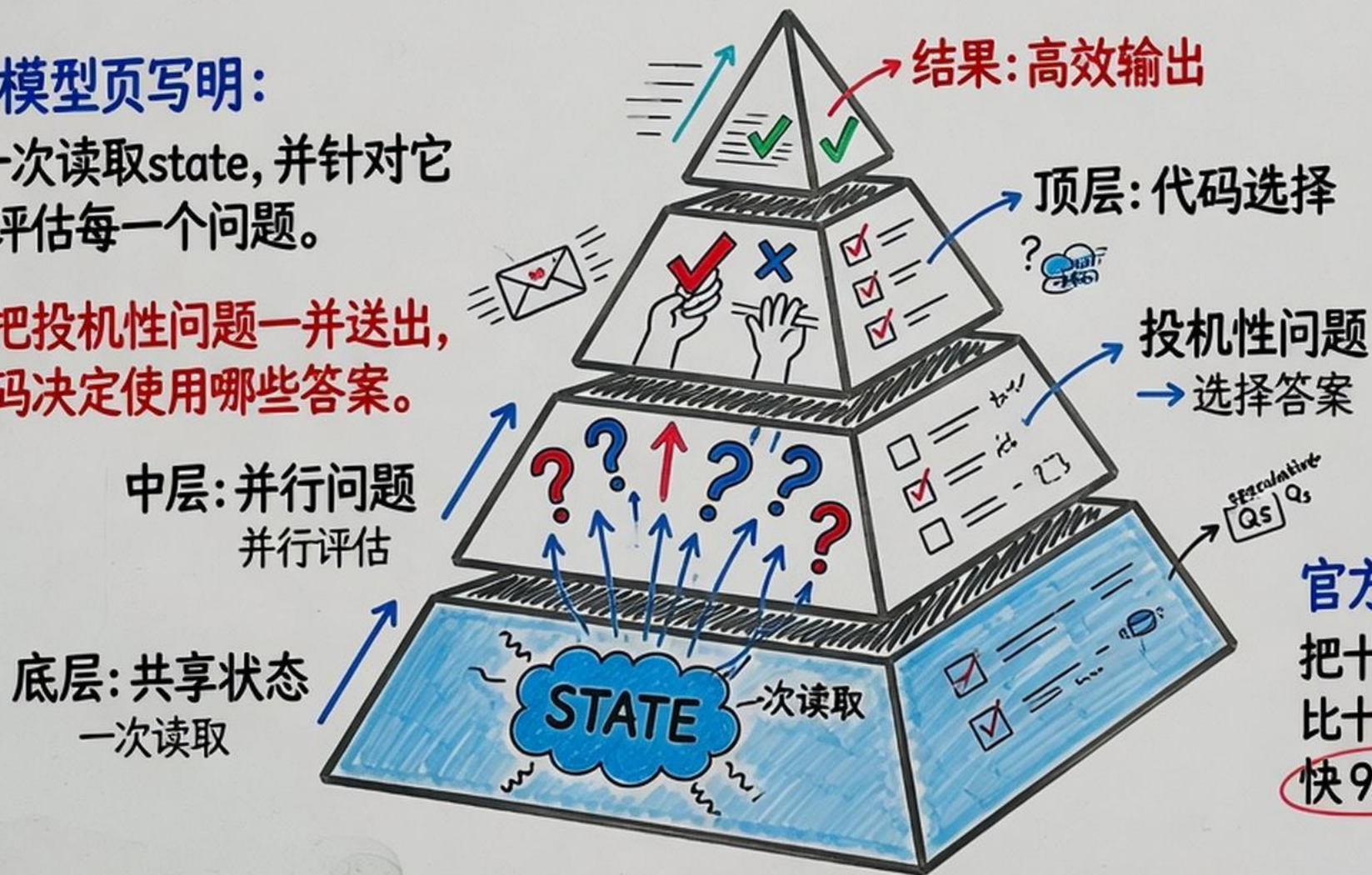


「一次读状态, 并行问多个问题」💡

官方模型页写明:

Jev一次读取state, 并针对它并行评估每一个问题。

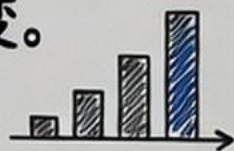
→ 可以把投机性问题一并送出, 由代码决定使用哪些答案。



新增问题主要增加 token, 不显著增加等待

官方原语页引用cookbook:

