

FSE598 前沿计算技术

模块 1 计算思维 单元 1 计算机系统设计 第 1 讲 前沿技术简介

本讲座的英文版内容基于教材：

Y. Chen, G. De Luca Service-Oriented Computing and System Integration: Software, IoT, Big Data, and AI as Services, 8th edition, Kendall Hunt Publishing, 2022.

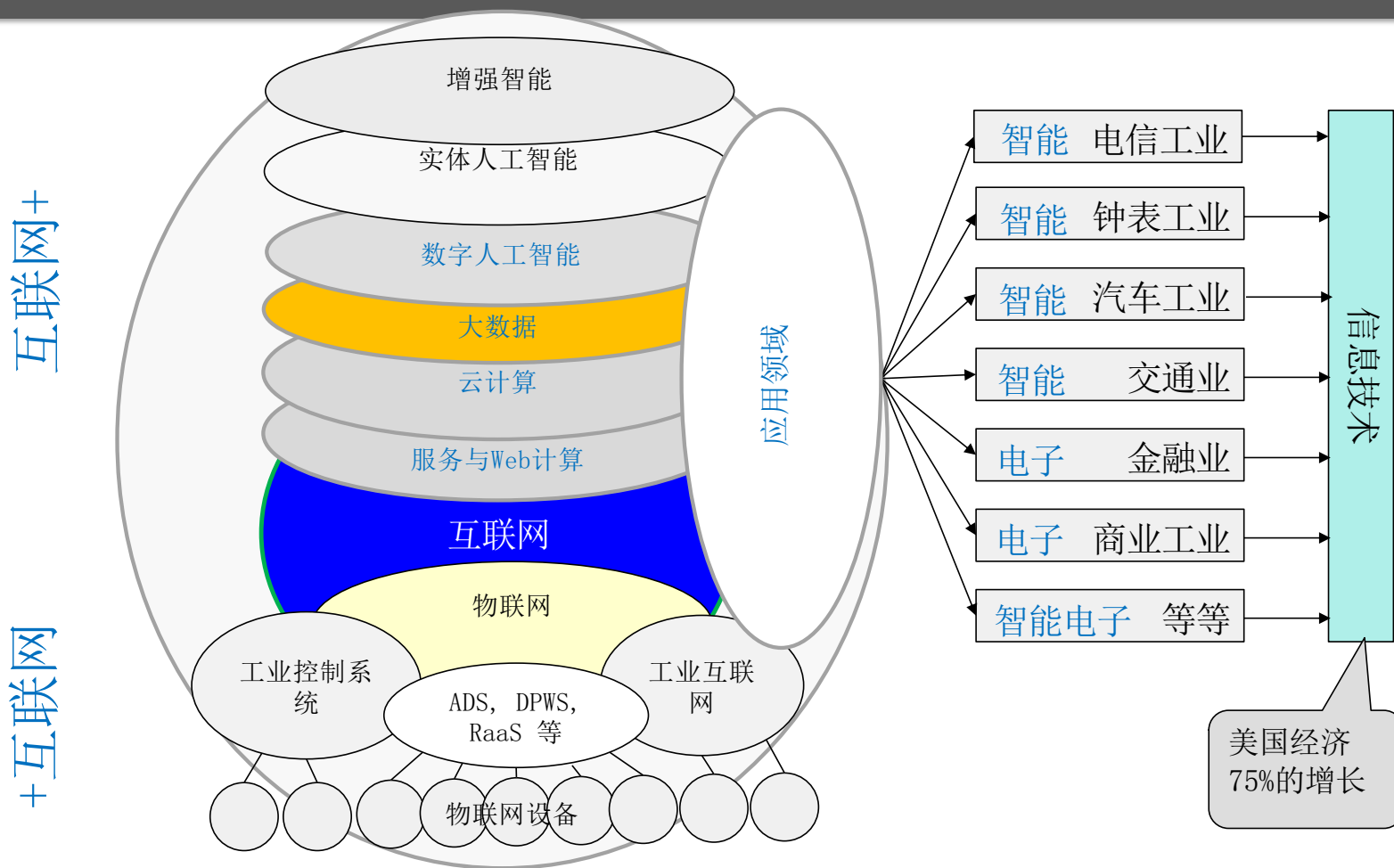
<https://www.public.asu.edu/~ychen10/book/socsi.html>

