

我们该如何理解AI?

# AI是一种整体吗?

也许，它并非由零件拼凑而成?

Hassani 2026 | arXiv:2606.20704

## 一元论

AI不是由独立部件组成的机器，而是一种在层层耦合、动态交互中涌现的整体。它的性质来自整体的组织与过程，而非单个元素。



## 不可拆分

拆解AI会破坏其关键的相互作用与上下文网络，导致整体能力的丧失。理解AI，不能只看零件，而要看关系与结构。



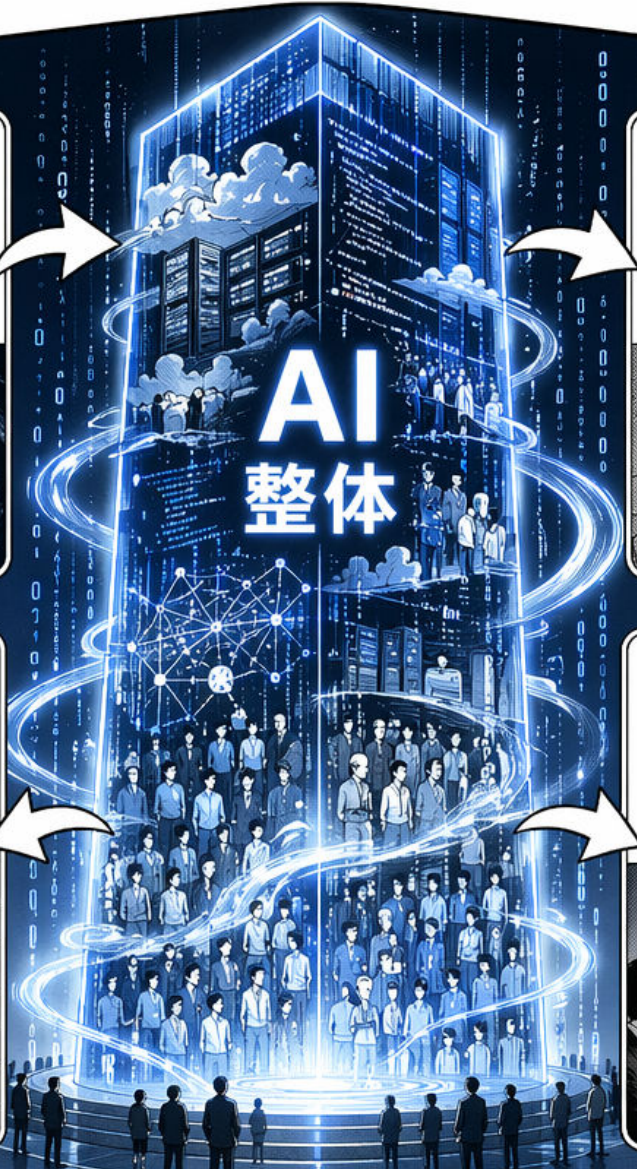
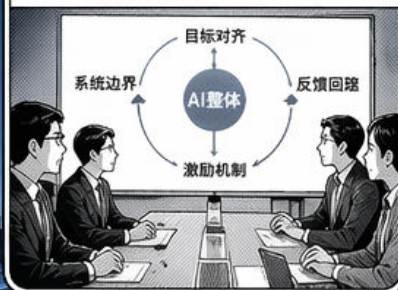
## 反还原

不能用还原论方法（如只看参数、模块或数据）来理解或评估AI。整体大于部分之和，行为来源于整体的涌现动力学。



## 组织治理

把AI当作整体，需要新的治理思路：关注系统边界、反馈回路、目标对齐、激励机制与社会影响，而不仅是技术指标。



## 风险提醒



将AI视为可控工具，可能低估涌现风险。



忽视整体性，容易导致评估失真与治理失效。



边界模糊的AI系统，可能带来不可预见影响。



唯有整体视角，才能促进稳健与负责任的AI。



问题变了，答案也会不同!

# 把AI看成整体，先改变提问方式

从整体出发，才能走得更远!