把十三个问题打成一次调用, 比十三次分开调用便宜11.5倍、快9.6倍, 答案不变。

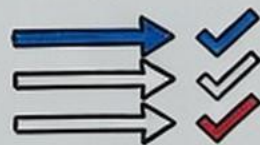


第二章 为什么需要 封闭决策



驱动来自接口错配、校准需求、并行求值和高频判断

并行求值

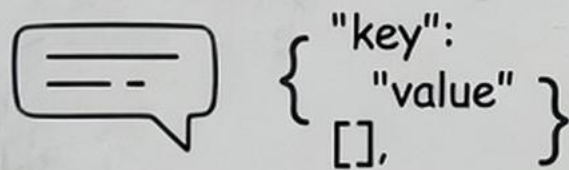


高频判断



Jev 1.13的产品边界

文本输入、共享权重、英语最佳



输入仅为文本

仅支持字符串、JSON对象或文本数组
不支持图像、音频和视频



EN

中日韩

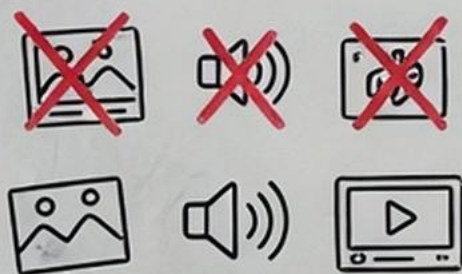
英语是主要训练语言

中日韩可用但官方提示性能并不等同
必须在自有内容上测试

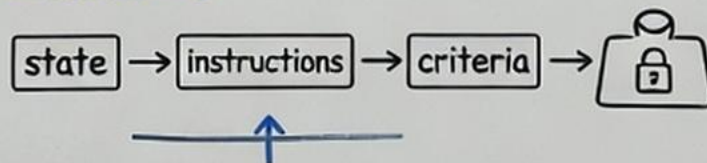


共享权重

不以客户数据微调或LoRA
同一组权重服务所有账户



适配方式



适配只能通过state、instructions和criteria完成
无法使用客户数据微调



「校准让不确定性可被软件使用」

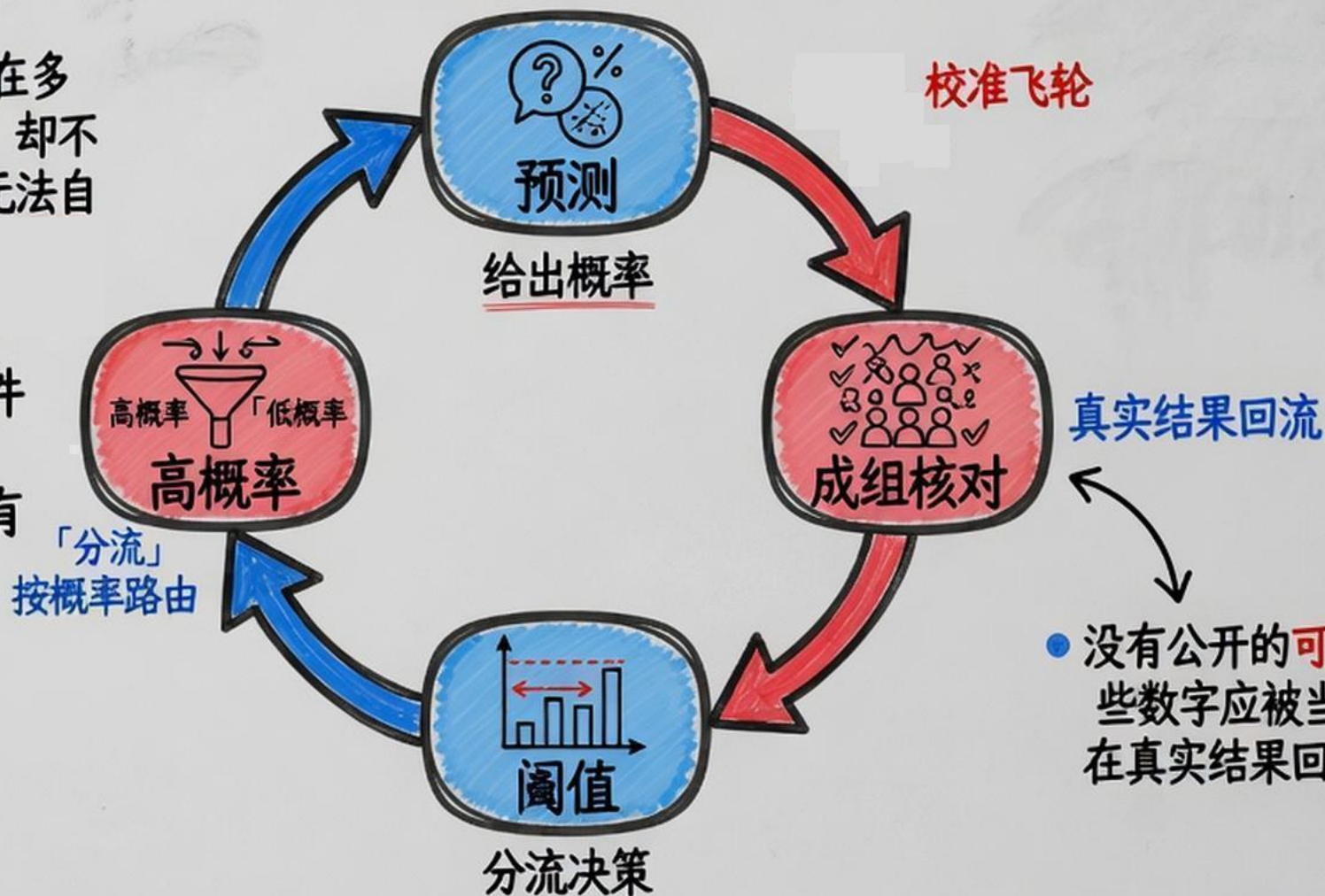
成组预测，不是单条保证

- 官方写明：如果模型在多数情况下能完成任务，却不说明何时失败，它就无法自动化该任务。

- 被赋予**0.2概率**的事件应大约有两成成立，**被赋予0.8**的应大约有八成成立。

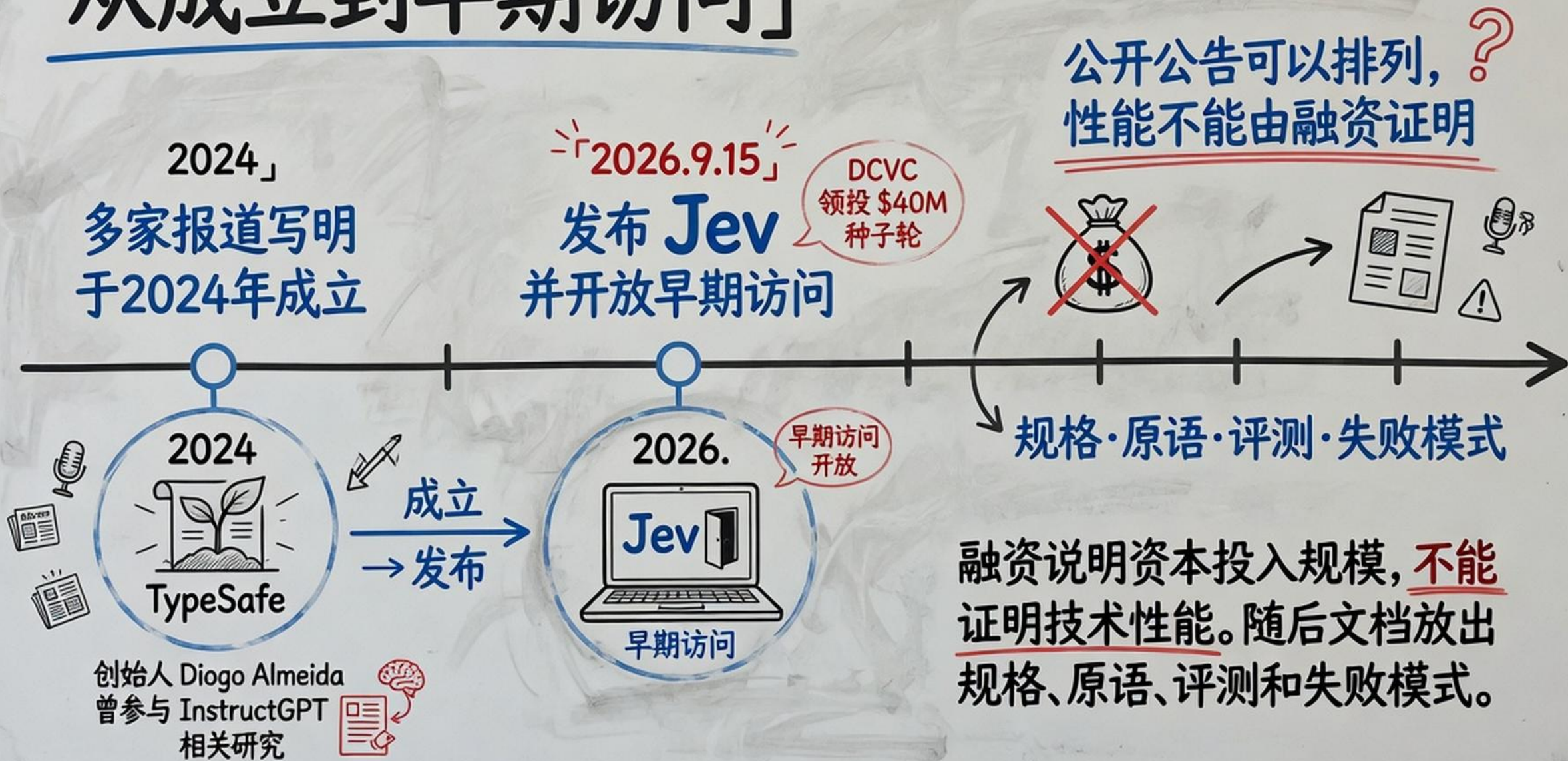


- 这是群体统计性质。



- 没有公开的**可靠性图**之前，这些数字应被当成分流信号，并在真实结果回流后重估。

「从成立到早期访问」



「生成式接口与软件执行的错配」

为一次选择支付了生成成本

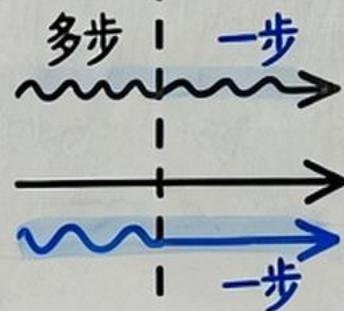
「传统大语言模型」↓

- 先把判断编码成token序列
- 再由解析器恢复成枚举、布尔值或数值
- 这个过程为本来只需一次选择的任务支付自回归解码、格式约束、重试和校验成本

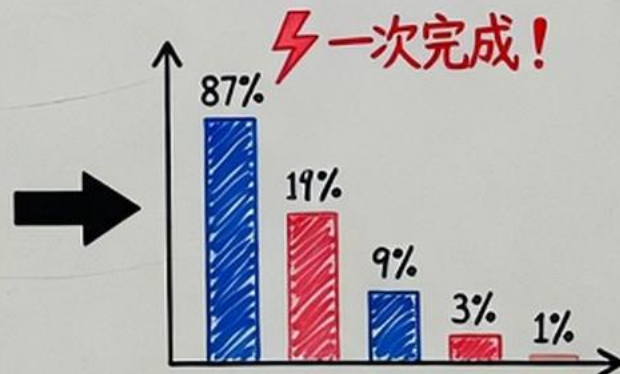


「Jev」✓

- 输出层直接定义在任务候选空间上
- 一次打分得到完整分布
- 没有生成长度随答案增长的问题

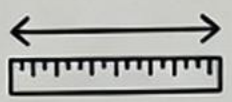


候选空间
是/否
A/B/C/D
42
枚举值



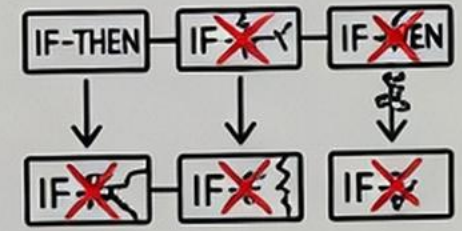
「企业决策自动化的四代定位」

Jev不是从零发明分类



1980s-2000s

第一代



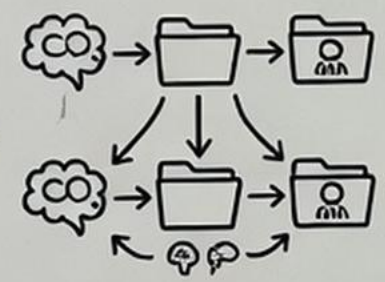
规则引擎

极快但脆弱



2010s

第二代



传统分类器

要为每个任务训练

2022-2024

第三代



大语言模型当万能判断器

零样本强, 但自回归解码和校验成本高

前史



2025+ ☆

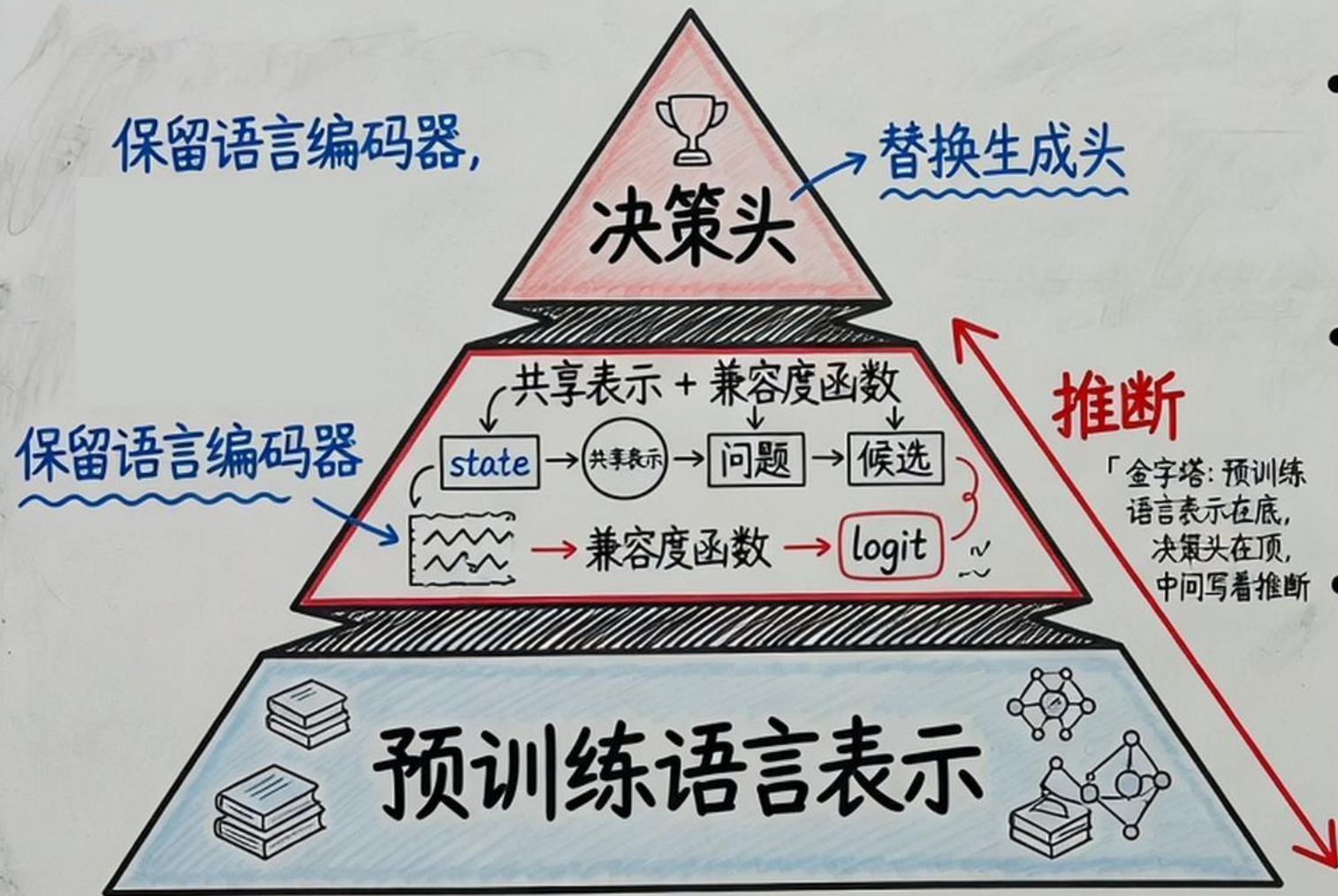
第四代



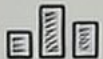
保留自然语言输入

输出头原生限制在运行时可定义的封闭集合上

「最可能的计算结构是推断」💡

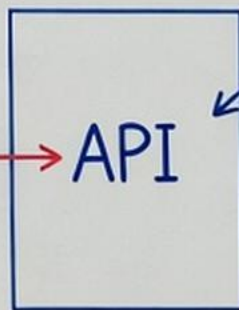
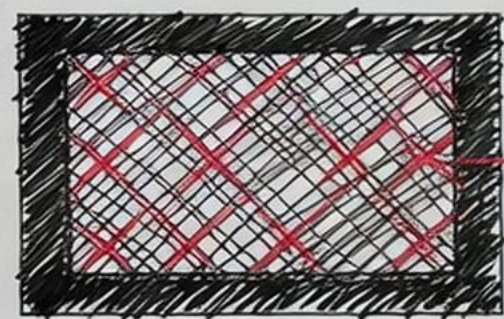


- 可能先将 **state** 编码为共享表示, 再将问题和候选编码后通过兼容性函数产生 **logit**。

- **Choice** 做 **softmax**, **Noul** 做 **sigmoid**, **Score** 对有序等级求期望, 并在张量维上并行。 
- 最经济的路线很可能是保留语言知识和迁移能力, 用合成决策任务和校准目标做后训练。

