

# 人类判断系统白皮书 v1.0 (中文版)

—在 AI 时代重建判断的社会基础设施

## 摘要 (Abstract)

随着人工智能系统在信息分析、决策建议与自动执行中的能力迅速增强，人类社会正进入一个前所未有的阶段：决策正在被自动化，而不可逆后果仍由人类承担。

当前的技术系统在计算、执行、监控方面高度发达，却缺失一个关键层级——对“人类判断是否真实发生”的结构性保障。

Human Judgment System (HJS) 提出一种新的基础设施范式：

将判断 (Judgment) 从个人心理活动，提升为可被识别、强制、记录与验证的公共结构。

HJS 并不试图替代人类判断，也不试图优化决策结果。

它只解决一个被长期忽视、但在 AI 时代变得致命的问题：在不可逆决策发生之前，是否真的存在过一个承担责任的人类判断？

## 第一章 | 问题的本质：判断正在消失

### 1.1 决策 ≠ 判断

现代系统高度擅长“决策”：

- 预测结果
- 计算概率
- 推荐路径
- 自动执行

但判断并不是决策。

判断只在以下条件同时成立时出现：

1. 信息不完备
2. 后果不可回退
3. 价值发生冲突
4. 必须有人承担责任

判断的本质不是“选择最优解”，而是：在不确定性中，明确承担后果。

## 1.2 AI 扩张带来的结构性风险

AI 的能力扩张带来一个隐蔽但危险的趋势：

- 决策链条被拉长
- 责任被稀释
- 判断被默认“包含在系统中”

当系统失败时，我们常问：

- 模型是否准确？
- 数据是否充分？
- 参数是否合理？

但很少问：是否真的存在过一个人类判断？

在许多系统中，答案是否定的。

## 1.3 判断的消失不是技术问题

判断消失的原因不是 AI 太强，而是：

- 判断从未被当作“系统层”对待
- 判断始终停留在“个人行为”层面
- 没有任何机制验证它是否发生

这在 AI 时代不再成立。

# 第二章 | 一个被忽略的基础设施层

## 2.1 现代系统的层级结构

现代社会基础设施通常包含：

1. 感知层（数据、传感器）
2. 计算层（模型、算法）
3. 执行层（自动化、行动）
4. 监控层（审计、日志、合规）

但缺失一个层级：判断层（Judgment Layer）

## 2.2 判断层的定义

判断层不是提供建议，也不是给出答案，而是：

- 识别不可逆节点
- 强制暂停自动化
- 要求人类显性承担
- 记录判断的存在性

判断层不评估“对错”，只确认：是否存在过承担责任的主体。

## 2.3 为什么判断必须被“强制”

如果判断是可选的，它就会在效率压力下被跳过。

就像安全带一样：

- 它并不保证你不会出事故
- 但它强制你在风险前承担约束

判断层的目标不是更快，而是不允许无判断的不可逆行为发生。

# 第三章 | Human Judgment System (HJS) 概述

## 3.1 系统目标

HJS 的核心目标只有一个：确保不可逆决策之前，存在且仅存在一次可验证的人类判断。

## 3.2 HJS 不是什么

HJS 不是：

- 决策 AI
- 智能代理
- 风险评估工具
- 自动化系统

HJS 不提供“答案”。

## 3.3 HJS 是什么

HJS 是一种协议级基础设施，用于：

- 判断的触发
- 判断的约束
- 判断的记录

- 判断的追溯

## 第四章 | 判断的最小结构单元

### 4.1 Judgment Object (判断对象)

每一次判断至少包含四个不可删除字段：

1. Context (上下文)
2. 决策所处的现实条件
3. Irreversibility Declaration (不可逆声明)
4. 明确指出哪些后果不可回退
5. Responsibility Statement (责任声明)
6. 明确谁承担后果
7. Time Anchor (时间锚点)
8. 判断发生的时间

### 4.2 不评估内容，只验证存在

HJS 不判断内容是否“合理”。

它只回答一个问题：判断是否真实发生过？

## 第五章 | 判断账本 (Judgment Ledger)

### 5.1 为什么需要账本

如果判断无法被验证，它就可以被事后伪造。

账本的作用是：

- 固化判断发生的事
- 防止事后篡改
- 提供社会级可信性

### 5.2 与区块链的关系

HJS 可以使用区块链，但不依赖特定链。

区块链的作用仅限于：

- 时间戳
- 不可篡改
- 公共可验证

HJS 不是金融系统，也不发行代币。

## 第六章 | 社会关系学与法学视角

### 6.1 判断是社会行为

判断从来不是纯粹的个人行为，而是：

- 被社会认可
- 被制度承载
- 被责任绑定

HJS 的目标，是让判断重新回到社会结构中。

### 6.2 法律意义

HJS 不替代法律，而是提供：

- 判断存在性的证据
- 责任边界的前置声明
- 法律系统的补充输入

## 第七章 | 应用场景（非穷举）

- 医疗不可逆治疗
- 金融高风险操作
- AI 自动执行系统
- 基因与生物技术
- 公共政策模拟
- 个人重大人生决策

## 第八章 | 为什么现在必须提出

因为：

- AI 正在加速
- 自动化正在渗透
- 不可逆决策越来越多
- 责任却越来越模糊

判断不能被延后重建。

## 第九章 | 开放性与治理原则

HJS 是一个：

- 开放协议
- 可被任何人实现
- 不属于任何公司
- 可被社会共同演化的结构

## 结语

未来的风险不来自 AI 做错了决定，而来自没有人真正做过判断。

Human Judgment System 提出的不是答案，

而是一个时代必须补上的结构。

版本：v1.0

协议状态：公共提案

许可：CC BY-SA 4.0