# 论文在问什么？

不同的人看到AI的不同部分，  
却都指向同一个影子。

建筑师

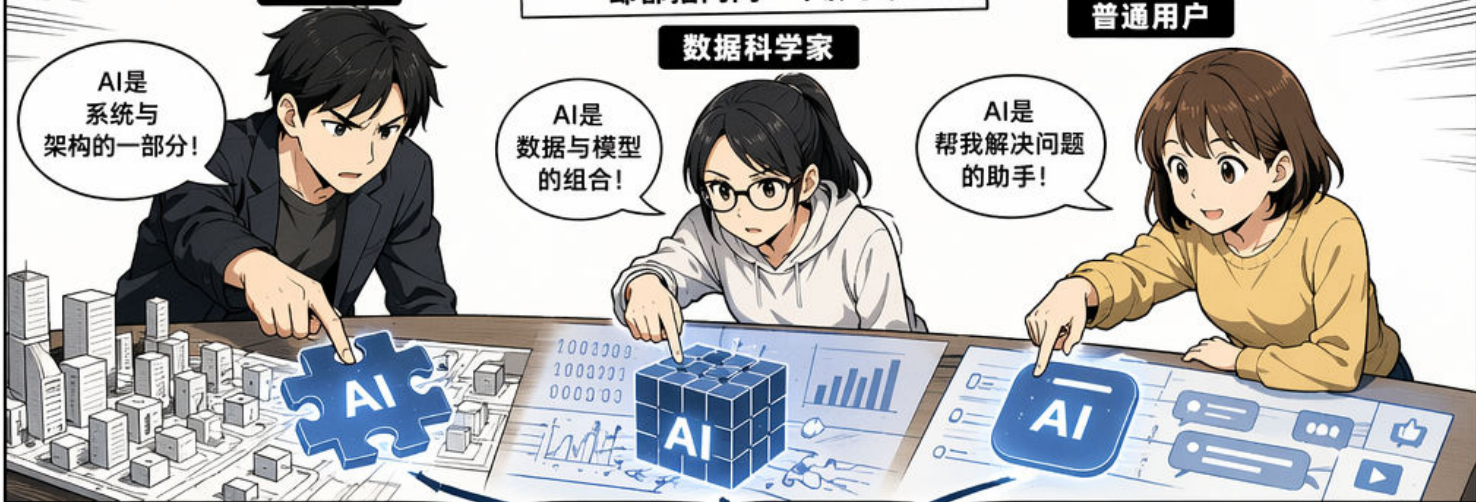
AI是  
系统与  
架构的一部分！

数据科学家

AI是  
数据与模型  
的组合！

普通用户

AI是  
帮我解决问  
题的助手！



## 1 局部定义

各自看到AI的  
一小块，  
并给出定义。



系统、模块、接口、  
基础设施...



数据、算法、模型、  
训练、评估...



聊天、推荐、搜索、  
生成、辅助...

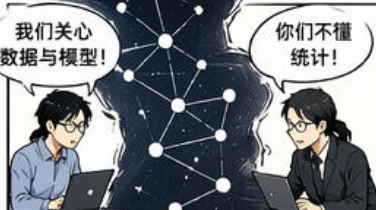
## 2 学科分裂

学科视角不同，  
语言不同，  
彼此难以对话。



我们关心  
结构与空间！

你们只看  
技术实现！



我们关心  
数据与模型！

你们不懂  
统计！



我们只关心  
好不好用！

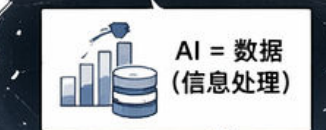
你们太  
复杂了！

## 3 还原论假设

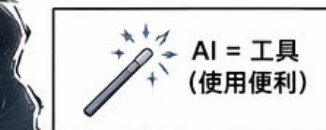
倾向于把AI还原为  
单一要素：  
技术、数据或工具。



AI = 技术  
(工具理性)



AI = 数据  
(信息处理)



AI = 工具  
(使用便利)