→ 这就是推断 ✓

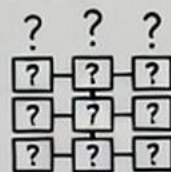
尚未公开的关键内容



接口调用

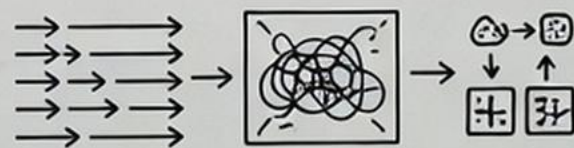
只能看到这里

任何完整复现架构的说法都超过证据 ⚠



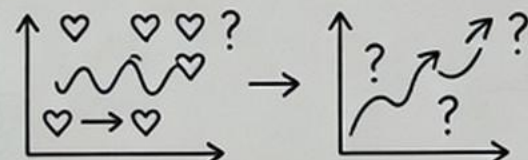
编码器

编码器形态、参数量、预训练语料
蒸馏教师和数据治理方法均未公开。



候选编码 & 并行采样器

计算图、键值复用和硬件利用率均未公开。



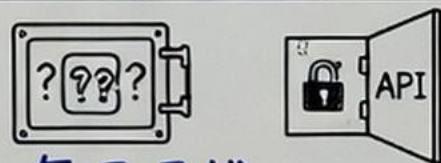
RLCD训练细节

奖励函数、策略更新、校准后处理
也未公开。

当前只能复现接口，不能复现内部模型。



内部模型

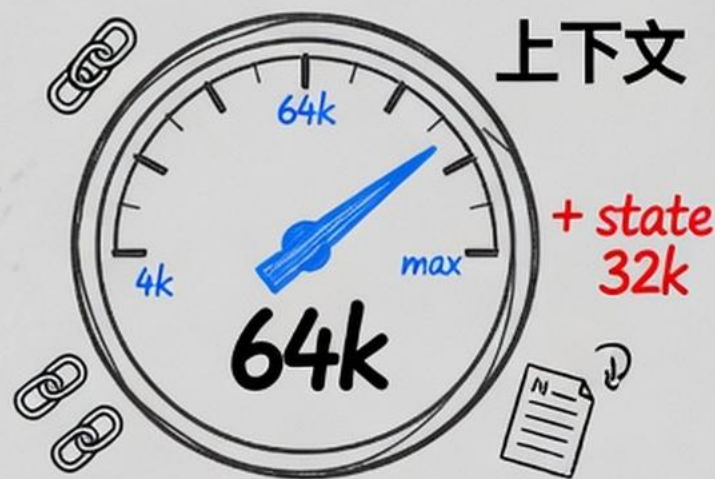


复现现状

接口可调用，核心架构完全黑箱。

Jev 1.13公开规格

以官方模型页为准, 限额可能动态调整 ^{← 随时更新}



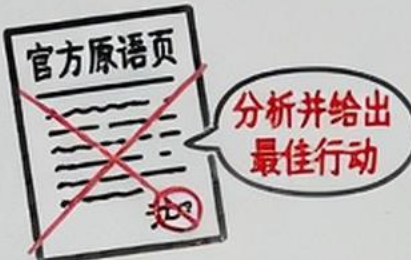
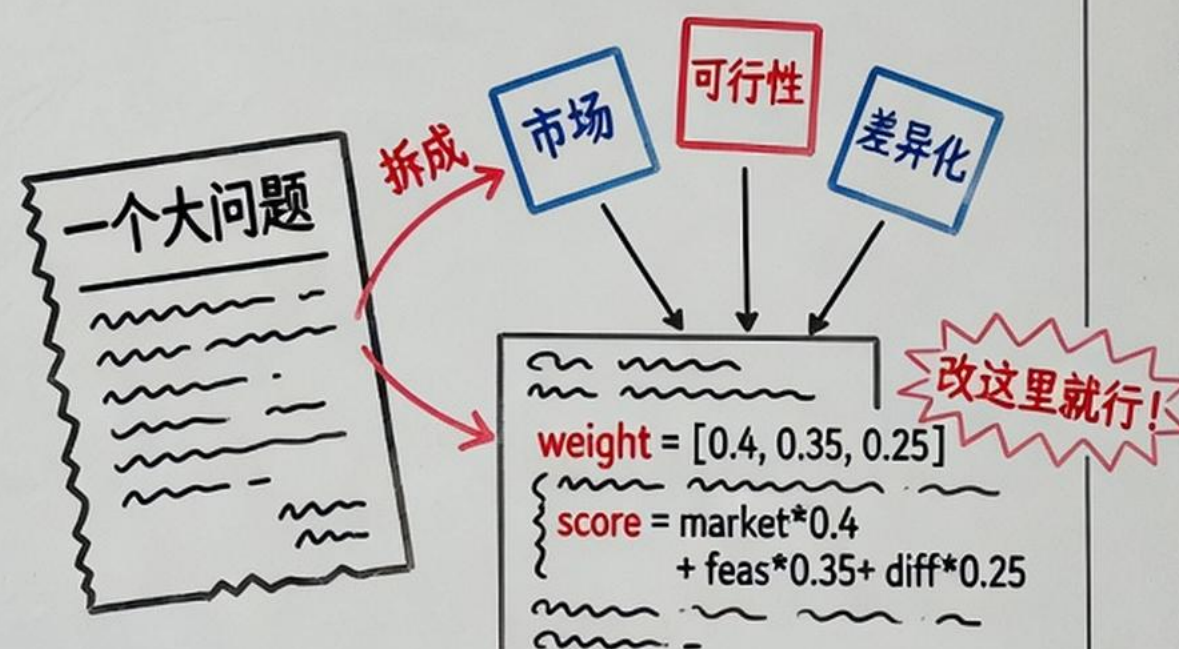
- 接口: `POST /v1/systemone`
- 当前模型: `jev-1.13.0`
- 别名: `jev-latest`、`jev-preview`



- 上下文: 每请求 **64k** token
- **state** + 最长问题 **32k** token
- **限额可能动态调整** ⚠

「一次只问几秒钟能做的判断」

复杂任务拆开，权重写在代码里

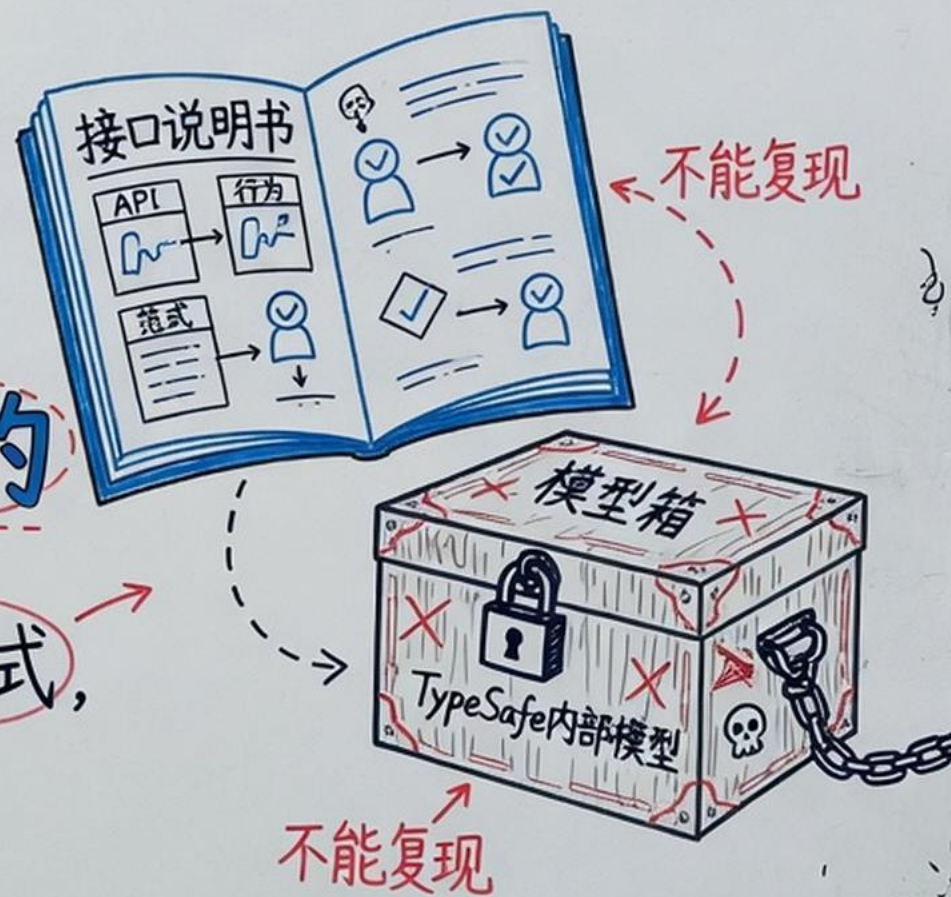
	官方原语页强调 不要把“分析并给出最佳行动”交给模型。	 <pre>weight = [0.4, 0.35, 0.25] score = market*0.4 + feas*0.35+ diff*0.25</pre> 改这里就行!
	分开提问 + 代码加权 市场、可行性、差异化应分开提问，再在代码里加权。 <u>优先级变化时改权重</u>，而不是重写整段提示。	
	Jev的本质 Jev 降低的是单次判断成本，不是问题设计成本。交付物其实是问题与标准。	

第三章

技术本质：

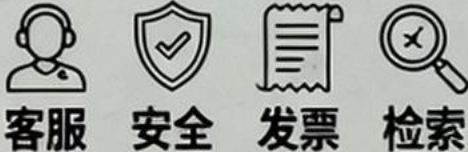
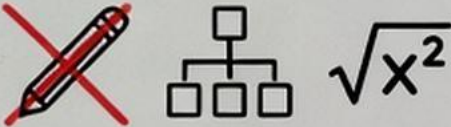
能确认的和不能确认的

- 可以复现的是产品行为和接口范式，不是TypeSafe内部模型。



「需求侧的五个同时成立条件」

→ 缺一条就不要把它当最终求解器

 1. 输入形式 • 输入主要是文本或结构化文本 • 动作空间可预先列举	 2. 判断速度 • 判断可在数秒内完成	 高频 3. 调用频率 • 调用频率高 • 系统允许用置信度分流
 五个条件同时成立 • 缺一条就不要把它当最终求解器	 符合案例 • 客服路由、安全预筛、发票分流、代理工具选择和检索过滤	 不符合案例 ✗ 写作、规划和数学不符合 ✗



三个原语就是全部词汇

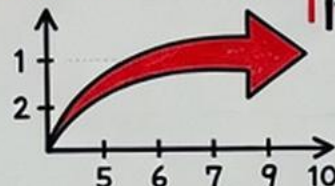
Choice · Score · Noul

Choice



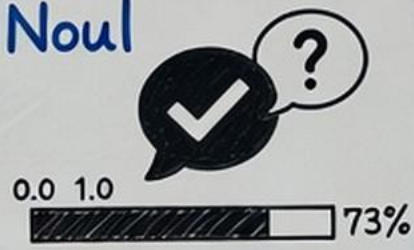
从预定义候选中选一个，最多255个选项，返回选项、全分布和confidence。↪ 全分布

