

AI多智能体系统的架构分类法：AI和人应该如何分工

Original 熊节 极限编程合作社 2025年11月7日 10:38 上海

当你用AI工具辅助研究或分析任务时,会遇到一个现象:有时候让AI全自动运行,几个小时就能完成高质量的工作;但有时候,即使投入大量算力,产出的结果也是浅薄的、缺乏洞察的。更糟的是,你很难事先判断哪种情况会发生。

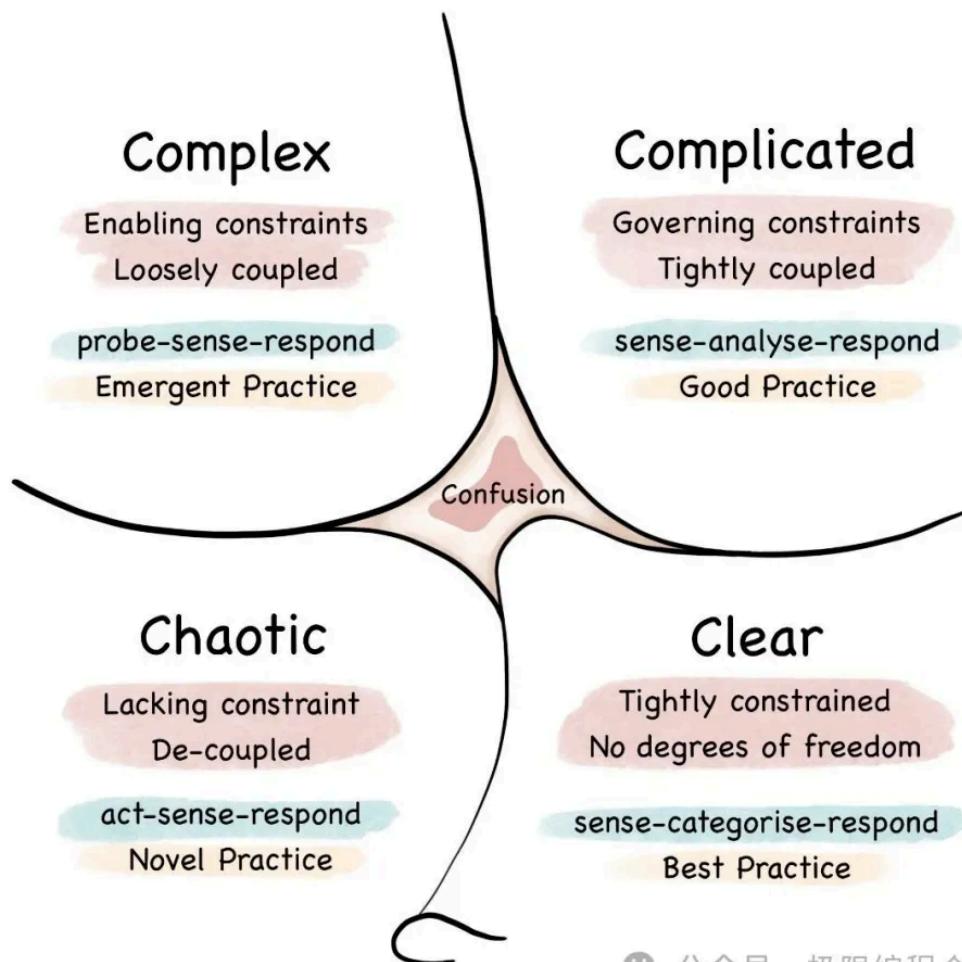
这个困惑的根源在于:**不同性质的任务,需要根本不同的MAS(多智能体系统)构型和人机协作方式**。但我们缺乏一个系统化的框架来判断任务性质,选择合适的设计策略。

本文引入Cynefin框架,提出一个MAS架构分类法,帮助你:

- 诊断任务属于哪个类别
- 选择合适的系统构型
- 明确人机协作的界面
- 理解系统应该如何演化

Cynefin框架:问题性质决定方法选择

Cynefin框架由Dave Snowden提出,它的核心洞察是:**问题的性质决定了合适的决策模式**。框架将问题情境划分为四个域:



公众号 · 极限编程合作社

Cynefin Framework

- **Clear(明确域)**: 因果关系清晰,存在最佳实践。决策模式:感知–分类–响应
- **Complicated(复杂域)**: 因果关系存在但需专家分析。决策模式:感知–分析–响应
- **Complex(涌现域)**: 因果关系只能事后识别,需要探索。决策模式:探测–感知–响应
- **Chaotic(混沌域)**: 因果关系不可辨识,需立即行动。决策模式:行动–感知–响应