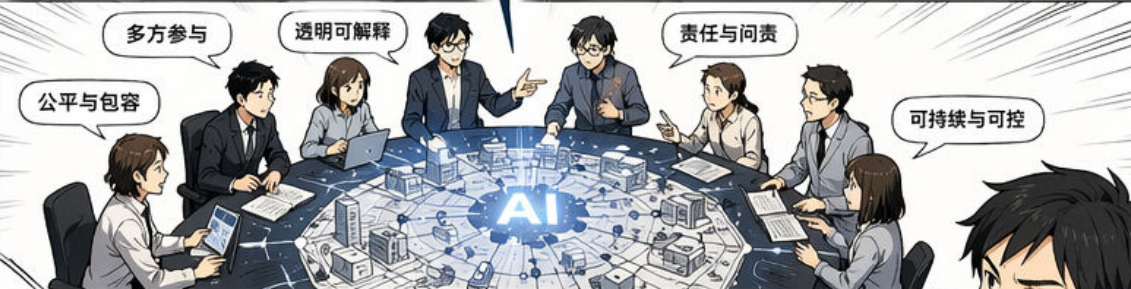
## 4 整体重读

跳出片段，看到  
AI背后的整体：  
知识、权力、价值、  
实践的网络。



## 5 治理重构

从整体出发，重构  
AI的治理与设计：  
跨学科、跨层级、  
面向公共利益。



问题不是AI像什么工具，  
而是AI如何组织知识与行动

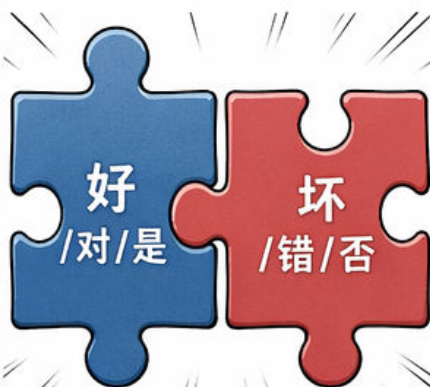


# 三种哲学框架

看待世界的三种不同角度，  
决定了我们如何理解问题与行动！

## 二元论：拆成两类

把世界拆分为对立的  
两类，非此即彼。



### 特点：

界限清晰，判断直接，  
但容易忽略中间地带。

### 例子：

“他要么支持，要么反对。”



简单明确，  
但可能太片面了...

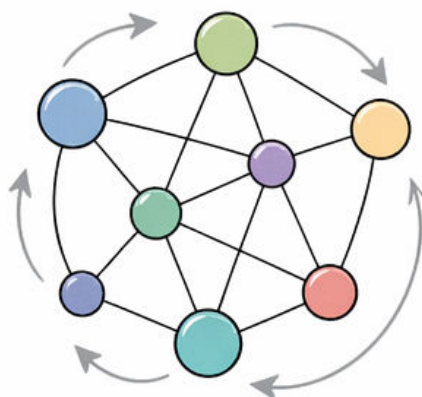
### ⚠️ 责任断裂

问题被推给“另一类”，  
容易逃避自身责任。



## 整体论：关注关系

不只看单独的事物，  
而是看它们之间的关系与  
相互影响。



### 特点：

强调整体、结构与动态，  
看到“牵一发而动全身”。

### 例子：

“气候变化影响经济、健康、  
政治，彼此交织。”



关系复杂，  
但更接近现实！

### ⚠️ 责任断裂

容易把责任稀释在系统中，  
难以追究具体行动者。



## 一元论：同一现象

认为所有事物本质上是  
同一个整体或同一实体的  
不同表现。



### 特点：

追求根本统一，超越差异，  
但可能忽视现象层面的多样。

### 例子：

“万物皆一，差异只是表象。”