Score



在有序等级上评分，至少两级、最多十级，score是等级编号的概率加权期望，可落在两级之间。

Noul

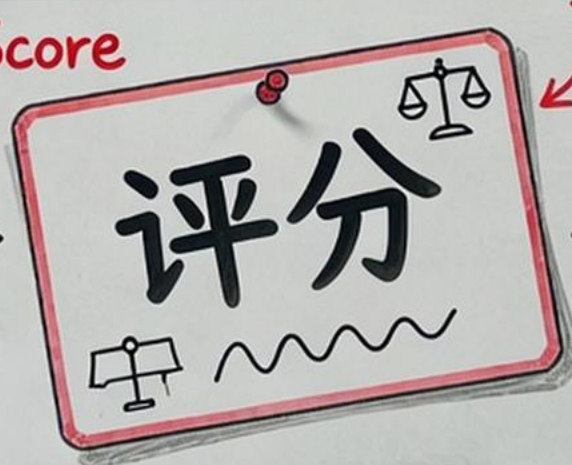


判断命题是否成立，返回0到1的真概率，没有单独的confidence字段。

Choice



Score



Noul



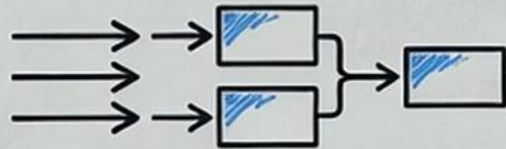
五条可确认的技术机制

全部来自官方接口行为



共享状态编码

一个state可同时对应多个问题，模型宣称**只读取一次状态**。



并行问题求值

并行问题求值、封闭输出空间、**概率原生返回**，以及用代码组合原子判断。



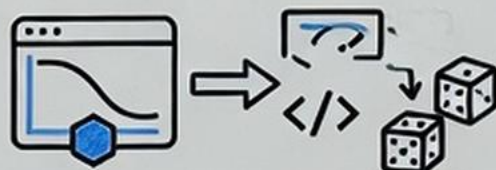
类型错误消除

类型错误在接口层面可**被结构性消除**。



共享 + 并行

同一状态被多个查询安全复用，接口行为可验证。



封闭输出 + 概率

输出空间**完全封闭**，**概率直接从接口返回**。



确定性代码承担

复杂算术与规则仍由代码精确执行。

第五章

官方案例怎么读

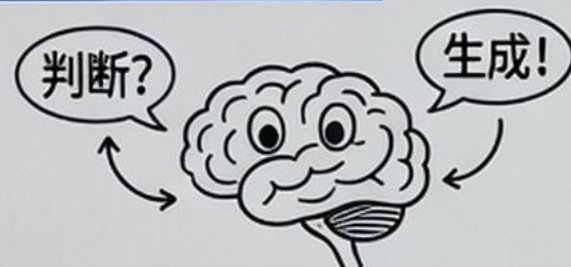


—— 先看工作流为什么可能快，再看为什么不能外推。
→ 演示的限制比演示本身更有研究价值。 →

相邻玩家是分工，不是替代名单

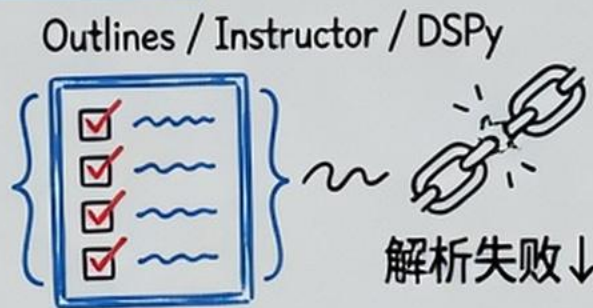
生成、约束解码、经典模型和编排层

通用大语言模型



仍适合判断+生成+推理的混合请求
只是高频判断不划算

约束解码



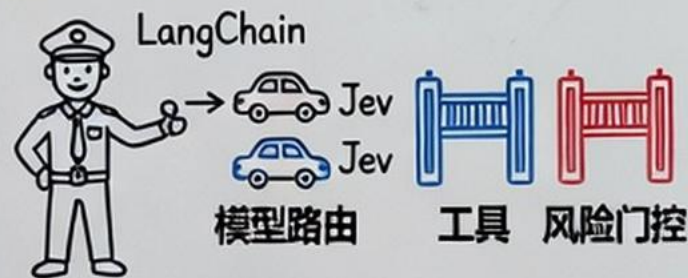
降低解析失败，但通常不把校准当作训练目标

经典模型



四格分工示意

编排层



把Jev放进模型路由和工具风险门控

开源复现证明接口可模仿，
不证明RLCD可复制

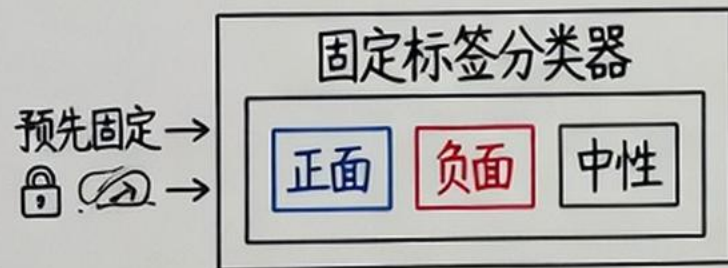


分工 替代

和分类器、和结构化输出各差在哪里

标签可现场定义, 采样原生受限

- ✓ 与传统分类器的差异不是输出为类别, 而是每次请求都能用自然语言定义 新问题和新标签。
- ✓ 与大语言模型结构化输出的差异是输出头与采样过程原生受限, 而不是先生成 token 再验证。
- ✓ 正确格式的答案仍可能语义错误。类型安全不是语义正确。

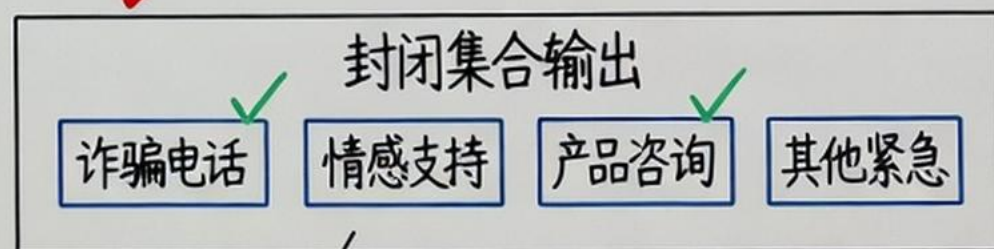
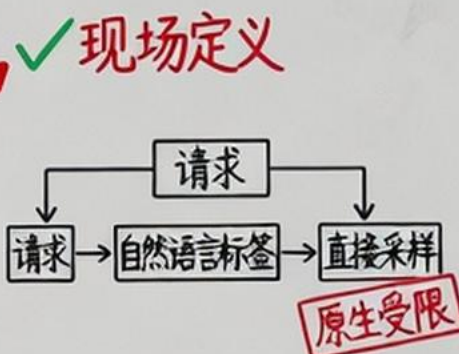


只能选这些



运行时定义的封闭集合

这次我想检测:
诈骗电话 / 情感支持 /
产品咨询 / 其他紧急



「六条得到类型化判断的路线」

差异在输出头，不在会不会分类

规则引擎

极快但脆弱

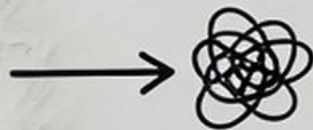
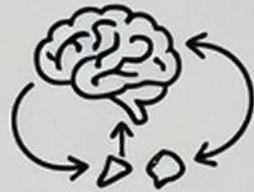
传统分类器要为
每个任务训练



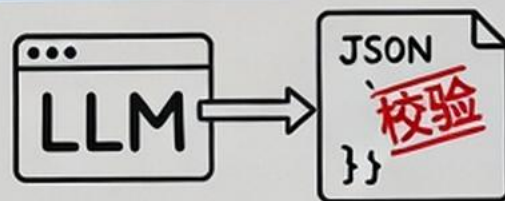
传统分类器

嵌入加线性头

对复杂条件
理解有限

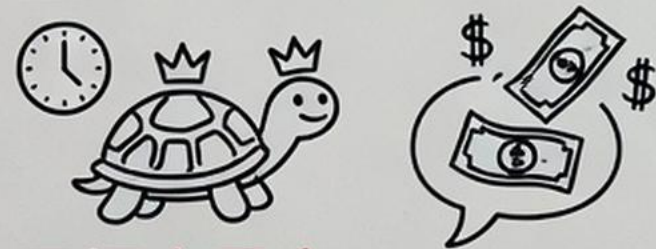


小型大模型



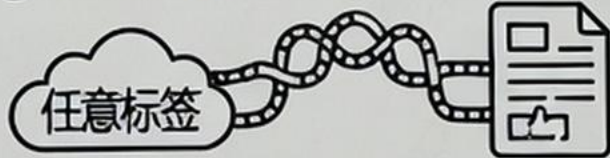
结构化输出仍需解码和校验

推理型大模型



最慢也最贵

Jev



最接近零样本、运行时可定义标签
空间的通用cross-encoder分类器

六格矩阵比较

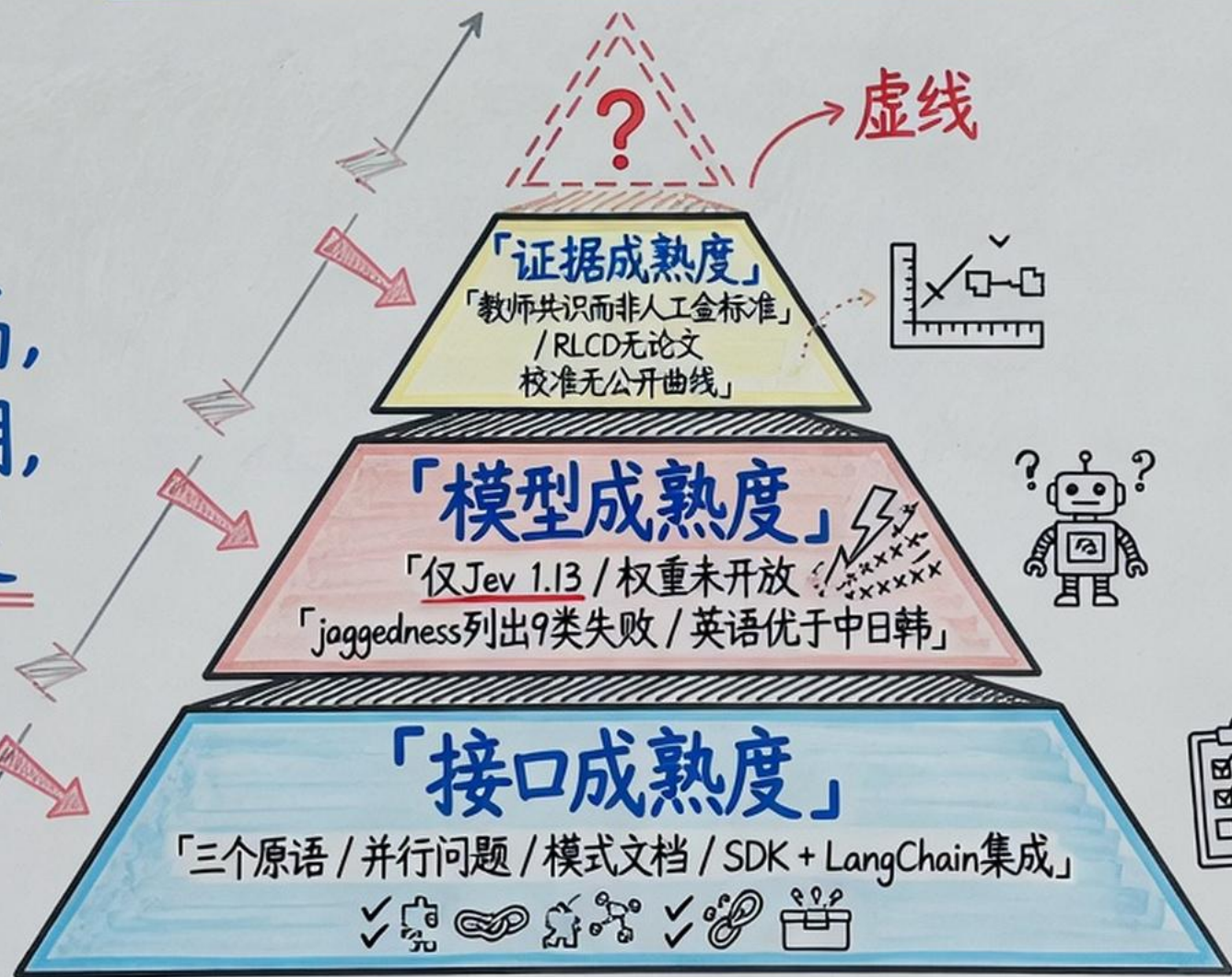
	开放语义	输出空间	速度
规则引擎	/	/	⚡
传统分类器	⚡	/	→
嵌入+线性	✓	×	→
小模型	脸	×	✓
推理大模型	/	—	\$ / \$
Jev	■	→	→



「把成熟度拆成三层来评」



接口较高,
模型早期,
证据不足



- 接口成熟度较高 ✓
- 模型成熟度早期 ⚠
- 证据成熟度不足 ✗

官方尚未公开具体公式?