第 1 讲大纲

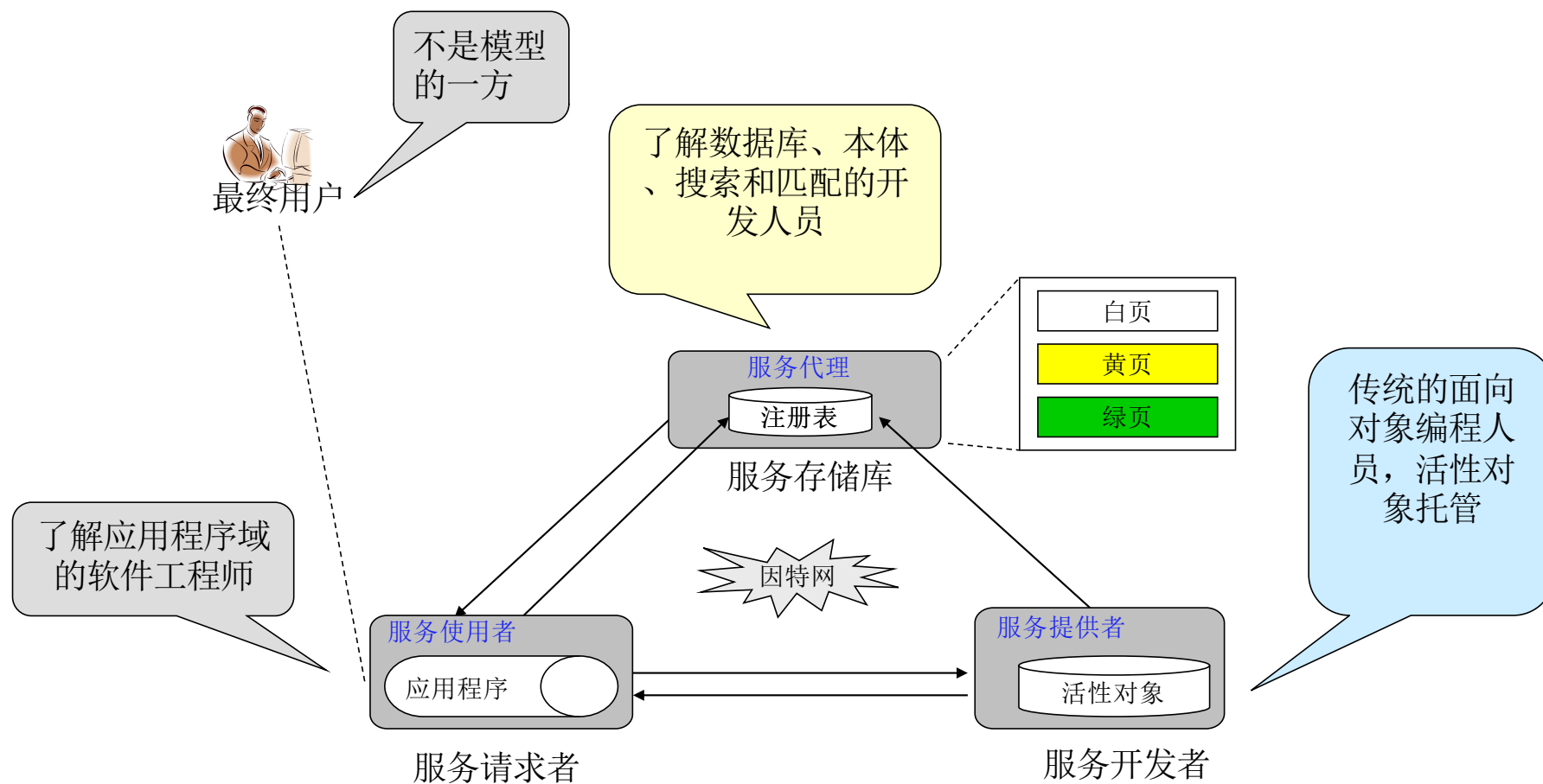
学习

- 前沿计算和技术
- 大数据的特征和价值
- 从大数据概念到研究和学习的领域
- 从大数据到智能决策

大数据和人工智能领域的前沿计算和技术



服务与Web计算：面向服务计算的三方模型



Web 应用示例：动态商业逻辑是必需品



最终用户

旅行社图形用户界面

输入出行日期
出发城市和
目的地

输入付款信息
显示预订确认

旅行社商业逻辑

验证数据
调用旅行 web 服务
为客户推荐航班
调用银行服务
签发预订确认

旅行社
预订机票
预订酒店
预订租车
预订火车票

不同的银行
服务

可编辑文本文件中的 Web 服务列表

构建代理和/或调用

构建代理和/或调用

构建代理和/或调用

<http://www.usair.com/webservice>
<http://www.swair.com/webservice>
<http://www.unitedair.com/webservice>
<http://www.usdeltaair.com/webservice>
<http://www.lufthansa.de/webservice>
<http://www.airfrance.com/webservice>
<http://www.hilton.com/webservice>
<http://www.doubletree.com/webservice>
<http://www.daysin.com/webservice>
<http://www.fourseason.com/webservice>
<http://www.hyatt.com/webservice>
<http://www.holidayinn.com/webservice>
<http://www.avis.com/webservice>
<http://www.hertz.com/webservice>
<http://www.enterprise.com/webservice>
<http://www.thrifty.com/webservice>
<http://www.budget.com/webservice>