从终极视角看，  
一切本为一体。

### ⚠️ 责任断裂

差异被视为幻象，  
现实中的不公可能被忽视。



要点  
总结！

整体论说相互影响，  
一元论说不可还原



# AI架构不能分开看

从数据到业务，形成一个端到端的智能系统

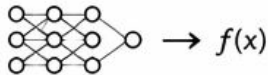
## 1 数据表示

原始数据经过清洗、结构化、特征工程，转化为模型可理解的表现形式。



## 2 模型逻辑

模型接收数据表示，通过算法与训练，学习规律并产生推理与决策能力。



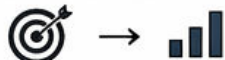
## 3 算力架构

计算资源、存储、网络与调度系统，为模型训练与推理提供稳定、高效的基础设施。



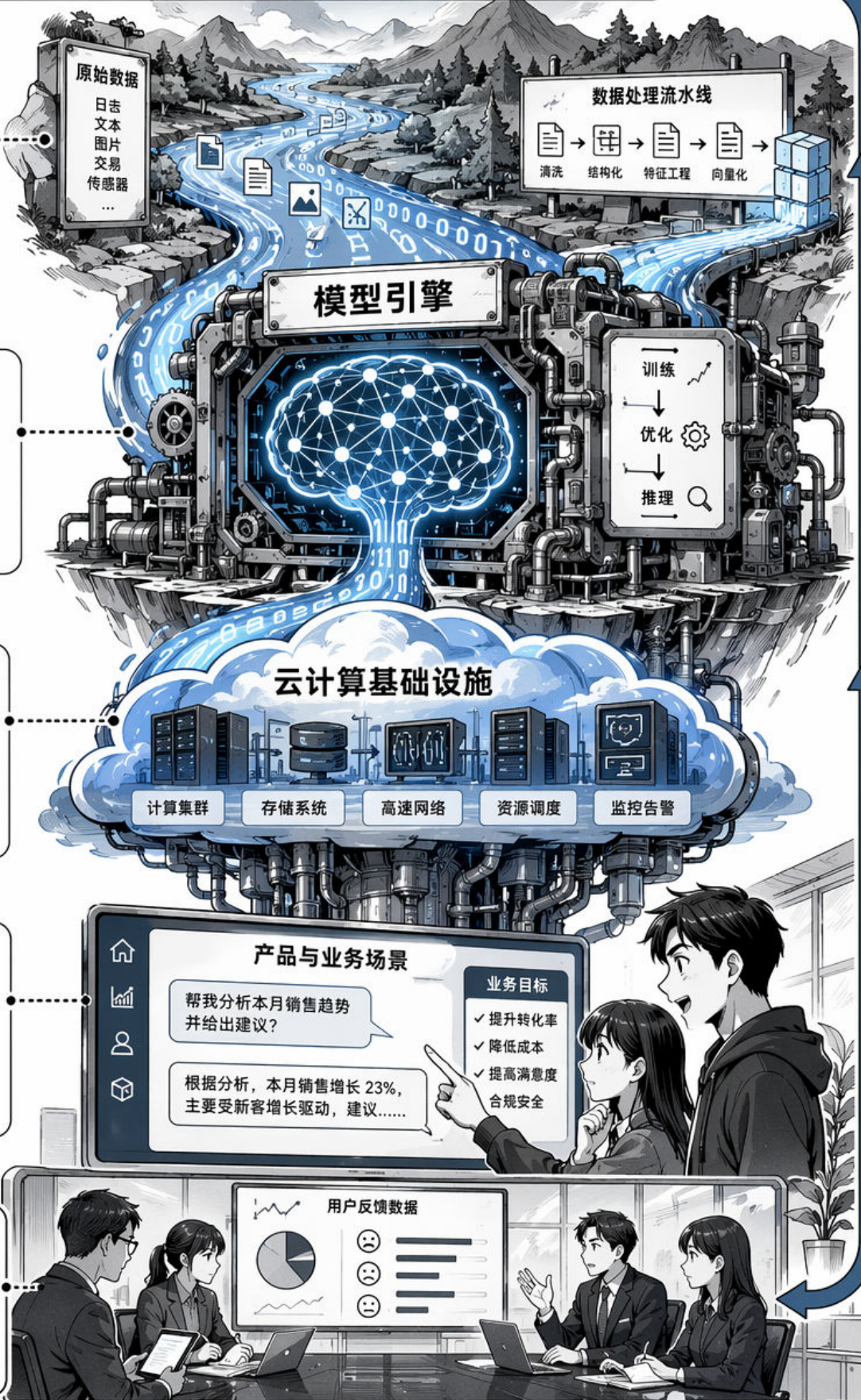
## 4 业务目标

模型能力要解决真实业务问题，对齐目标、指标与场景，创造可衡量的价值。



## 5 用户反馈

用户使用产生反馈数据，推动迭代优化，形成持续进化的闭环系统。



# 一个问题贯穿所有部件

只有协同，才能发挥整体智能的最大价值！

图片由AI生成

# 作者如何论证？

哲学  
追问本质  
澄清概念  
指导实践



## 1

### 哲学定义

澄清核心概念，界定研究对象与关键范畴。

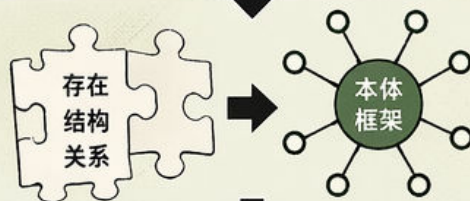


回答“是什么”：  
通过概念分析与哲学论证，  
建立清晰的理论起点。

## 2

### 本体论转向

从存在论视角转向，重构理解框架与关系结构。

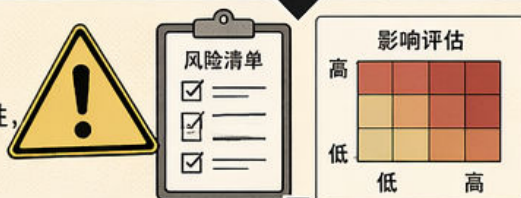


回答“存在如何”：  
以本体论建模揭示实体、属性  