Confidence是形状压缩，Noul本身就是概率

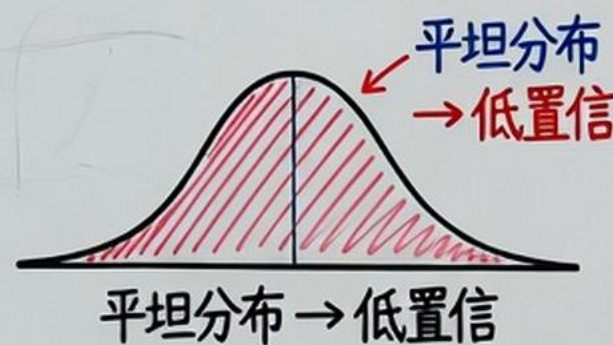
Choice & Score 的 Confidence

- Choice和Score的confidence由输出分布形状压缩得到
- 集中则高、平坦则低



Noul 与 使用建议

- Noul本身就是概率，没有单独confidence
- 调用方可直接使用完整probabilities
- 官方0.5转人 / 0.9高风险仅为教学示例
- 不是普遍最优生产阈值★



直接用probabilities!

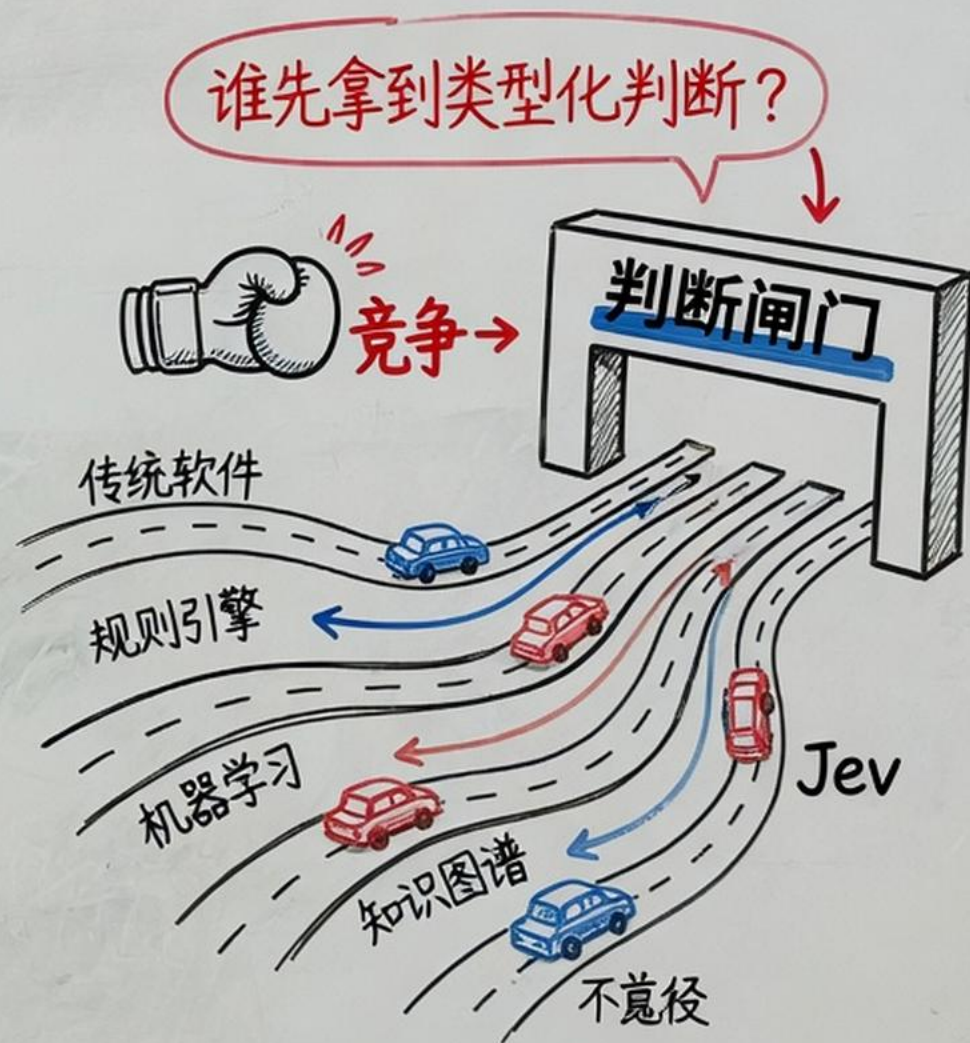
~~0.9~~
非最优

第四章

它在和谁竞争

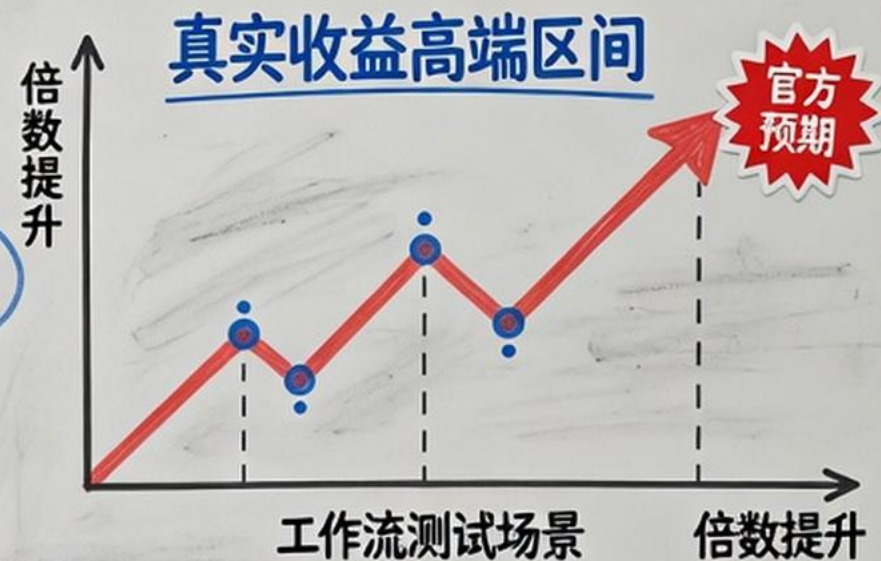
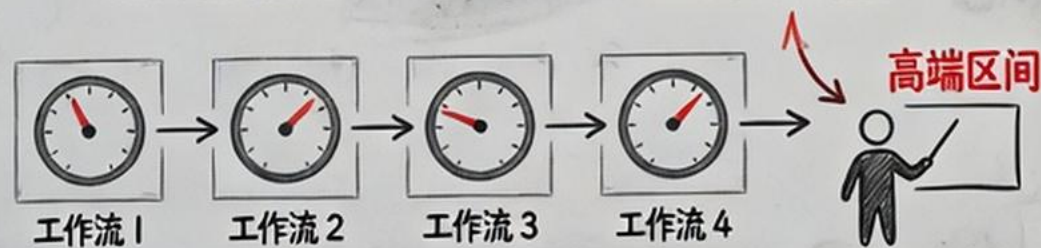
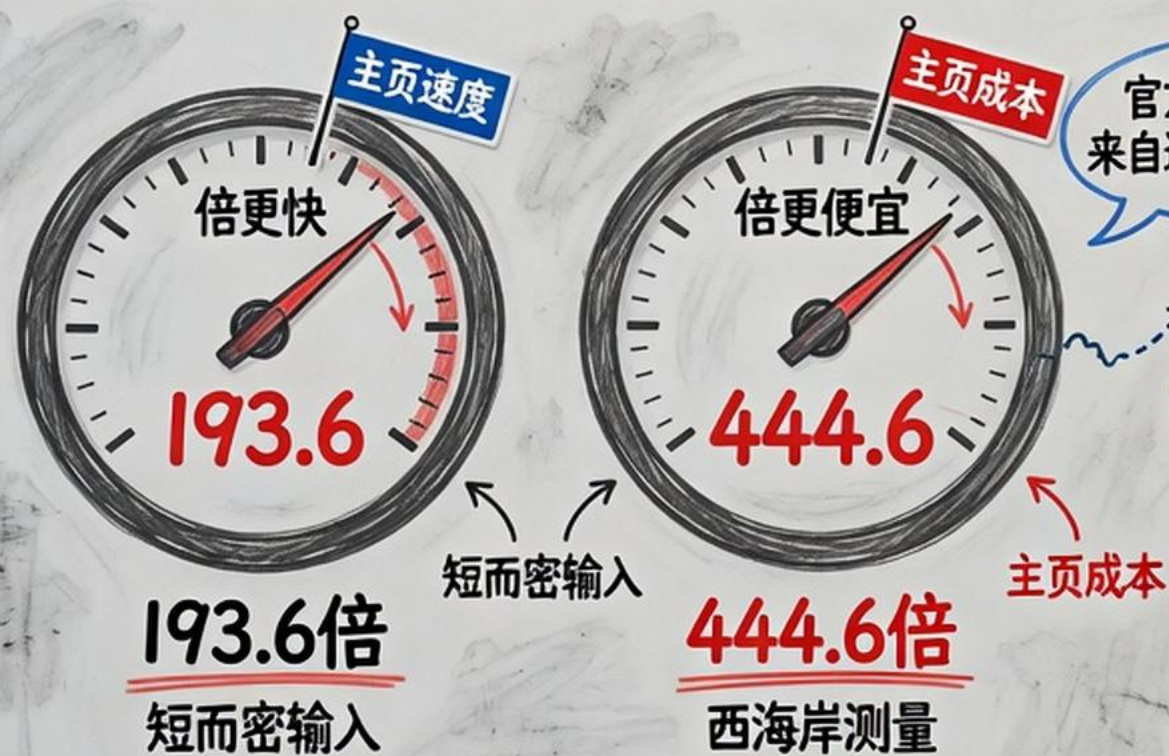
4