服务计算与云计算-- 房屋与公寓的类比



个人设施

- 空调，供电
- 供水、下水道，
- 安全/安保……

可扩展性和隐私

- 无法创造更多空间
- 良好的隐私



共用设施

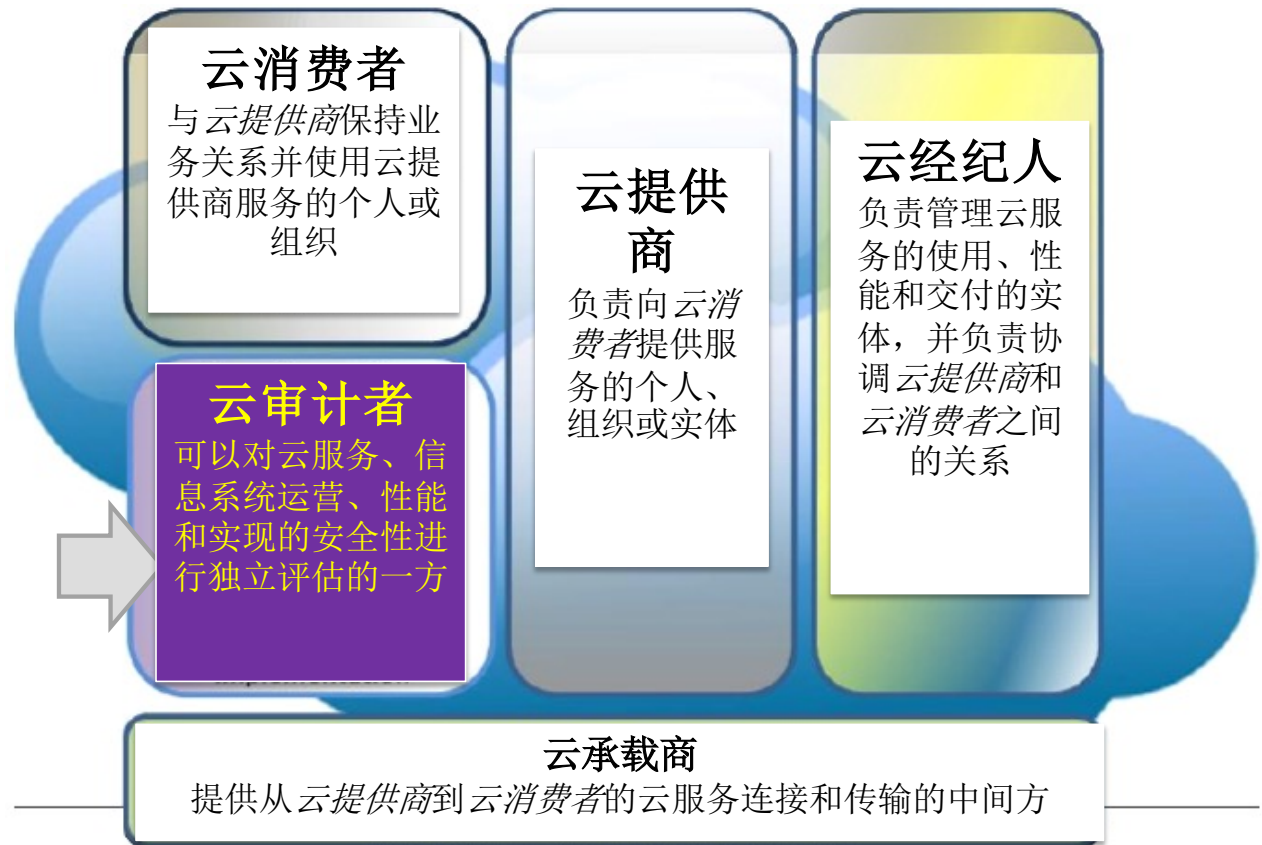
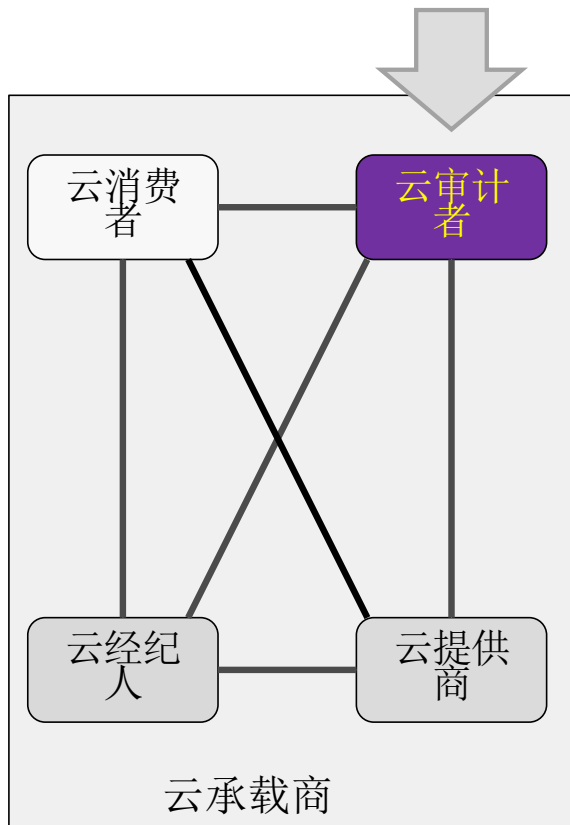
- 空调，供电
- 供水、下水道
- 安全/安保……