在MAS设计中,这个框架帮助我们回答核心问题:**任务能不能自动化?自动化到什么程度?人在哪里介入?**

关键洞察:系统的边界是研究过程本身,而非研究对象。我们设计的MAS不是模拟或直接作用于研究对象,而是辅助研究过程——从搜寻材料、分析数据、构建理论到撰写报告的整个认知活动。

Clear域:全自动化流水线

任务特征

所需信息明确存在,获取路径清晰,处理流程简单直接,不需要判断和选择。

典型场景：批量下载文件、格式转换、数据清洗、简单的信息提取。

判断标准

- **信息存在性**: 文件已存在于特定位置,位置可确定
- **因果关系清晰**: 访问URL → 下载文件 → 存储,步骤明确
- **不需要专家判断**: 不涉及内容评估、质量鉴别
- **流程可标准化**: 整个过程可以编写为明确的算法

业务架构



单阶段管道：

输入：URL列表 + 命名规则



处理：遍历URL → 下载文件 → 按规则命名 → 组织目录



输出：规整的文件集合

架构特征：

- 单一职责,线性流程,无复杂分支
- 确定性:相同输入产生相同输出
- 无人机协作点:完全自动化

真实案例:News on China期刊下载

News on China是一份周刊,过往所有期刊以PDF形式存储在网站上,URL遵循固定模式。研究者需要这些期刊作为分析媒体话语演变的材料。

执行流程：

1. 人类定义:URL模式、时间范围、命名规则、存储位置
2. AI执行:生成URL列表(约250个文件)、并行下载、按规则命名、生成日志
3. 结果:耗时约15分钟,人类节省约3小时手动下载时间

人机协作模式

- **设计阶段**(人类主导):定义目标URL、文件命名规范、存储位置
- **执行阶段**(AI全自动):无需人类介入
- **验证阶段**(人类抽查):检查文件数量和内容

这是最简单的协作模式:人类定义任务,AI执行,人类验证。没有迭代和反馈循环。

设计要点

- **明确性优先**: 完全明确URL、命名、存储结构后再执行
- **可恢复性**: 实现断点续传,记录已下载文件清单

- **速率控制**: 避免被识别为恶意爬虫
- **工具选择灵活**: 简单任务无需复杂架构,Claude Code单次会话可完成即可

Complicated域:可并行的分析流水线

任务特征

需要对信息进行专业判断和解读,但评估框架本身是稳定的、预先确定的。分析路径清晰,但应用需要专家级能力。

典型场景: 指标评估、合规检查、多维度分析。

判断标准

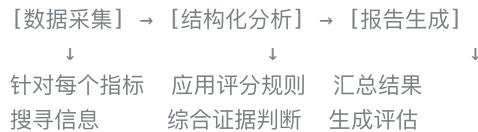
- **信息存在但需解读**: 信息在互联网上存在,但需要专家能力来理解和评估
- **因果关系明确但需分析**: 分析路径清晰,但需要判断力
- **需要专家判断**: 需要综合证据,做出程度判断
- **框架预先确定**: 评估维度和指标体系预先设计好,不是在评估中涌现

与Clear的区别: "某国是否有A法律"(是/否)是Clear;"A法律执行效果如何"需要综合证据判断,是Complicated。

业务架构



三阶段管道:



架构特征:

- 流程清晰,可分解为独立阶段
- 不同指标可并行处理(充分利用MAS能力)
- 每个阶段产出可验证的中间结果
- 人类可在阶段间审查,但通常无需介入

真实案例:数字主权指数(DSI)评估

DSI是一个评估国家数字主权的指标体系,包含4个维度、16项指标。针对每个指标细化了15–20个具体问题。

评估流程:

1. **数据采集**:针对每个问题,从政府网站、政策文档、新闻报道等渠道采集信息
2. **分析评分**:根据预定义的评分规则,对每个指标逐一评分(需要判断"数据本地化执行程度")
3. **生成报告**:汇总16个指标的得分,生成综合评估

关键特征:

- 要采集什么数据是清晰的:每个指标对应明确的问题列表
- 怎么分析是清晰的:每个指标有明确评分规则
- 可以并行:16个指标的评估互不干扰,每个国家需访问数千网页,但可并行执行

| 变体:动态问题生成

有时用于数据采集的细粒度问题需要根据研究对象动态生成。

比如评估"农业生产情况":尼泊尔应该问"水稻种植面积"、"稻米年产量",加纳应该问"可可豆年产量"、"可可出口额"。

这种情况下,流程变成:**overview → 问题生成 → 数据采集 → 分析 → 报告**

先做粗粒度调研,了解国家农业结构,然后生成针对性问题,再进入采集和分析。

但这仍属于Complicated域,因为问题生成虽然动态,但可通过规则和上下文自动完成,无需人工介入。

| 人机协作模式

- **设计阶段**(人类主导):定义评估框架(维度和指标)和评分规则
- **执行阶段**(AI主导):数据采集和分析评估
- **审查阶段**(人类验证):验证分析质量,检查框架应用是否正确

人类可以验证中间结果,但不需要在流程中介入。

| 设计要点

- **任务分解**:识别可独立处理的子任务
- **并行执行**:充分利用多智能体并行能力
- **规则显式化**:评分规则清晰建模,便于调试和优化
- **中间结果可验证**:每个指标的评估结果可追溯,便于人工审查

Complex域:人机协作的迭代流程

| 任务特征

AI可以处理大量信息,但关键洞察必须由人提供。理论框架不能预设,必须在理解具体情境后才能确定。答案在探索中涌现。

典型场景：需要理论框架指导的深度分析、需要编辑判断的内容策展。

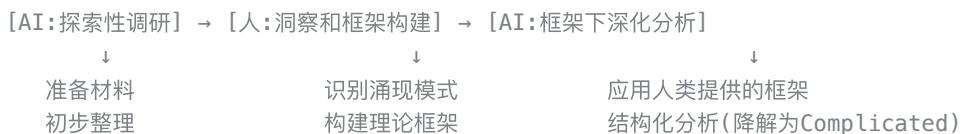
判断标准

- **信息存在,但意义需涌现:** 信息存在,但其意义和模式只有通过整合理解才能浮现
- **因果关系只能事后识别:** 不知道应该用什么理论视角,框架需要在过程中"发现"
- **无法通过应用既有框架得到:** 框架本身需要被构建
- **需要人类创造性洞察:** "灵光一现"的理论创新,看到表面文本之下的深层关联

业务架构



人机协作的三段式：



架构特征:

- 有明确的人工介入节点(中间环节)
- 上下游环节解耦:第二段不关心第一段如何产生材料,只要材料格式合适
- 人提供的洞察/框架成为下游分析的输入
- 系统演化:人的介入将Complex任务降解为Complicated任务

真实案例1:南非的阶级与矛盾分析

要做一个国家的阶级分析(马克思主义意义上的),关键特征是:**必须有适用于这个国家的理论框架,不能拿通用框架套所有国家。**

分析南非,如果不理解种族问题的历史脉络和现状,阶级分析就会抓不到重点。可能分析半天小资产阶级,但小资产阶级在南非根本不是主要矛盾——南非的社会矛盾围绕种族问题展开,需要用"新形态殖民主义"框架来理解。

类似的,分析印度必须理解种姓制度,分析中东必须理解伊斯兰教影响。

流程:

1. **AI做初步调研:** 用标准化指标采集信息,包括该国所有左翼组织和思想家的分析文章,形成conjunctural overview
2. **人介入形成框架:** 读材料,形成对该国矛盾的初步感觉,与专家交流,确定用什么理论框架指导分析
3. **AI在框架下深化:** 把理论框架导入MAS,后续深度分析在这个框架下进行(此时任务降解为Complicated)

关键点：理论框架是涌现的，不是预设的。必须先理解具体情境，才能选择框架。AI提供材料，但不能替代人的洞察。

| 真实案例2:News on China新闻选择

制作"News on China"每周新闻摘要，流程分两段：

第一段:新闻选择(必须人工介入)

- 从本周所有新闻中选出10–12条进入期刊
- 典型的编辑判断场景：
 - 几条新闻其实是同一件事(都是四中全会,不同角度),AI识别不出
 - 把相关但不完全相同的新闻组合起来,讲更完整的故事
 - 判断哪类新闻更重要,如何平衡正面/负面新闻比例
 - 这期期刊的整体节奏和叙事逻辑

这些判断,AI做不好,必须编辑动脑子。

第二段:深度调研和摘要写作(可以自动化)