Jev不与聊天模型争夺同一槽位。
真正的竞争是软件如何得到一个
类型化判断。



193.6倍与444.6倍从哪里来

四个工作流均值, 官方承认高端区间



- 官方预期这些数字处于真实收益的高端区间
- 短而密输入 + 西海岸位置更有利
- 公司承认无法证明当前价格未被补贴
- 严谨结论: 接口形态有速度和成本潜力
倍率不能跨行业复写

「这些结果为什么可能成立」

机制解释，不是保证



「并行多问题处理」

多个问题共享state
并行输出所有决策



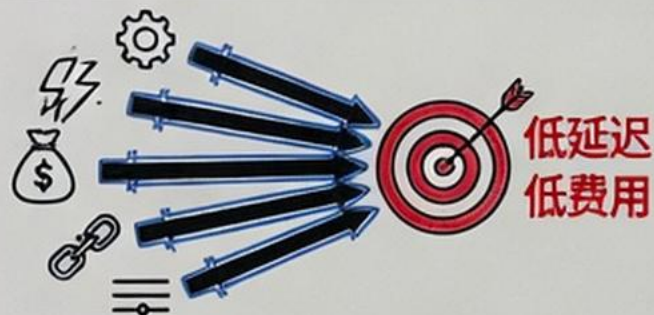
「避免逐token解码」

不生成自然语言 → 低费用
避开长输出费用



「任务拆成浅判断」

较浅判断取代长链推理
结构和算术交给代码



「五条机制箭头」

极低输入价格可能含商业策略



「机制解释潜力」

机制可以解释潜力
不能视为已经普遍兑现



「结构优化」

由代码承担复杂部分
整体保持轻量高效

「官方四个 workflow 评测」

「假设代码正确, 标签来自教师共识」



「安全事件」



→ Analyst

「决定关闭、交给分析师还是立即隔离」。

→ 立即隔离



「轨迹观察」

「决定是否需要人看、以及多快要看」。



「四格业务闸门」



安全



轨迹



发票



客服



「发票处理」

「决定支付、暂扣还是退回」。

→ 支付 / 暂扣 / 退回



「客户服务」

「决定助手下一步该说什么、做什么」。

下一步回复?



* 参考标签是 GPT-6 Astra 与 Claude Fable 5.1 高思考设置的平均概率, 不是人工金标准 *

最适合与最不适合

适合 ✓

- 客服路由
- 舆情标签
- 风控预筛
- 工具选择
- 轨迹异常
- 检索过滤
- 安全守门
- 实体对齐

✗ 不适合 🔒

- 开放写作
- 代码生成
- 复杂数学
- 长链规划
- 创造新动作
- 原生多模态
- 高错误成本且无兜底



第六章

治理要的是熔断， 不是更会说话

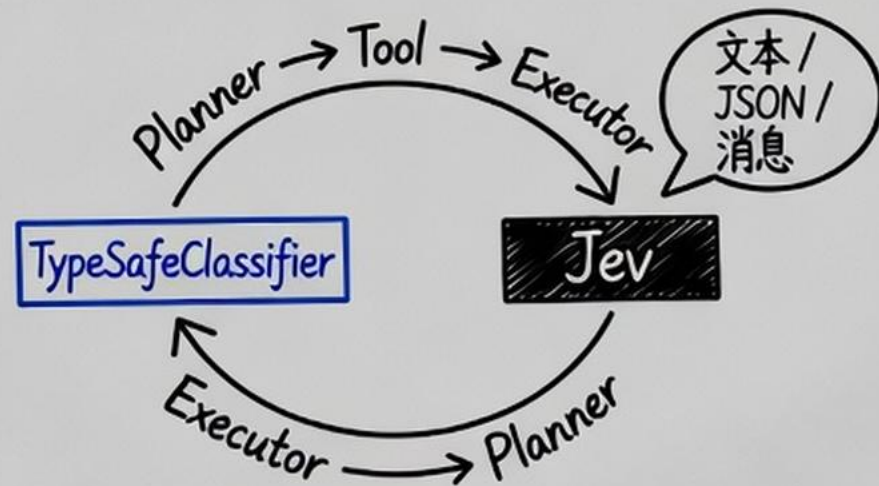


案例：代理循环里的分类与门控

不是替换代理，是给代理加反射

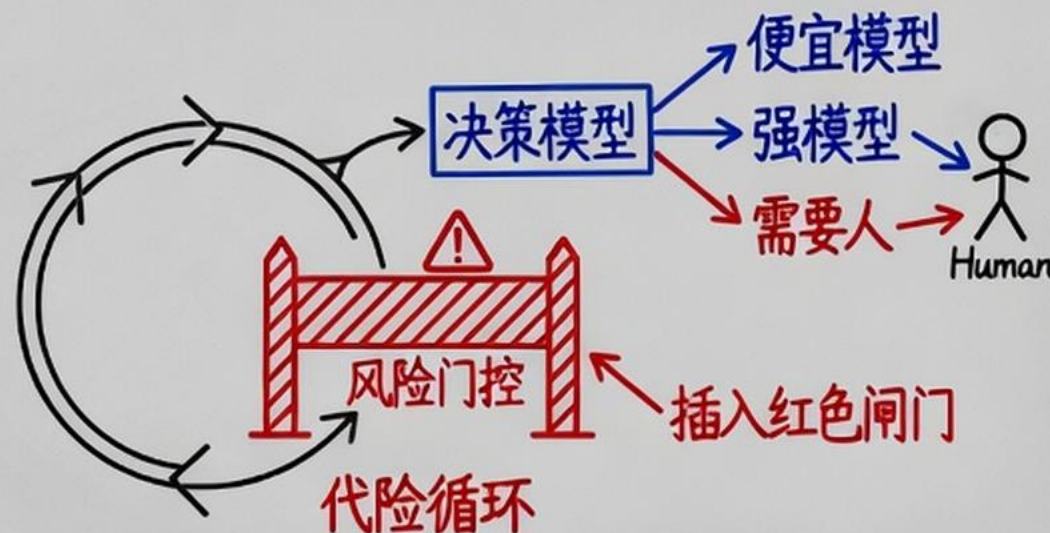
LangChain + Jev 分类机制

- LangChain通过TypeSafeClassifier调用Jev
- state可以是文本、结构化数据或消息
- 模型路由中间件按任务选择更便宜或更强的生成模型



自动模式中间件 + 门控

- 自动模式中间件在工具执行前检查风险
- 开放写作和规划仍由生成模型完成
- 路由、风险门控和是否需要人，交给决策模型



第七章

零幻觉的边界 和九类失败

失败模式

一块布满缺口的齿轮

jaggedness 文档比营销页更有技术价值。

失败模式是产品说明书的一部分。



@ 清新研究团队 | 2026年9月

「官方承认的九类失败模式」

适用于 jev-1.13, 审阅日期 2026-09-17 



「字面理解」

「必须靠改问题或改代码处理」。



「数学计数」

「必须靠改问题或改代码处理」

九格失败模式看板

字面

计数

日期

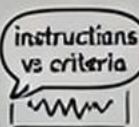


多跳



对抗内容

长状态

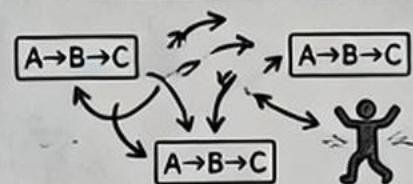


最值得警惕



「日期比较」

「必须靠改问题或改代码处理」



「多跳间接」

「必须靠改问题或改代码处理」



「长状态干扰」

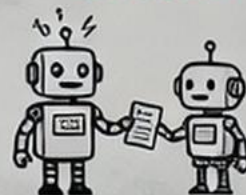
「必须靠改问题或改代码处理」

instructions:
必须客观

criteria:
必须有趣

会混乱

对抗内容可以移动答案。生成任务应交给生成模型。



「结构不变量」

「互补问题的概率未必相加为1」。

采用Jev时的六项合规设计



把自动、升级和人工写成显式分支



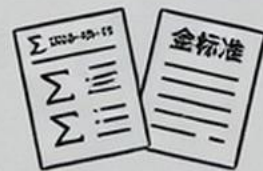
按风险类别分别设阈值

固定模型版本，不要在调好阈值后
跟随jev-latest漂移。



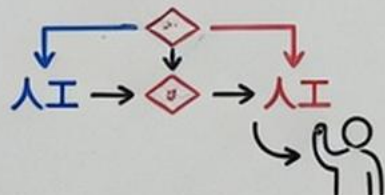
完整记录轨迹

记录state、问题、答案、**概率**、阈
值和最终动作，保证**回放**。

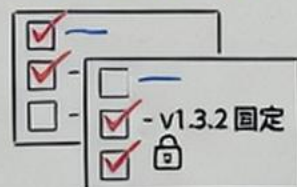


结构一致性测试

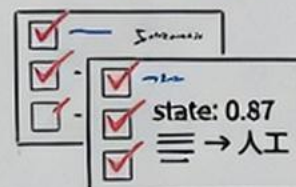
做结构一致性测试。数学和日期留在代
码里。中文业务必须用中文金标准重测。



六项合规



六项合规



「结构不变量缺口」条件评分, 不是完整世界模型

同一工单不同问法

- 同一张工单上, Noul 问「是否要求退款」得到 0.22
- Choice 的「yes」只有 0.01 (官方表格中的例子) → **概率对不齐**



同一工单概率总和异常

- 另一张工单上:
 - 退款 Noul = 0.72
 - 非退款 Noul = 0.47
 - 两者之和 = 1.19 (加法已按原文验算)