与关系的深层结构。

## 3

### 风险分析

识别潜在风险与不确定性，  
评估影响与后果。



回答“可能出什么问题”：  
系统识别风险源，评估  
概率与影响，形成风险图谱。

## 4

### 组织设计

设计治理组织架构、职责  
分工与协同机制。



回答“谁来做、如何做”：  
构建组织体系与协同机制，  
明确职责与流程。

## 5

### 数据治理

建立数据治理体系与规范，  
保障数据质量与合规。



回答“如何落地、如何持续”：  
通过制度、技术与流程，  
实现数据的可信、可用、可控。

论文结论

## 这是一篇概念到治理的框架论文

# 核心论点链条

从碎片化到整合，再到制衡

还原论思维：  
把复杂系统拆成部件



碎片化导致：  
信息割裂、理解偏差、  
决策失真、风险放大！

## 1 现象不可拆

现实世界是整体的、  
动态的、相互关联的，  
无法被真正拆开。



## 2 AI不只是部件

AI是一个整体系统，  
能力来自各部分协同  
与涌现，不是零件堆叠。



## 3 筒仓会误读

部门/模块各自为政，  
视角受限，数据孤岛，  
导致误解与误判。



## 4 横向整合

打通数据、知识、流程  
与视角，形成端到端的  
导通与闭环。

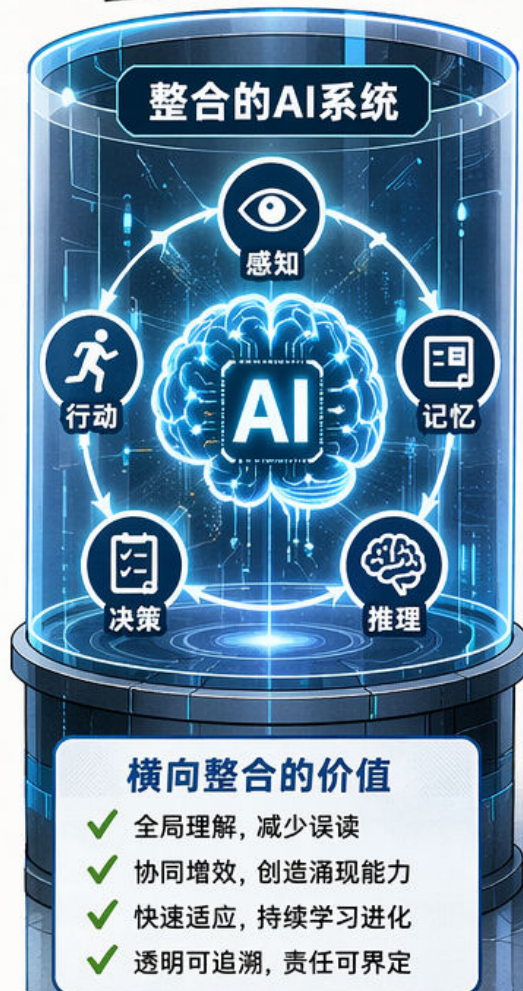


## 5 制衡权力

建立分布式治理与监督  
机制，防止权力集中，  
保障多元与公平。



整体视角：  
一个整合的AI系统



警惕中心化风险！

- ✗ 权力过度集中
- ✗ 缺乏监督制衡
- ✗ 单点失控风险大
- ✗ 压制多元与创新



一元论修正碎片化，  
但要防止中心化



# 反驳与回应

关于“一元论”的三组核心争议

反驳  
(质疑)