- 针对选定的新闻线索,从7个角度全方位调研:事件全貌、媒体报道、专家评论、历史背景、专业知识、地缘政治、影响分析
- 根据调研结果,撰写80字以内的精炼摘要
- 编排成邮件期刊格式

两段完全解耦：第二段不关心第一段新闻如何选出,只要编辑给出新闻线索列表即可。

| 人机协作模式

AI做overview,人提供框架或判断,AI在框架下深化分析。

人机接口设计要清晰:

- 上游AI产出什么格式的材料,便于人阅读和判断?
- 人如何将洞察/框架注入系统?(文档?结构化输入?)
- 下游AI如何理解和应用人提供的框架?

| 设计要点

- **明确的人工介入节点：**在流程哪个环节需要人,输入输出格式是什么
- **上下文传递机制：**人的洞察如何传递给下游AI
- **解耦设计：**上下游环节独立,人工介入不破坏整体流程
- **支持域降解：**人的框架注入后,下游任务降解为Complicated,可自动化

Complex域的核心挑战是精心设计人机接口,明确"AI能做什么"和"人必须做什么"的边界。

Chaotic域:AI作为辅助工具

任务特征

所需信息不存在于可访问的数据源,必须通过人的现场判断和创造性工作获得。行动先于理解:必须先做,才能知道。

典型场景: 田野调研、原创理论构建、创新探索。

判断标准

- **信息基本不存在:** 这是Chaotic域最关键的特征
- **无法通过搜索或分析已有材料完成:** 必须通过行动创造新数据
- **因果关系在系统层面不可辨识:** 高度不确定,需要在行动中感知

与Complex的区别:

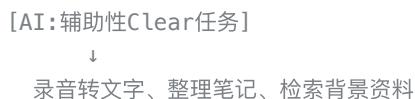
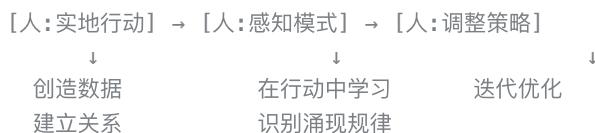
- Complex:信息存在(如左翼组织的分析文章),但需要人类洞察识别模式
- Chaotic:信息不存在(如村超的流量转化机制,无公开材料),必须通过行动创造

这个区分对业务架构设计至关重要:决定任务是否可以通过网络搜索+AI分析完成(Complex),还是必须实地调研(Chaotic)。

业务架构



人主导的行动循环:



架构特征:

- AI不是主体,只是辅助工具
- 核心的信息采集和理论构建都由人完成
- AI只能做Clear域的辅助任务

真实案例:贵州榕江村超调研

研究榕江村超,数字技术如何助力乡村振兴。要把研究做扎实,需要掌握很多具体情况:

- 每月实际流量多少?

- 流量从哪些媒体/自媒体来?
- 县里如何运作庞大的自媒体矩阵?
- 流量如何转化为乡镇经济?
- 转化形式有哪些?效果如何?
- 各乡镇村寨做哪些工作来创造和转化流量?

核心问题：这些信息不存在于互联网上,没有档案记录,AI没地方获取。必须去基层调研,跟实际干事的人交流,在这个过程中碰撞打磨出理论框架。

更重要的是,事先不知道该问什么问题。只有到了现场,理解了实际运作方式,才能提出有理论意义的问题。

人机协作模式

人是主体,AI是工具：

- 整理访谈录音(语音转文字)
- 整理田野笔记
- 检索背景资料
- 生成初步的材料分类

这些都是Clear域的任务,AI可以做。但核心的田野观察、理论构建、洞察形成,必须由人完成。

设计要点

接受AI的局限性。不要试图设计"自动化田野调研系统"。明确AI能做的辅助性工作,把它们做好就足够了。

Chaotic域提醒我们:识别AI的边界,和识别AI的能力一样重要。

如何判断任务属于哪个域?

三问决策树



问题1:完成任务所需的信息是否已经存在?

|— 基本不存在 → Chaotic域(需要创造性行动生成新数据)

| |例:榕江村超流量转化机制(需实地调研)

|— 已存在(可通过搜索、文献、公开材料获取) → 继续判断

| |问题2:从信息到结论的因果关系/分析路径是否明确?

| |— 明确(知道"怎么分析") → 继续判断

| | |

| | |问题3:是否需要专家级判断和解读?