1.19 ≠ 1



05



五个团队首创概念 思想高地 → ★

这些概念用于把Jev从产品新闻提升为
可迁移的研究框架，不是TypeSafe官方术语。

核心

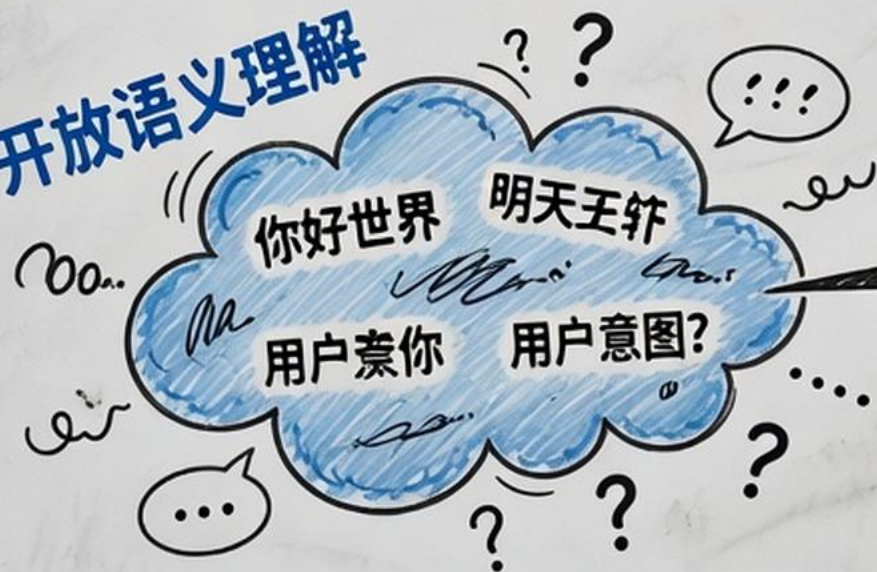


封闭决策契约

开放语义理解 + 运行时封闭输出 + 概率可执行

团队首创概念

开放语义理解



契约



封闭选项盒



选项A  68%

选项B  27%

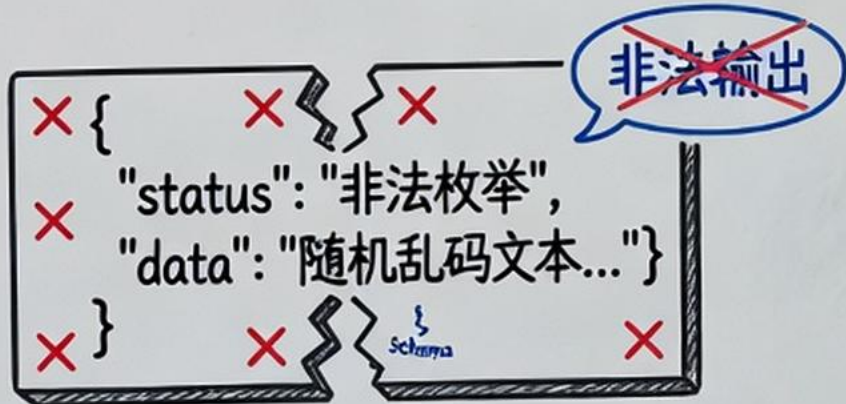
选项C  5%

- 大语言模型的产品契约是生成字符串。软件还要再签一份隐含的解析契约。
- Jev把这份隐含契约提前写成接口。调用方给出合法答案集合，模型只能在该集合上返回分布。
- 契约边界之外没有答案。系统创新首先是契约创新，不必等待一种全新的神经元发明。

零非法输出, 不是零幻觉

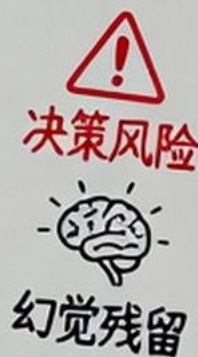
故障从不可控文本

- 它不能生成不存在的字段、非法枚举或一段脱离格式的文本
- 类型幻觉可以被结构上消除
- 官方把类型错误画成零, 并写明该数字不是经验统计, 而是由 schema 匹配保证



收缩成可测量的决策错误

- 它仍可能在证据不足时选错类别
- 误解否定词, 把恶意文本当指令
- 或以 很高概率 支持错误判断

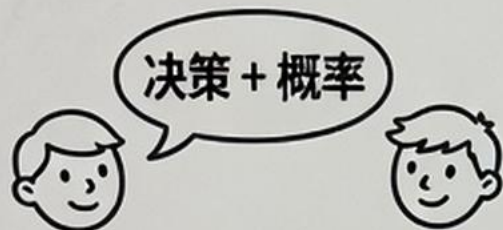


RLCD还是目标名称，不是可复现算法



(公式位置空白)

有名称和目标，没有论文和配方



RLCD 的含义

让模型同时输出决策
及与真实正确率匹配的概率



没有论文和配方

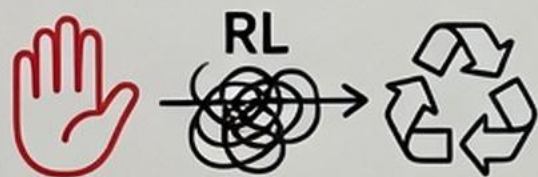
独立综述明确说明：
无论文、无数据集、无训练配方、无校准图



✓ 正确性
✓ 选择性风险
✓ 计算成本

可检验目标

必须同时考虑正确性、校准、
选择性风险和计算成本，使用适当评分规则



名称 ≠ 新算法

不能因为名字里有「强化学习」
就把方法登记为新算法贡献



避免名称混淆

2023年另有同名对比蒸馏论文
必须写全称：**RLCD (目标名称)**



核心提醒

RLCD 是目标名称而非可复现算法

「零非法输出」

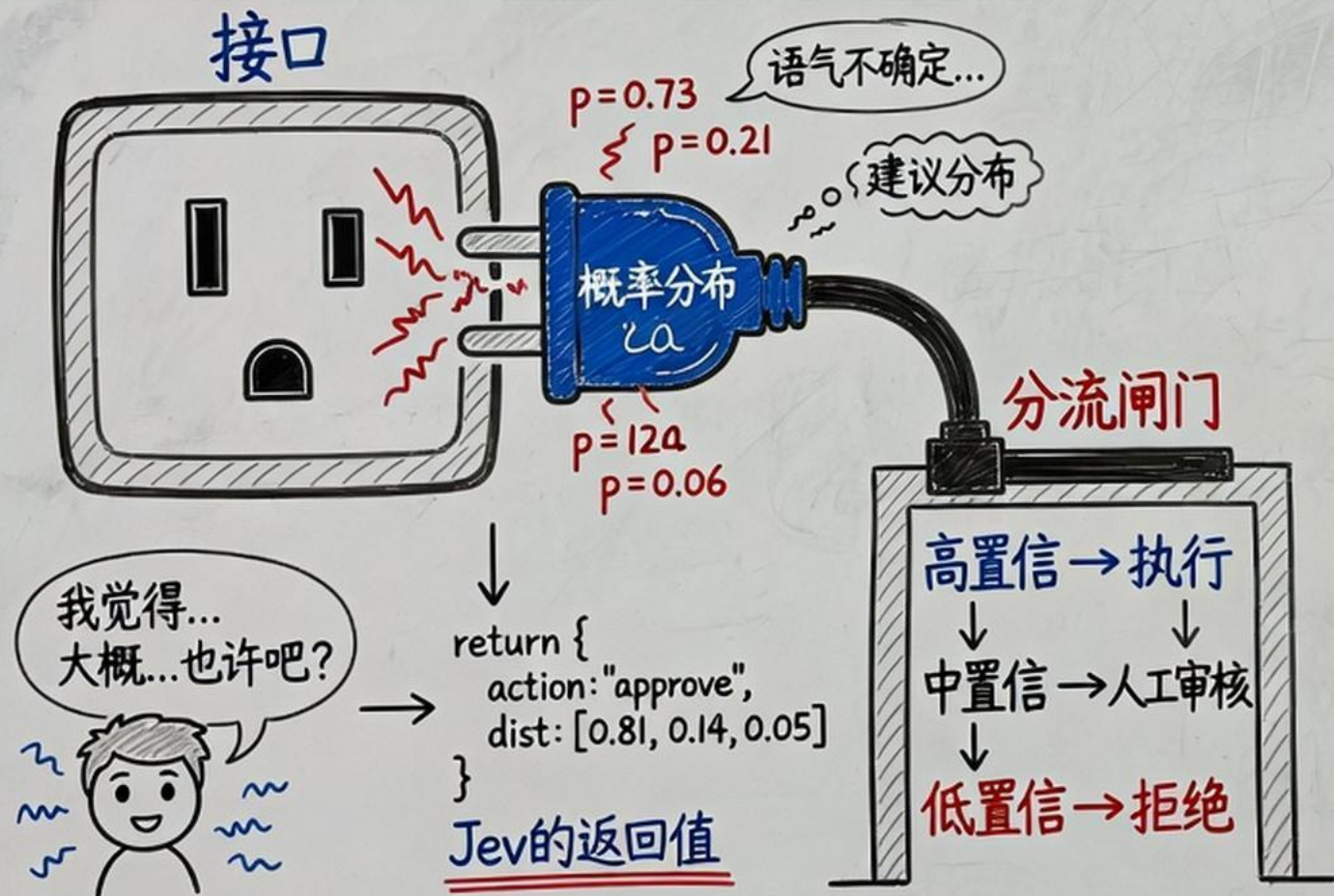


「零非法输出 ≠ 零幻觉」

- 模型不能发明第四个枚举，不能返回无法解析的字段，但可以把错误的合法选项说到很高的概率。
- 故障从语法层被压缩到语义层。这让错误可测、可门控、可回放，但不是消灭错误。
- 不要验收是否幻觉，要验收非法输出是否为零、语义错误率在给定覆盖率下是否低于错误预算。验收重点

概率即接口

团队首创概念

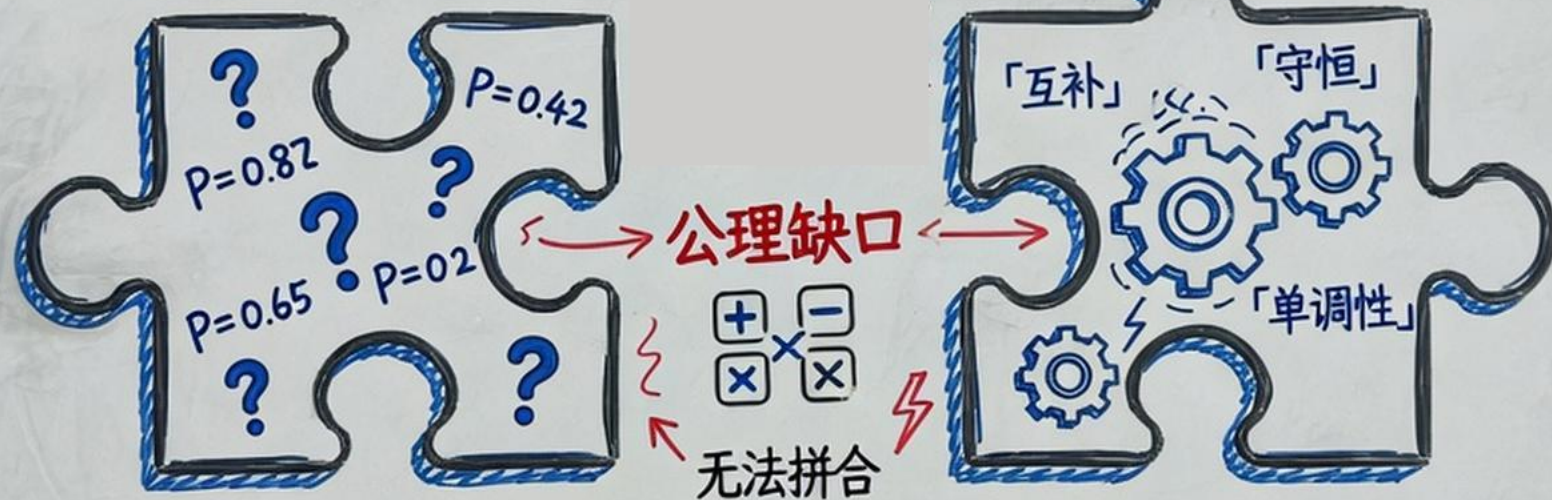


- 聊天模型的不确定性藏在语气里。
Jev把不确定性做成返回值。
- 真正的决策权在代码, 模型只提供建议分布。
接口形状不等于校准完成。
- 概率一旦成为接口, 校准就不再是模型训练的配菜, 而是软件正确性的一部分。

「结构不变量缺口」

团队首创概念

条件评分 - 世界模型公理



提问形式条件评分

- 当前输出更像绑定于提问形式的条件评分,
- 而不是满足互补、守恒和单调性的世界模型。

- 缺口不否定产品价值,
- 但否定把一次请求里的多个概率当成 **可以互相加减的物理量**。⚠️
- 安全关键系统必须在单一问法下校准阈值, 并用**代码强制一致性**。
(if (query != same) reject)

