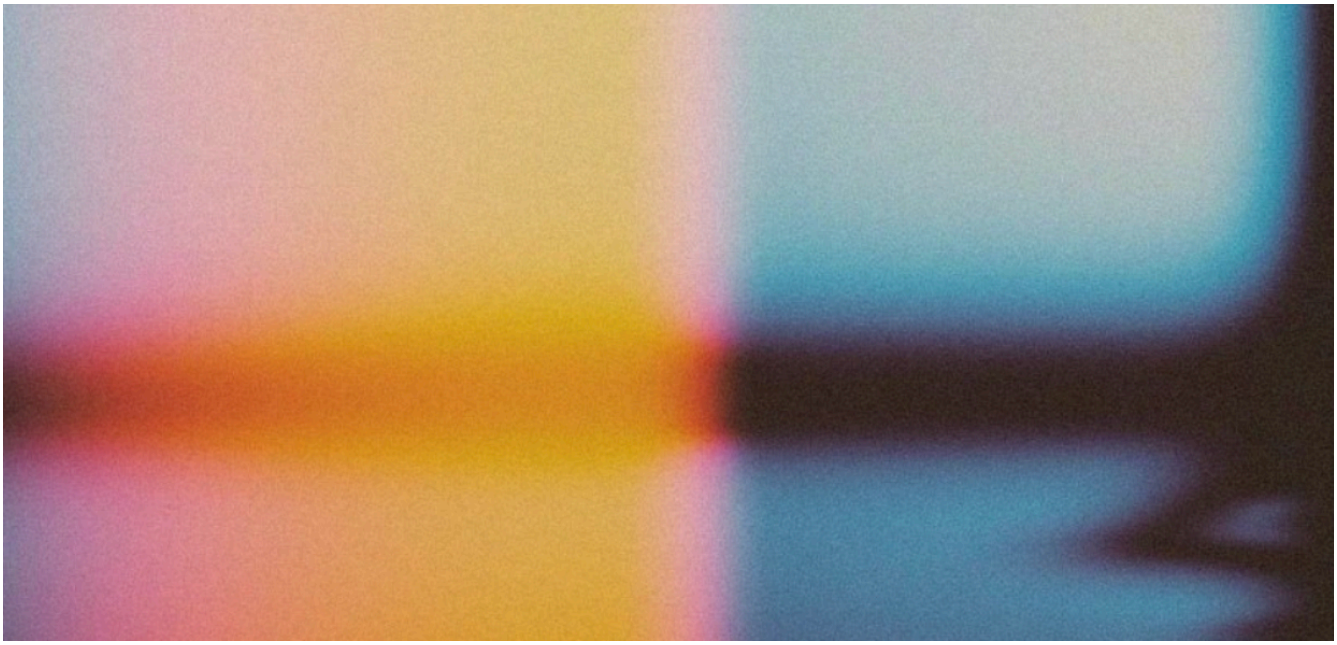
超越人在环路

May 17, 2026



图片来源: Rik Oostenbroek - @RikOostenbroek

AI 协调工具可以帮助组织更有效地对齐、审议和协作。但它们也会改变人类合作所依赖的信任信号和团队动态。

因此，领导者面对的挑战，不只是把 AI 部署到所有能节省时间的地方，也不是坚持每一个 AI 流程都必须有人在环路中。真正的问题，是有意识地设计人类、AI 系统和混合团队各自在何处最能创造价值。

Google DeepMind 的 *Organising Intelligence* 在第 9 个观点 “AI-driven Coordination” 中很好地提出了这个问题。它认为，围绕 AI 协作的研究正在汇聚到三个问题上：AI 是否具备成为良好协作者所需的社会认知能力；AI 是帮助还是伤害人类协调；以及人类-AI 混合团队是否优于纯人类团队。

本文的核心观点是：AI 的组织价值不仅在于自动化，更在于重新设计合作本身。

从自动化到协调

大多数关于 AI 的高管讨论，仍然从生产率开始：哪些任务可以自动化，可以节省多少时间，哪些工作流可以加速。这个框架有用，但并不完整。组织失败并不只是因为任务太慢。它们失败，是因为团队彼此误解，激励不一致，知识困在孤岛里，分歧在被有效呈现之前就已经僵化。

这正是为什么 AI 更深层的组织承诺，可能不是自动化，而是协调。Sangeet Paul Choudary 在《哈佛商业评论》中提出，AI 的巨大回报可能来自降低协调成本，而不仅仅是替代工作 (Choudary, 2026)。在这种视角下，AI 成为团队之间的翻译层、分歧中的调解者，以及让隐含假设显性化的机制。

