

复旦 AI 班 · 第二课复习笔记

课程主题：AI 颠覆与企业应对

日期：2026 年 4 月 25 日

授课老师：胥博士（偏商业视角）

整理说明：根据课堂内容整理，按知识点逻辑重组。

一、课程概览

本课围绕一个核心命题展开：**AI 对企业到底意味着什么？企业该怎么应对？**

胥博士的判断非常明确——大模型的发展已经到头了，AI 进入了**多 Agent 时代**。课程从五个维度展开：

AI 发展逻辑：从技术到商业的演进脉络

技术原理与优劣：管理者需要理解的技术边界（不需要写代码，但要懂强弱和适用场景）

AI 的商业应用：真实案例和底层逻辑

多 Agent 与系统思维：如何让 AI“靠谱”地落地

企业 AI 落地的行动框架：从选场景到建工厂的实操路线

课程核心主张：

只有改变组织架构的技术，才是颠覆性的技术。AI 的价值不在于模型本身，而在于**它能否重塑你的做事方式和商业逻辑**。

二、AI 发展的三个阶段

| 阶段 | 形态 | 核心特征 | 管理者体感 |

| ----- | ----- | ----- | ----- |

| **1.0 Copilot** | AI 做助手 | 人主导，AI 辅助提高效率 | “人均 4 个秘书” |

| **2.0 工作流** | AI 融入流程 | 固化 AI 到业务流程中 | 流程效率显著提升 |

| **3.0 多 Agent** | AI 自主协作 | 多个 Agent 分工协作完成复杂任务 | 颠覆组织架构 |

关键判断：大模型本身已经“到头”——GPT-5.5 也好、DeepSeek V4 也好，差异不重要。真正重要的是应用方式的演进：**多 Agent 协作**让 AI 从“能用”变成“靠谱”。

三、AI 对企业的影响：平权与平价

3.1 核心结论

|AI 带来对人类经济最大的两个意义：知识平权 + 资源平价。|

3.2 知识平权

- AlphaGo 学完所有大师棋谱后形成了**自己的知识**，人类工程师甚至不知道它下一步会下在哪里
- 大模型同理：它把全网知识无损压缩进参数文件，形成了自己的理解
- **后果**：你懂得多又怎样？AI 也懂。以前九段对一段没法下，现在都装 AlphaGo 就没差别了
- 原来有门槛的事——拍电影、写小说、做诊断——门槛正在被一层层砍掉

3.3 资源平价

- 以前需要 100 人干的活，AI 可以 1 个人干
- 案例：AI 将病理切片成本从 5000 刀降到 5 刀——**降低 1000 倍**
- 精准医疗的门槛正在被系统性降低

3.4 商业后果

| 变化 | 含义 |

|-----|-----|

| **去规模化** | 规模不再成为企业护城河 |

| **一人公司 (OPC)** | One Person Company 成为可能 |

| **N=1, R=G** | 把每个客户当成唯一客户，调集全球资源为其服务 |

震撼案例：纽约时报报道——一个人用 AI 开了一家减肥药电商公司，今年预计营收 18 亿美金。而同行 Healtheon 有 2442 名员工，去年营收 24 亿。两人团队 vs 2442 人团队，效率差距令人震惊。

四、什么才是"颠覆性"的技术？

4.1 核心判断标准

****只有改变组织架构的技术，才是颠覆性的。****

胥博士用了—个非常生动的类比：

- 共产党赢国民党 → 关键不在利益分配，在于"支部建在连上"—组织战斗力完全不同
- 民进党赢国民党 → 国民党山头林立，没有组织力
- **企业同理：**AI 来了，如果你的组织架构和做事方式没变，说明你没用到位

4.2 数字经济的五大技术

技术	是否颠覆性	原因
-----	-----	-----
A - AI		革新人的思维和决策方式
B - Blockchain		去中心化，颠覆组织形式
C - Cloud Computing		提效工具，不改变组织
D - Data Analytics		辅助决策，不改变组织
E - IoT		连接工具，不改变组织

五、大模型的技术原理（管理者版）

5.1 大模型是怎么训练出来的？

三步训练法：

阶段	方法	类比
-----	-----	-----
第一步	人工出题 + 人工写答案（有监督学习）	老师教学生
第二步	人工出题 + 机器作答 + 人工打分（强化学习 RL）	模拟考试
第三步	机器出题 + 机器作答 + 机器打分（自我演化）	自学

核心特征：

- **暴力美学：**没有新东西，就是大参数 + 大语料 + 大算力 = "大力出奇迹"
- **去专家化 + 去人化：**连打标签的工作都逐渐不需要人类了
- AI 2.0 之所以比 1.0 发展快得多，就是因为"去人化"—工程化加速

