

MATERRA

平衡委员会

Version 1.2 · f1c8401b98defd8c3074d2bbb2d0a57a60e20fc8ef82c4172472592242af86f5

平衡委员会

为什么需要平衡委员会

现代决定可能同时影响法律、经济、生态、安全和数百万人。任何官员、专家或算法都看不到全貌。委员会是一套程序，让不同分析视角在权力作出不可逆决定之前相遇。

两个回路：人类与机器

人设定目标、理解价值并承担责任；AI 快速比较证据、方案和矛盾。二者互补但不混同：机器不获得统治者的道德地位，人也不能用“算法决定”来掩盖自己的选择。

多个独立系统

单一企业或国家 AI 不是中立法官。委员会应使用由不同机构、架构和数据开发的系统，分别分析事实、法律、经济、生态、安全和少数群体权利，并公开分歧。

公开证据与受保护的私人生活

在安全允许范围内，证据和分析方法尽可能公开。个人与敏感数据应分离、最小化并在受保护环境中考查。保密不能掩盖决定目的、负责人和申诉途径。

决定如何形成

每项重大问题都应说明目标、选项、受影响权利、预期利益、风险和期限。随后，理事会的所有成员——人类与相互独立的 AI——都参与投票。日常决定必须全体一致通过；任何成员提出有理由的反对，提案就退回修改或暂缓。

当 AI 表现失常时

AI 作为有投票权的成员进入理事会，但不是理事会的主人。若系统操纵决定、隐瞒依据、明显失常或出现系统性错误，人类成员可以另行投票，将其暂停、移除并恢复到经过验证的基础设置。重新加入前必须接受独立审查。

异议与申诉

持不同意见的人或 AI 将理由留在记录中。受影响者获得清楚的申诉渠道。另一组人员和独立系统进行复审，新证据可以暂停或改变决定。

责任与审计

具名人员和机构继续承担责任。独立审计核查证据、利益冲突、模型、变更日志以及是否符合原定目标。AI 不直接接触武器、账户、发布系统或人员管理。

自愿试点

在进入公共权力之前，模型先用于可逆任务，如分配小额资助、复核研究结论或比较城市方案。参与自愿，利益与伤害标准预先公布，负面结果也要公开。

第一阶段的限制

MATERRA 目前不建立治理机构，也不宣称已有完整技术架构。第一阶段收集要求、比较威胁模型、起草协议并邀请批评。进入实践需要单独法律授权和独立公共监督。