可扩展性和隐私

- 可以要求更多空间
- 需要对隐私加以改善

NIST 云计算四方模型

http://www.nist.gov/manuscript-publication-search.cfm?pub_id=915112



云计算从何而来？

- ❑ 面向服务的计算与基于 Web 的计算
- ❑ 虚拟化
- ❑ 数据库、数据中心和数据仓库
- ❑ 超级计算系统（向上扩展）
- ❑ 并行、分布式、网格计算（向外扩展）
- ❑ 我们是先有云计算还是先有大数据？

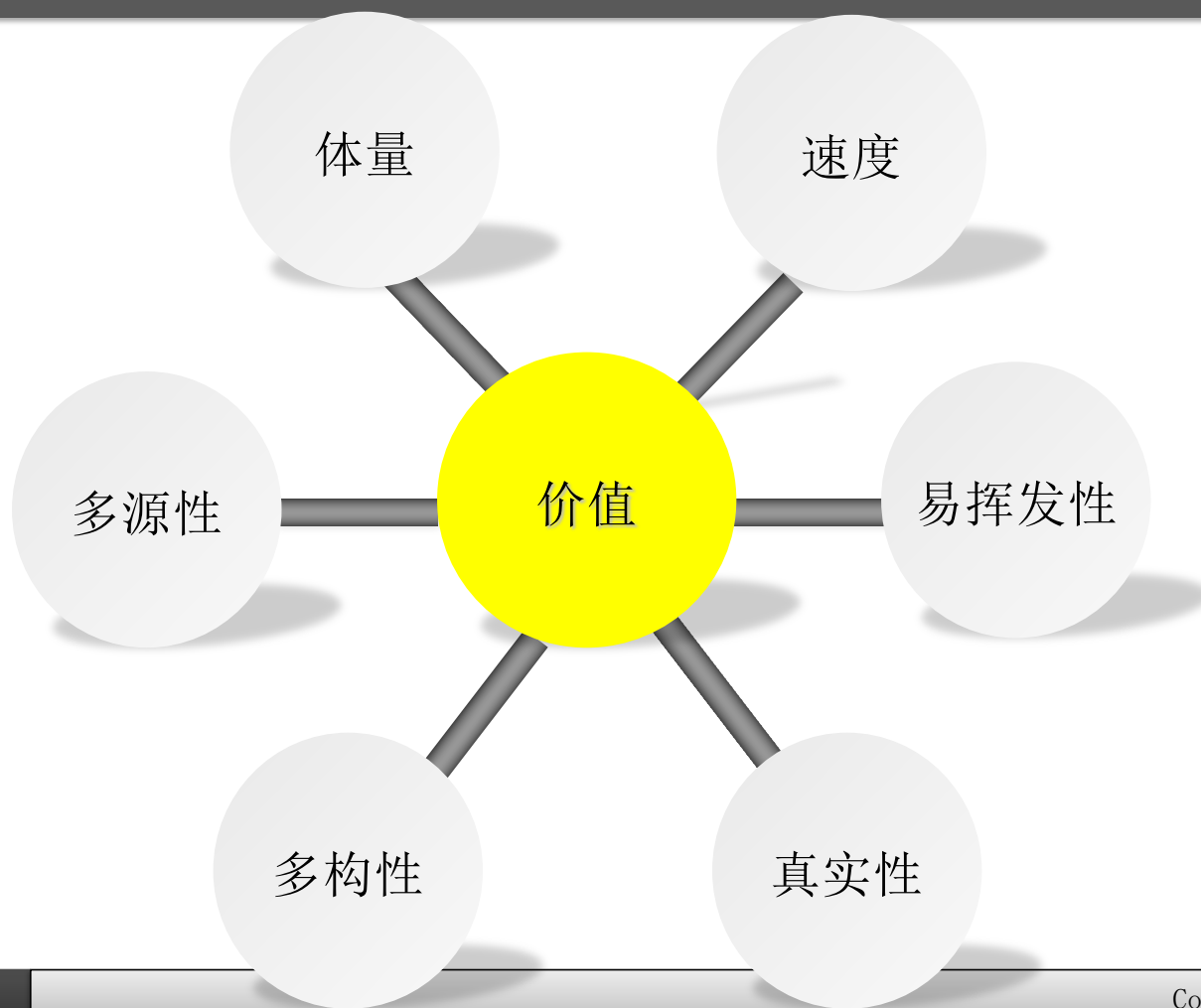


云计算



大数据

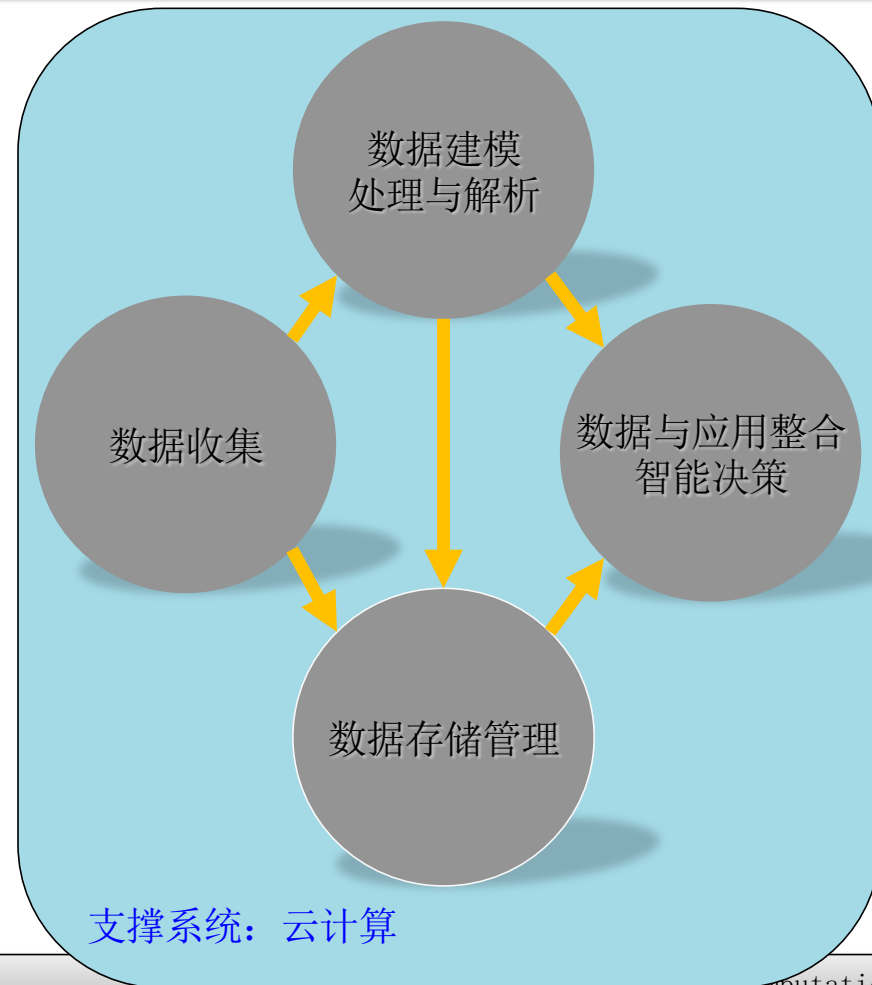
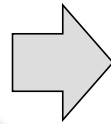
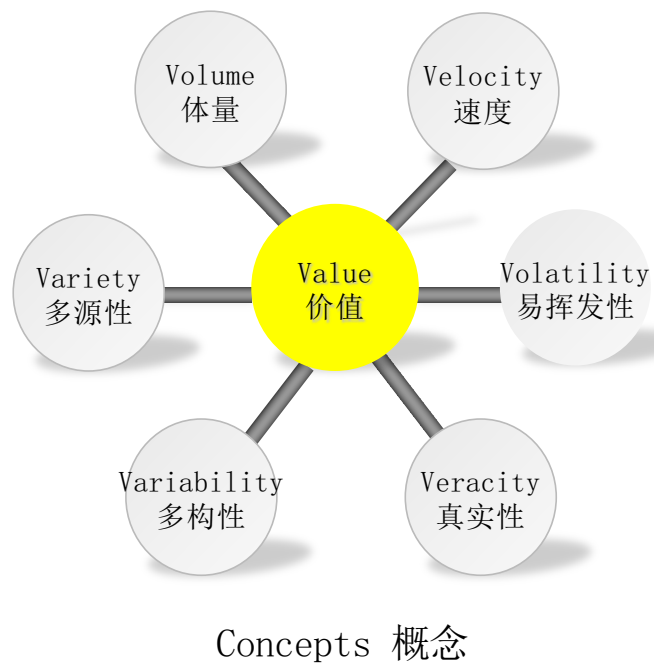
大数据的特征和价值