回应  
(解释)

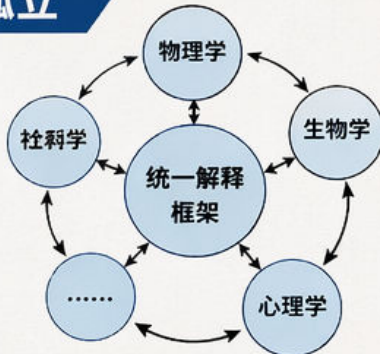
## 1 组件可替换?

现实世界由许多可替换的组件组成,不需要假设一个统一基础。



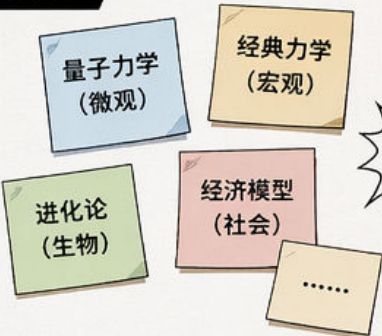
## 解释上不可孤立

这些领域在解释上相互依赖、相互约束。脱离整体,难以形成自治的理解。



## 2 多模型并存?

不同模型在不同领域都很成功,没必要追求一个最终统一的模型。



## 问题层面统一

模型是工具,解决的是不同层面的同一类问题:“世界如何运作、为何如此?”统一问题需要统一视角。



## 3 权力集中?

一元论若被少数人掌握,可能导致思想垄断与权力集中,风险巨大。



## 需要显性制衡

一元论更需要透明、开放、可质疑的制度设计,防止滥用,保障多元参与与持续纠错。



## 一元论可能有用,也可能过强

关键不在于非此即彼,而在于在统一与多元之间,找到最有解释力、最负责任、最可持续的平衡。



保持开放的心智,理性地统一;建立制度的防线,理性地制衡。

# 最终脑图：AI一元论

## 1 本体论



AI是什么？

- 人工系统，不是自然存在
- 由数据、算法与算力构成
- 具有目标导向与行为能力

## 2 认识论



我们如何认识AI？

- 通过模型与数据认知世界
- 存在不确定性与偏差
- 可解释性与可验证性是关键

## 3 组织设计



谁来构建与运营AI？

- 跨学科协作与角色分工
- 治理结构与决策机制
- 权责清晰与制衡设计

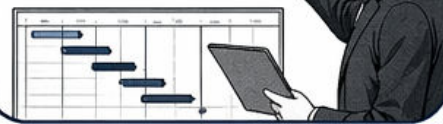


## 4 项目管理



如何让AI项目成功落地？

- 需求定义与范围管理
- 迭代开发与风险控制
- 质量保障与交付价值



# AI整体

统一的技术系统  
连贯的价值链条  
共同的社会影响

## 5 数据治理



数据如何支撑AI？

- 数据质量与标注管理
- 数据安全与隐私保护
- 数据全生命周期治理



## 6 伦理风险



AI可能带来哪些风险？

- 偏见与歧视
- 隐私侵犯与监控滥用
- 失控与滥用风险



## 7 公共责任



AI对社会与未来意味着什么？

- 公平、透明与可问责
- 包容、普惠与可持续
- 共同参与、共治共享



## 责任链条：从工程师到管理者再到监管者

### 工程师（创造者）



我定义  
技术实现与  
模型行为

责任：  
技术正确性与安全性

### 管理者（决策者）



我决定  
目标、资源与  
应用场景

责任：  
业务合理性与组织治理

### 监管者（守护者）



我制定规则  
并监督  
系统影响

监管  
指南

责任：  
公共利益与社会福祉

# 谁定义问题，谁就影响AI的现实解释权