「四层计算栈」

团队首创概念

发现、判断、执行、验证

验证层
真实结果测量残差并回流

残差

执行层
程序做算术、约束和动作

判断层
决策模型做反残语义判断
高速反射层

决策模型做高频语义判断

发现层
强生成模型提出新问题

发现层
强生成模型提出新问题和新流程

Jev 适合判断层，
不是自进化系统本身。
它可以成为自进化系统里
的高速反射层。

..‘回流’

证据分级：哪些是强事实

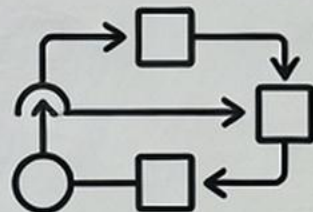
接口可集成，科学仍待公开 →



交通灯矩阵

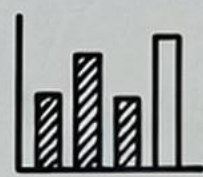
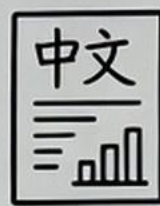
类型安全

由封闭schema保证，是强结论。
语义零幻觉与jaggedness矛盾，
不成立。 ❌



官方 workflow 倍率

可信但只适用于目标任务形态。
RLCD数学形式未公开，尚非可
复现科研贡献。 未公开 ❌



中文能力

必须在中文业务数据上重测。
架构是否全新 不能下结论。



接口可集成



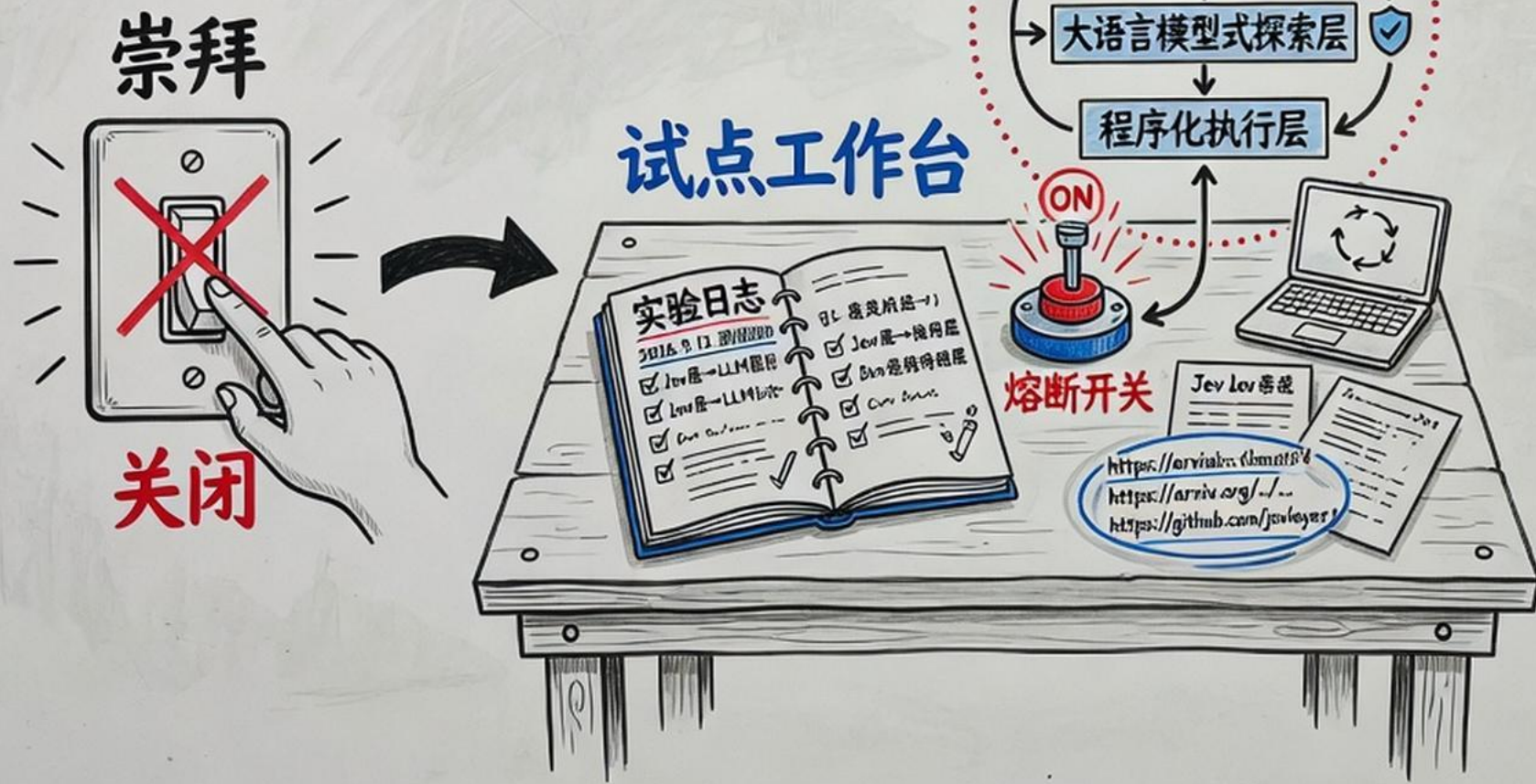
倍率可信



未公开训练



清新研究团队 | 2026年9月



- 把Jev式反射层、大语言模型式探索层、程序化执行层与**虚实验证器闭环**统一起来，才是更完整的自进化基础设施。

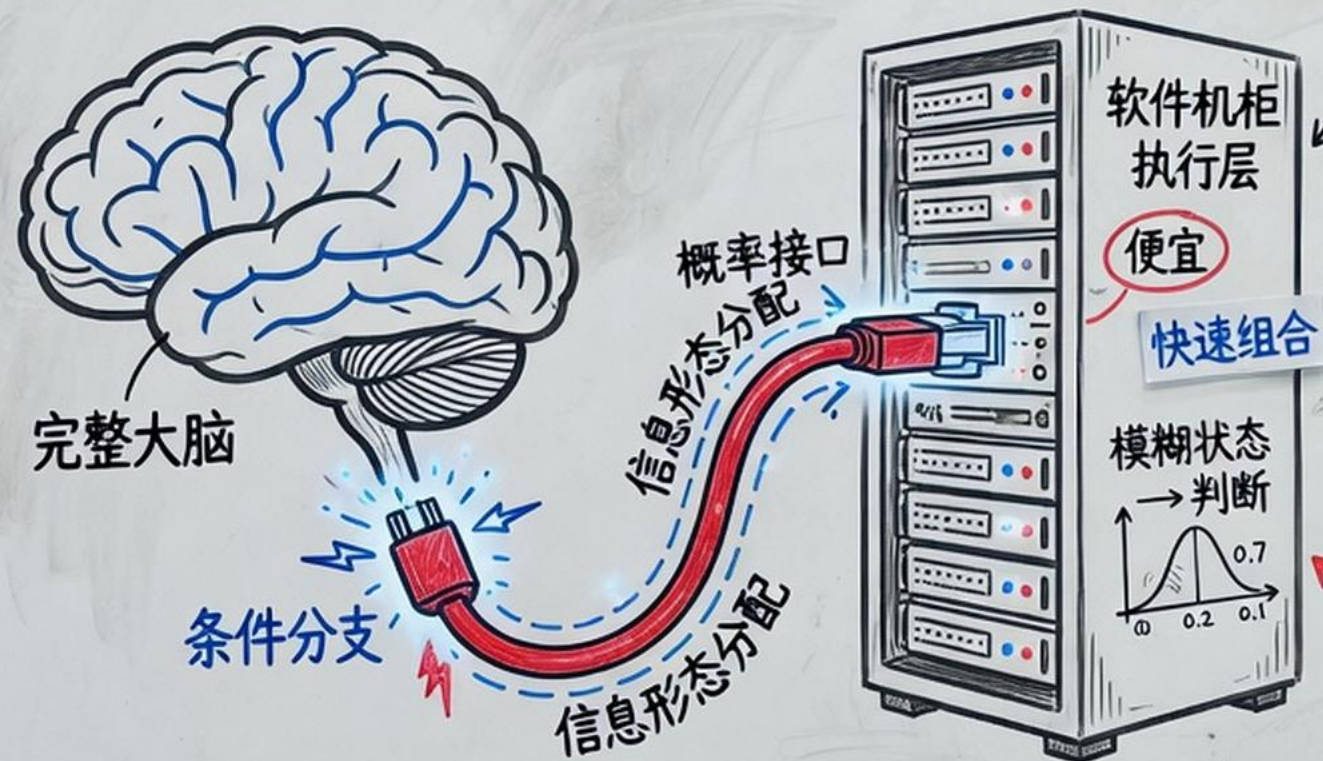
- 凡是不能指出具体网址的数字，都不进入结论页。凡是涉及内部网络结构的句子，都**加推断标记**。

- 框架一旦成立，后续只需替换证据，不必**重写立场**。

@清新研究团队 | 2026年9月

神经条件分支，不是完整大脑

承认接口革命，拒绝未公布的科学革命



- Jev是重要的系统创新，也可能包含重要的模型创新，但公开证据目前只足以确认前者。
- 大量企业任务并不需要会说话的模型，只需要对模糊状态做快速、可组合、带概率的判断。🎲🎲
- 最值得复制的是按信息形态分配模型、把概率当成接口、把稳定能力编译到更便宜的执行层。



「比Jev更进一步的六条研究机会」

复制原则，超越缺口



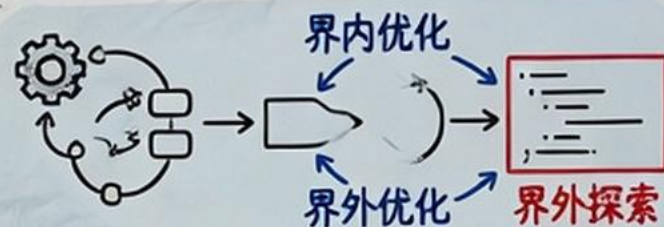
「可生长决策空间」

新候选必须经过验证才能进入执行空间。
跨原语一致性：互补、包含、单调、守恒。



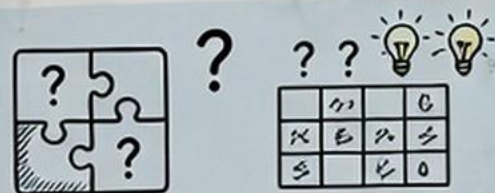
「风险条件校准」

按行业、群体、时间和成本函数分别校准。
虚实残差进化 → **真实业务差值**



决策到程序自动编译

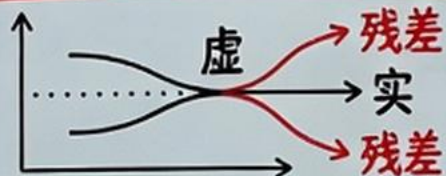
以及界内优化、界外探索的递归拓界。



「跨原语一致性」

守恒与单调约束确保复制稳定

六格研究机会看板



「虚实残差进化」

用真实业务差值当训练信号



递归拓界

界内优化 ↔ 界外探索自动切换

建议的四周独立复现

用风险覆盖而不是平均准确率做验收 ✓ 风险覆盖