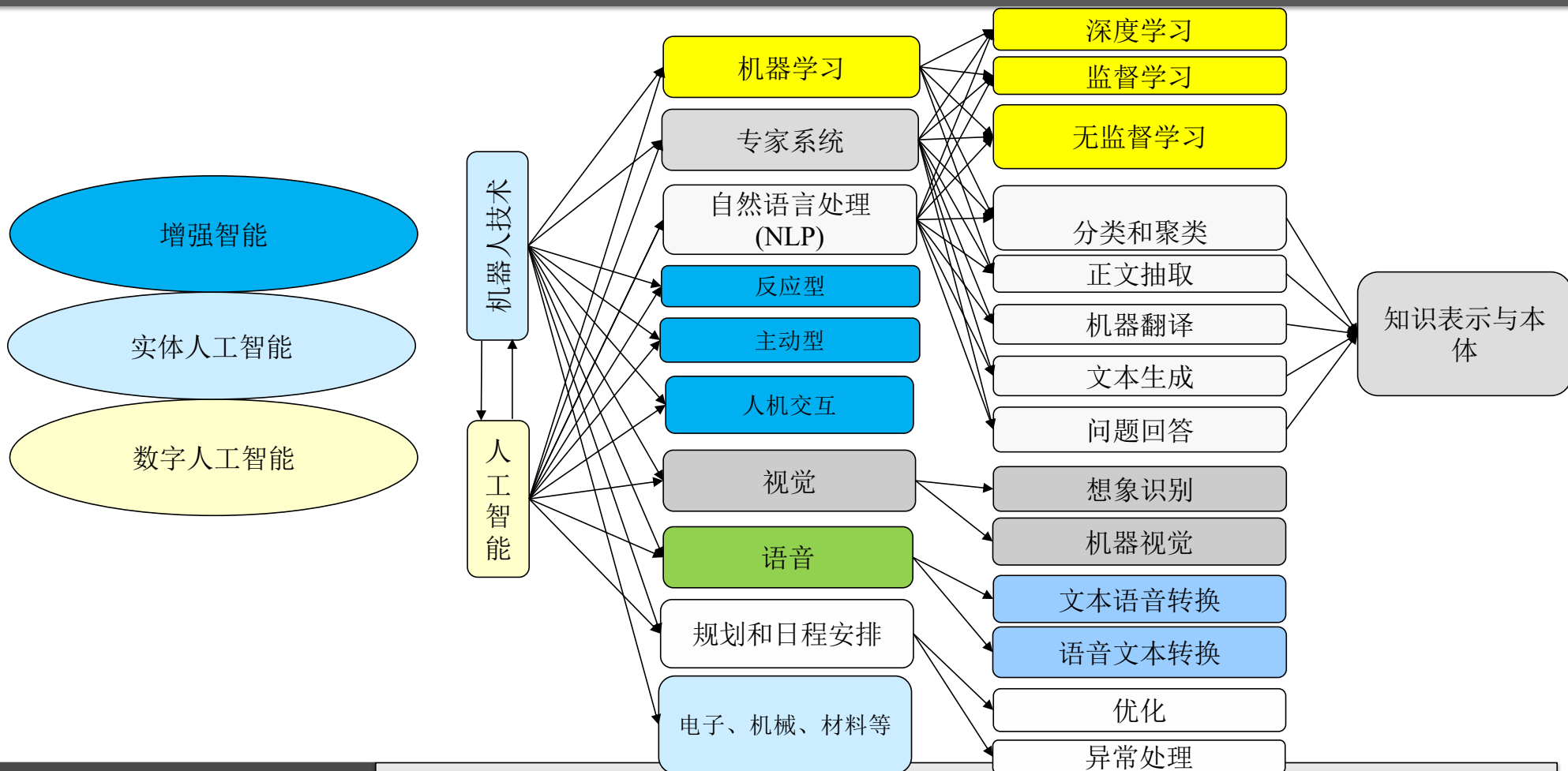
大数据系统的各个方面：几个 V

- ❑ **价值**：大数据是继互联网（**通信**）和云计算（**计算**）之后的下一件大事。大数据事关**数据**。
- ❑ **体量**：数据量从拍字节（ 10^{15} 字节）、艾字节（ 10^{18} ）到泽字节（ 10^{21} ），越来越多.....
- ❑ **速度**：实时数据要求实时响应。
- ❑ **多样性**：来自不同**来源**、具有不同**语义**的数据被集成到不同的应用程序中。
- ❑ 数据结构的**多变性**：多结构数据
- ❑ **真实性**：准确性问题：噪声消除和容错
- ❑ **易挥发性**：并非所有数据都会被存储，有些数据会被永久删除，因此需要大数据处理系统选择性地存储和组织数据，以最大限度地发挥其价值。

从大数据概念到研究和学习领域



当前的人工智能(AI)分支



机器学习算法

□ 监督学习算法

- 分类：已知的类属性包括离散值，将新数据分配到预定义的类中
 - 决策树
 - K-NN: k -最近的邻居
 - 神经网络和深度学习
 - 应用：检测数据库操作通常找不到的隐藏模式
 - 垃圾邮件检测
 - 网络攻击检测
- 回归：已知的类属性是实数
 - 预测给定数据实例的实数值
 - 根据过去的历史数据预测给定月份的降雨量
 - 根据周边已售房屋数据预测给定房屋的价格

