



清华大学 计算机科学与技术系  
Department of Computer Science and Technology, Tsinghua University

# 迈向通用人工智能

清华大学计算机系、人工智能研究院 张 钊



# 一、在干什么

## 分布式语义 (word embeddings)

- 语义定义: “You shall know a word by the company it keeps.”
  - [J.R. Firth](#) (1957)
- 从共现中学习意义
- 把稀疏的高维共现信息变成致密向量空间中的几何结构 → 语义关系

## 二、要干什么

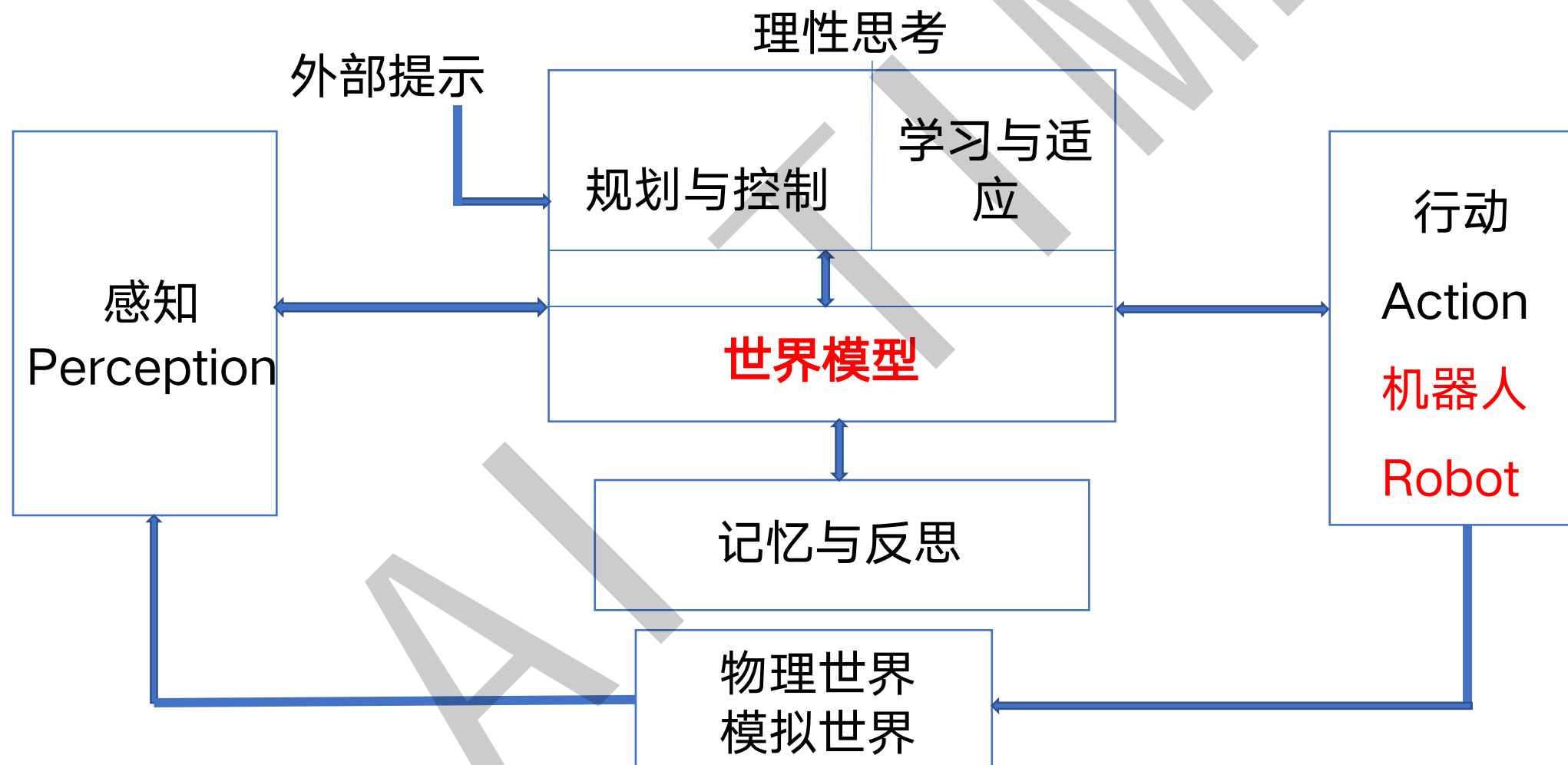
从LLM 到在现实环境下（数字、物理）执行复杂任务的Agent

AGI Next: 可执行与可验证的定义

- （1）时空一致的多模态理解与落地，
- （2）可控的在线学习与适应
- （3）可验证的推理与长期规划与执行，
- （4）可校准的反思与元认知
- （5）跨任务强泛化

并在上述维度上达到预设的人类基准阈值与可靠性约束。

# 从LLM到智能体（Agent）的跃迁



### 三、怎么干

多模态落地， 具身与交互落地， 检索与证据落地，  
结构化知识对齐， 工具与执行落地， 对齐与约束落地

### 四、愿景

从工具（Tool）到主体（Subject）

- （1）功能-行动主体（agency subject）
- （2）规范-责任主体（moral/practical subject）
- （3）体验-意识主体（phenomenal subject）

## 五、AI时代的企业家

重新定义价值创造，  
建立可信与可治理的 智能系统，  
推动产业协同与生态共赢，

组织与释放新生产力，  
构建长期韧性，  
实现普惠与可持续增长



谢谢!