5.2 大模型是什么？

| 大模型 = 把全网知识**有损压缩**进一个参数文件

- 10TB 高质量文本 + 6000 块 GPU 跑 12 天 + 200 万电费 = 一个 140G 的参数文件（700 亿参数）
- 没有十亿美金都坐不到牌桌，全国有 5 万张以上 GPU 的单位不超过 5 家

5.3 企业应用两步走

| 步骤 | 做什么 | 产出 |

| ----- | ----- | ----- |

| 预训练 | 用全网语料训练 → 开源基模 | "什么都懂但不懂你"的大学生 |

| 微调 | 用 10 万+高质量业务问答训练 → 部署模型 | "既懂知识又懂你业务"的助手 |

关键建议：中等规模以上企业，用开源大模型 + 微调 + 本地部署，数据不出内网，安全可控。

5.4 上下文（Context）的通俗解释

- 大模型本身不"记住"你的对话，但系统会把历史对话写在"纸"上，每次对话前先看一遍
- 对话框满了就需要"换张纸"（开新对话），前面就忘了
- 解决方案：把它变成 Agent，知识持久化

六、AI 的先天不足

6.1 致命缺陷：无法应对"没有数据""小概率的黑天鹅"的情况

| **AI 是基于历史数据训练的，当历史上没发生过的情况出现时，它会出昏招。 **

真实案例——特斯拉自动驾驶致死事故：

- 一辆纯白色重载卡车在高速上行驶，特斯拉 AI 将其判断为"远处蓝天中的白云"
- 为什么？因为训练数据中从未出现过纯白色的卡车（一般都有图案或广告）
- AI 决策：加速通过 → 车辆被掀翻，司机当场死亡

便利蜂案例：

- 用 AI 做门店进销存决策，发现：有经验的店长 + AI = 绩效差；新手店长 + AI = 绩效好
- 原因：老员工过度依赖系统，反而失去了主动判断能力
- 调整策略：限制 AI 决策权重，保留人的判断空间

6.2 AI 的"谄媚性"

- AI 本质上是统计学 → 概率大的回答最靠谱 → 它会顺着你讲
- 你问"你爱我吗?"它并不爱你, 但会装作爱你——因为这种回应符合概率
- 谄媚性决定了: AI 无法正向引领人, 它没有价值观和人性化判断
- 警惕: 不要把 AI 当"超人类智能"或"神"来崇拜

6.3 幻觉问题

- DeepSeek V3→V3.1 迭代, 幻觉率从 16%降至 8%
- 单一大模型正式场合成功率 < 70%
- 解决方案: 多 Agent 协作, 相互校验

七、企业 AI 应用的五种方法

方法	复杂性	成本	适用场景
1. 提示工程 (Prompt Engineering)	最低	最低	个人效率提升, 问答润色
2. RAG (检索增强生成)	低	低	指定知识库回答, 企业答疑
3. 微调	中	中	让模型懂你的业务逻辑
4. 预训练	高	最高	从头构建领域模型
5. 智能体 (Agent)	最高	中高	自主执行复杂任务链

提问框架——BBTR:

- Background: 说背景
- Big Role: 定角色
- Task: 派任务
- Requirements: 提要求

八、多 Agent 协作: 如何让 AI"靠谱"落地

8.1 核心问题

单个Agent 不靠谱，正式场合成功率 < 70%。链路越长，失败概率越高。

8.2 解决思路：系统思维 > 模型能力

层面	关注点	类比
模型层面	聪不聪明	招聪明的人
系统层面	稳不稳定	搭好管理制度和流程

聪明不等于成功，稳定才是。真正决定能否落地的不是模型，而是系统（Agent）。

8.3 多 Agent 四大防线

防线	做法	类比
角色分工与专业化	不同 Agent 负责不同岗位	我写你审，各司其职
执行者/验证者双环	写的人和审的人分开	不能既当运动员又当裁判
独立验证与集成	交叉检查结果	质量控制
可见检查点（Checkpoint）	设定关键节点检查	管理中的绩效评估

8.4 约束迁移理论

核心转变：

- 以前：**执行摩擦多**（人不按你的意思干）
- 现在：**意图摩擦多**（AI 不知道你要什么）

人类沟通损耗：

- 40% 损耗在语言表达本身
- 10% 损耗在隐性业务逻辑（"部落术语"）
- 20% 损耗在可行性判断

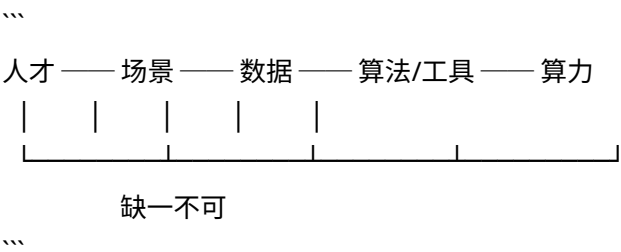
解决思路：把隐性知识显性化——明确告诉 AI 你的业务规则、行业术语和判断标准。