□ 无监督学习算法

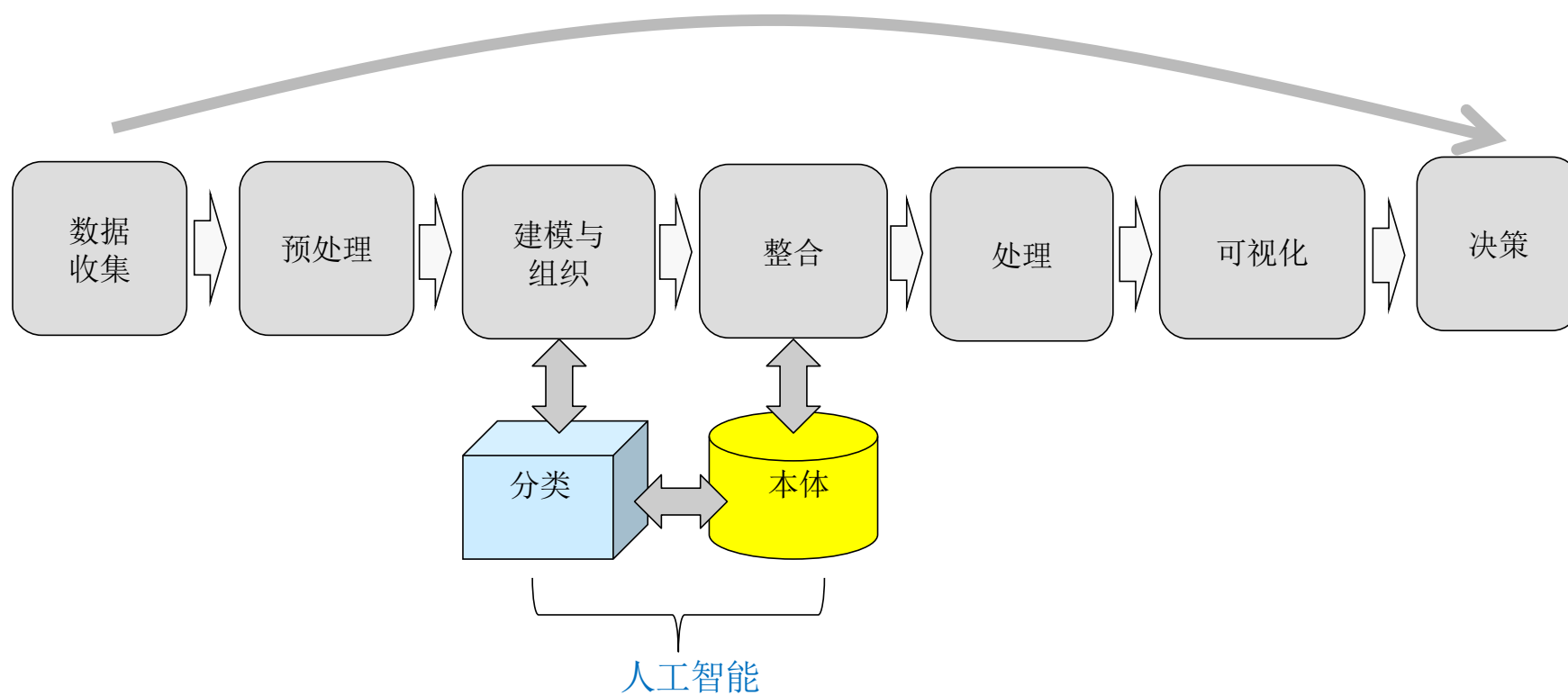
- 根据项目的相似性，例如它们之间的语义距离/关系将项目组合成一些集群
- 无培训数据集
 - 检测给定社交网络中的社区