近期研究支持这种更宽的协调视角。在民主审议中，研究显示 AI 系统可以帮助人们找到参与者共同认可的共识表述 (Tessler et al., 2024)。另一些研究表明，AI 来源可以提高人们对相反观点的开放性 (Lu, Tormala, & Duhachek, 2025)。放在组织语境中，这指向一种强大的可能性：AI 或许能帮助群体从立场对抗，转向对假设、证据和取舍的结构化比较。

AI 还可以丰富组织理解自身社会系统的方式。Amir Goldberg 和 Sameer Srivastava 认为，AI 可以通过识别语言、意义和互动中的模式，加深我们对组织文化的理解，而这些模式在大规模情境下通常很难观察 (Goldberg & Srivastava, 2024)。这很重要，因为协调很少只是正式流程。它受到文化塑造：人们说什么，避免说什么，哪些群体容易跨边界翻译，以及共享语言在哪里断裂。

社会认知问题

如果 AI 要支持协调，它就不能只是检索信息。它必须参与工作的社会结构。这要求它在某种程度上能够推理信念、意图、情绪和视角。

因此，关于大语言模型心智理论能力的研究与此相关。Street 及其同事发现，LLM 在高阶心智理论任务上可以达到成年人类水平，这类任务需要递归地推理别人知道什么、相信什么或会推断什么 (Street et al., 2024)。关于共情沟通的相关研究也表明，在特定条件下，大语言模型或许可以可靠地判断共情沟通 (Kumar et al., 2026)。

但这不应被解读为 AI 在人类意义上“理解”组织。社会意义总是嵌入文化语境之中的。关于跨文化面部表情的研究显示，情绪信号的解释有多复杂、多依赖情境 (Brooks et al., 2024)。对领导者而言，含义是：AI 协调工具应被视为强大但局部的社会工具。它们适合用来呈现视角、总结分歧和支持共情，但不应被当作人类意义的最终解释者。

隐藏的信任问题

AI 协调的乐观论据很强。但合作不仅是信息问题。它也是信任问题。

当人们协作时，他们依赖可见的努力、判断和意图信号。我们判断是否信任一位同事，往往不只看答案质量，也看他们如何得到这个答案：他们是否真正和问题搏斗过？是否考虑了取舍？是否理解后果？是否暴露了自己的假设？

Wojtowicz 和 DeDeo 将这称为 “mental proof”：一种可观察的思考证据，使他人能够推断技能、用心和正直等原本不可见的特质 (Wojtowicz & DeDeo, 2024)。AI 可能会削弱这种证明，因为它让思考变得更容易、更快速，也更不可见。一个由 AI 辅助生成的精致答案或许很有用，但它也可能遮蔽了呈现这个答案的人是否真正行使了判断。

这正是 AI 协调的核心悖论。AI 可以在任务层面让协作更容易，却在社会层面让合作更困难。它可以减少产出过程中的摩擦，却移除那种帮助人们相互信任的可见挣扎。

这并不意味着组织应该避免 AI，恰恰相反。它意味着领导者需要有意识地设计信任信号。AI 支持下的工作可能需要新的溯源形式：可见的假设、审计轨迹、异议记录、置信水平、来源轨迹，以及人类判断与 AI 生成综合之间的明确区分。重点不是为了放慢组织速度而放慢。重点是保留协作所依赖的社会证据。

混合团队并不会自动更好

“人在环路”这个说法，常常被当作一种普遍安全原则。它听起来令人安心：AI 可以行动，但人类仍会参与其中。然而，证据表明，人类参与并不总是自动带来增益。

一项关于人类-AI 组合的系统综述和元分析发现，混合团队经常优于纯人类团队，但并非总是如此。在某些情况下，尤其是当 AI 系统在某项任务上已经优于人类时，加入人类反而会相对于 AI 单独执行降低表现 (Vaccaro, Almaatouq, & Malone, 2024)。这使那种“每个重要流程都应该有人类检查点”的简单治理答案变得复杂。

正确的问题不是：“是否应该有人在环路中？”更好的问题是：“人类在这里到底要做什么？”

有时，人类应该决策。有时，人类应该审计。有时，人类应该设定目标、价值和约束。有时，人类应该监控异常。有时，AI 应该生成选项、批判假设，或在专家群体之间进行翻译。而在某些低风险、边界清晰的场景中，AI 可能最适合在没有持续人类打断的情况下执行。

这也是 AI 团队研究变得重要的地方。Schmutz 及其同事认为，AI teaming 要求我们在数字时代重新思考协作本身，包括角色、相互依赖、信任和团队设计 (Schmutz et al., 2024)。AI 不是一个简单放进现有团队的工具。它会改变团队本身。

任务张量：一种更丰富的设计语言

我与 Anil Doshi 关于 “Human–AI Task Tensor” 的工作，提供了一种有用方式，让我们跳出粗糙的“人在环路”思维。这个框架从八个维度组织人类-AI 工作：任务定义、AI 集成、交互方式、审计要求、输出定义、决策权、AI 结构和人类角色。