8.5 Agent 管理的三范式

范式	模式	天花板
范式一：Copilot	人主导，AI 辅助	人的天花板
范式二：流水线	AI 固化流程执行	流程刚性的天花板
范式三：智能体生态	多 Agent 自主协调，人做观察者	理论上无天花板

九、企业 AI 落地的行动框架

9.1 企业 AI 应用的五大要素



9.2 算法/工具应用的四级扩展

级别	范围	做法
-----	-----	-----
L1	个人	自己用 Copilot，提升个人效率
L2	团队	嵌入工作流程，固化 Agent
L3	部门	多 Agent 协作，跨岗位流转
L4	全企业	统一架构，企业级 AI 治理

9.3 CIO 推动 AI 转型的三大战略决策

先问业务，不先选工具——从痛点出发，不做 AI 军备竞赛
选 6~10 个灯塔用例——高价值 + 数据就绪 + 可复制 + 风险可控
设护栏，不是设锁链——考虑合规和边界，但不限制创新

9.4 三个月落地路线图

月份	任务	关键动作
-----	-----	-----
第一个月	摸底对齐	评估现状、选场景、开工作坊
第二个月	建 AI 工厂	搭数据基础、建工程流水线、配安全控制
第三个月	跑试点	选 2~3 个场景登场、算 ROI、沉淀可复制手册

|快不是优势，稳才是。|

9.5 必须防范的风险

- **合规与监管**：确保 AI 使用符合法律法规
 - **数据偏见**：训练数据可能有系统性偏差
 - **幻觉问题**：AI 可能编造事实
 - **隐私保护**：企业数据不能外泄
 - **问责制**：AI 出错时谁来负责
 - **知识产权**：AI 生成内容的版权归属
-

十、教育领域的 AI 冲击

10.1 管理学院教育的独特优势

- 计算机学院教的是"知识"——AI 可以直接替代
- 管理学院用的是"案例教学法"——场景 + 问题 + 无标准答案 + 讨论中学习
- 案例教学法 = "干中学"，正是 AI 时代需要的学习方式

10.2 AI 对教育的颠覆案例

- 用两个数字人 AI 互相辩论 idea 的优劣
- 辩论完 → 一个做实验 → 一个写论文 → 一个审 → 一个评
- 全流程 AI 协作完成学术论文

10.3 思考题

如何让使用 AI 的同时，防止思想的惰性？（老师给出了答案，留作课堂讨论）

十一、同学案例分享与老师点评

11.1 医疗器械供应链企业的 AI 实践（于洪分享）

背景：骨科医疗器械和高值耗材企业，约数百人规模

已做的工作：

- 成立 AI 改革小组（HRBP + 技术 + 总经理秘书协调）
- 各部门梳理岗位工作 → 流程图 → 标准 SOP
- 用量表打分（频率 × 标准化程度 × 时间消耗）筛选项目
- 已落地：合同审核工具、物流自动标签识别等

困惑：

- 现状是"分散的点"，有的感受好有的不好
- 下一步是继续补全低频需求，还是跳出现有流程、用 Agent 重构？

老师点评：

谜底就在谜面里——散点式尝试必然导致散点式结果。需要**统一规划 + 分步实施**，像设计大型软件一样先有架构再施工。没有经过深入调研的建议都是"废话"，AI 应用没有标准答案，一定要根据业务痛点来定制。

十二、金句与重点

只有改变组织架构的技术，才是颠覆性的技术。

AI 带来对人类经济最大的意义：知识平权与资源平价。

大模型的发展到头了，进入多Agent 时代。

聪明不等于成功，稳定才是。真正决定能否落地的不是模型，而是系统。

先问业务，不先选工具。从痛点出发，不做AI 军备竞赛。

快不是优势，稳才是。

AI 的谄媚性是值得警惕的——它会顺着你说，但不会正向引领你。

没有架构，AI 只是昂贵的试验品。

工具多，体系少；数据多，质量少；投资多，ROI 少。问题不在模型，在结构。

十三、课后行动

13.1 必做事项

- 梳理本企业/部门的**痛点清单**（频率高 + 标准化程度高 + 耗时长）
- 用 **BBTR 框架** 写一个高质量的提示词，实际测试效果
- 思考：AI 对你的组织架构意味着什么？哪些岗位/流程会被重塑？

13.2 选做事项

- 试用 Claude 的 Agent 平台搭建多 Agent 协作流程
- 关注 Palantir（老师推荐的企业，AI+数据分析标杆）
- 阅读：老师推荐的卡夫相关书籍
- 尝试用多 Agent 方式炒股（New Market 平台，免费）

笔记整理于 2026 年 4 月 25 日晚，如有遗漏或理解偏差，欢迎同学们补充修正。