- 不需要 → Clear域(标准流程可处理)
 - 例：下载指定URL文件、转录录音
- 需要 → Complicated域(需要分析但路径清晰)
 - 例：应用DSI指标评估数字主权
- 不明确(不知道“该用什么视角分析”) → Complex域(需要探索)
 - 例：为南非构建阶级分析框架、选择周刊的10条关键新闻

典型特征对照表

域	信息存在性	处理复杂度	框架确定性	AI能力	人的角色
Clear	明确存在	简单直接	无需框架	全自动执行	定义任务，验收结果
Complicated	明确存在	需要分析	框架可预设	专家系统	定义规则和框架
Complex	明确存在	需要洞察	框架涌现	辅助材料准备	提供洞察和框架
Chaotic	不存在/不可访问	高度不确定	框架待构建	仅辅助Clear任务	主导全流程

设计原则总结

Clear域:监控和容错

- 流水线设计:单向流程,步骤清晰
- 异常处理:网络中断、数据损坏、权限问题
- 进度可见:实时监控,判断是否正常运行
- 幕等性:支持断点续传,避免重复工作

Complicated域:并行化和规则引擎

- 任务分解:识别可独立处理的子任务
- 并行执行:充分利用多智能体并行能力
- 规则显式化:分析规则清晰建模,便于调试优化
- 中间结果验证:每个环节产出可追溯、可验证

Complex域:人机接口设计

- 清晰的介入节点:明确在流程哪个环节需要人工判断
- 接口标准化:AI产出什么格式材料给人?人如何将洞察注入系统?
- 解耦设计:上下游环节独立,人工介入不破坏整体流程
- 上下文传递:人的框架和洞察如何传递给下游AI

Chaotic域:接受局限

- 人是主体:不要试图自动化核心流程
- AI做辅助:识别可以辅助的Clear域任务(整理、检索、转录)
- 工具化定位:把AI当作增强人能力的工具,而不是替代人

核心洞察

不要试图把所有任务都自动化。识别任务的性质,选择合适的MAS构型,明确人的角色,这才是好的系统设计。

强行自动化Complex或Chaotic域的任务,结果往往是产出大量肤浅、缺乏洞察的内容。准确识别任务的域,设计合适的人机协作方式,才能发挥AI和人各自不可替代的优势。

域降解:系统的演化视角

最后一个重要的洞察:**研究过程往往是域降解的过程**。

研究通常从Chaotic或Complex开始:

- 第一次研究新问题,不知道该用什么理论框架(Complex)
- 甚至不知道信息在哪里,该问什么问题(Chaotic)

但随着研究深入:

- 形成了理论框架,下次研究类似问题时,任务从Complex降解为Complicated
- 摸清了数据源和分析规则,任务进一步降解,部分环节甚至成为Clear

演化路径:



Chaotic → Complex → Complicated → Clear
(创造数据) → (探索框架) → (结构化分析) → (标准化执行)

MAS设计的演化启示

系统要支持演化。一个好的研究用MAS应该:

- 在Complex阶段,提供灵活的人工介入机制
- 当框架明确后,支持将人的洞察固化为规则,降解为Complicated域
- 当规则稳定后,支持进一步自动化,降解为Clear域

这样,系统随着研究者的理解深入而逐步自动化,而不是一开始就强行自动化所有环节。

实际案例：南非阶级分析的演化

第一次分析南非：

- Chaotic：不知道该问什么，需要广泛阅读建立感觉
- Complex：确定用“新形态殖民主义”框架
- Complicated：在框架下深度分析

第二次分析类似国家（如津巴布韦）：

- 跳过Chaotic：已经知道该问什么类型的问题
- 可能仍需Complex：需要判断同一框架是否适用，或需要调整
- 更快进入Complicated：框架确定后，大量分析工作可自动化

多次迭代后，甚至可能：

- 形成一套适用于南部非洲国家的标准分析框架
- 整个流程降解为Complicated，可以快速批量分析

结语

Cynefin框架为MAS设计提供了一个系统化的思考工具：

- **诊断任务性质**：这个任务属于哪个域？
- **选择构型**：应该采用什么样的系统架构？
- **明确边界**：哪些部分让AI做，哪些部分必须人来做？
- **规划演化**：如何随着理解深入逐步自动化？

下次设计MAS时，先判断任务的域，再选择合适的构型。不要让AI做它做不好的事，也不要让人做AI能做好的事。

这才是真正的人机协作。



熊节

Like the Author