这很重要，因为不同工作需要不同的人类-AI 配置。高容量、定义清晰、容易审计的任务，可能适合高度自动化。战略性、模糊、政治敏感的任务，可能需要 AI 支持，但决策权仍由人类掌握。创造性或审议性过程，可能受益于 AI 生成替代方案，但仍需要人类保留问责、解释语境和协商意义。

我们提出了人类-AI 二元关系中 20 个决策权层级。例如，它区分了人类在环路中给予批准、人类在环路之上拥有否决权、人类在接近环路的位置进行观察，以及更高自主性的配置。

这个区别至关重要。人在环路之上 并不等同于 人在环路中。在人在环路中的流程里，人类是主动决策周期的一部分。人在环路之上的流程中，AI 可以拥有更高自主性，而人类负责监督、干预或在必要时否决。这为领导者提供了更丰富的设计词汇，用来根据任务风险、可审计性、表现和问责要求匹配监督方式。

领导者应该设计什么

实践层面的教训是：AI 协调需要被设计，而不是临时拼凑。

第一，领导者应该映射任务，而不是只映射岗位。一个角色中可能包含需要不同人类-AI 配置的任务。有些任务适合自动化；有些需要增强；有些需要审议、争辩和人类问责。

第二，领导者应该定义 AI 的协调角色。AI 是职能之间的翻译者，分歧群体之间的调解者，假设的批判者，决策的记录者，选项的推荐者，还是自主执行者？每一种角色都会产生不同的风险和信任要求。

第三，领导者应该保留 mental proof。当 AI 帮助生成输出时，团队仍应能够看到推理过程：假设、考虑过的替代方案、使用的来源、作出的判断，以及不确定点。在高信任团队中，AI 可能加速协作。在低信任团队中，若 AI 辅助不可见，人们可能更加怀疑，除非过程被设计得足够清晰可读。

第四，领导者应该把混合团队当作真正的团队来设计。这意味着设计角色、交接、激励、升级路径和问责规范。问题不是 AI 是否在场。问题是这个人类-AI 系统是否协调良好。

最后，领导者应该抵制自动化与人类控制之间的虚假二分。任务张量指向一种更细致的架构：人类可以定义目标，AI 可以生成选项；人类可以审议，AI 可以模拟后果；人类可以决策，AI 可以监控执行。在其他场景中，AI 可以在边界内决策，而人类在环路之上进行监督。

结论

下一代 AI 赋能组织，不会是那些自动化最多决策的组织。它们会是最善于设计人类、AI 系统和混合团队之间认知劳动分工的组织。

AI 协调工具可以帮助组织审议、对齐、翻译和协作。但它们也会重塑一种社会证据，人们正是通过这种证据判断他人是否有能力、是否真诚、是否值得信任。天真地使用 AI，可能会让工作更快，却削弱合作。深思熟虑地使用 AI，它可以成为一种新的协调架构：既提升集体判断，也保留让协作成为可能的人类信号。

目标并不是默认更快地工作。目标是设计得更好的合作。

参考文献

- Brooks, J. A., et al. (2024). Deep learning reveals what facial expressions mean to people in different cultures. *iScience, 27*(3), 109175.
- Choudary, S. P. (2026). AI's big payoff is coordination, not automation. *Harvard Business Review*.
- Doshi, A. R., & Moore, A. P. (2026). Toward a Human–AI Task Tensor: A Taxonomy for Organizing Work in the Age of Generative AI. In *Handbook of Artificial Intelligence and Strategy*. Open version: SSRN.
- Goldberg, A., & Srivastava, S. B. (2024). How Artificial Intelligence Can Enrich Our Understanding of Organizational Culture. *Management and Business Review, 4*(2), 32–37.
- Google DeepMind. (2026). *Organizing Intelligence: 16 big ideas from frontier AI research that will redefine how we build, lead, and scale*.
- Kumar, A., et al. (2026). When large language models are reliable for judging empathic communication. *Nature Machine Intelligence, 8*(2), 173–185.
- Lu, L., Tormala, Z. L., & Duhachek, A. (2025). How AI sources can increase openness to opposing views. *Scientific Reports, 15*, 17170.
- Schmutz, J. B., et al. (2024). AI-teaming: Redefining collaboration in the digital era. *Current Opinion in Psychology, 58*, 101837.

- Street, W., et al. (2024). LLMs achieve adult human performance on higher-order theory of mind tasks. *arXiv*.
- Tessler, M. H., et al. (2024). AI can help humans find common ground in democratic deliberation. *Science, 386*(6719), eadq2852.
- Vaccaro, M., Almaatouq, A., & Malone, T. W. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293–2303.
- Wojtowicz, Z., & DeDeo, S. (2024). Undermining mental proof: How AI can make cooperation harder by making thinking easier. *arXiv*.