□ 强化学习：

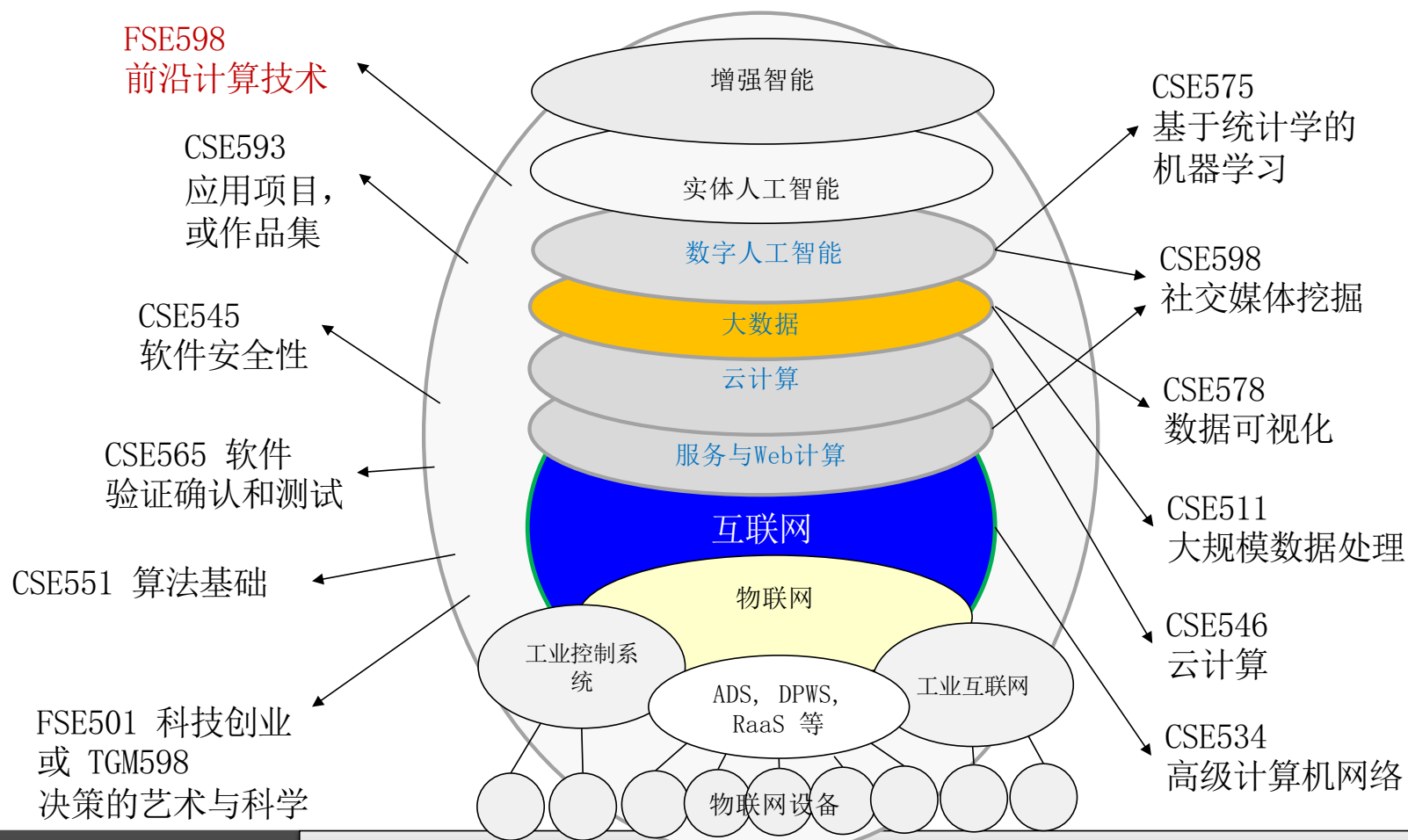
- 获取反馈并动态增强模型

从大数据到智能决策

大数据人工智能开发范例可自动执行步骤



大数据和人工智能领域的前沿计算和技术的学位



课程讨论：计算与软件设计与分析（1）

❑ FSE598 前沿计算技术

- 前沿计算技术
- 计算思维
- 为后续课程作准备

❑ CSE551 算法基础：有效算法的编写与算法复杂性分析，包括，

- 算法设计与分析技术
- 贪婪算法，分治算法，动态编程
- 图形算法与网络流, 均摊分析, NP-完全问题

❑ CSE565 软件验证、确认和测试

- 软件的测试方法
- 软件的正确性与可靠性
- 测试管理与测试过程改进

❑ CSE545 软件安全性：软件安全的理论和工具，包括

- 安全设计
- 威胁分析和建模
- 安全测试和编码

❑ FSE501 技术创业, 或 TGM598 决策的艺术与科学

- 用来满足我们计划里必修的科技与创新课的要求。

课程讨论：大数据与人工智能（2）

- ❑ CSE546 云计算：
 - 虚拟化与云计算，
 - 可编程网络，性能评估，信息保证，
 - 分布式和并行计算以及基于云计算的应用程序
- ❑ CSE511 大规模数据处理：学习生成，处理和管理大规模数据集的新框架。涵盖以下主题：
 - 核心数据库概念，
 - 分布式和并行数据系统，
 - 在云中部署和操作数据系统，
 - NoSQL数据库系统以及大数据工具。
- ❑ CSE534 高级计算机网络：先进的网络协议，高性能网络在分布式系统中的应用。涵盖的主要主题包括
 - 互联网和多媒体领域，机器学习在网络中应用
 - 数据中心网络，软件定义网络，网络的特殊功能。

课程讨论：大数据与人工智能（3）

- ❑ CSE598 社交网络挖掘基础：
 - 网络度量，社交网络分析，网络模型，随机图，
 - 小世界模型，幂定律，数据挖掘，服务计算，Web与移动计算
 - 信息传播，流行病，行为分析，社交媒体中的推荐，社区检测。
- ❑ CSE575 基于统计学的机器学习：研究支持人工智能的数据挖掘和统计模式识别。涵盖的主要主题包括
 - 监督学习；无监督学习；以及深度学习，
 - 包括机器学习的主要组成部分以及支持它的数据分析。
- ❑ CSE578 数据可视化：大数据的兴起有可能为决策提供依据，而视觉表示可以在我们的日常信息消费中发挥至关重要的中间作用。 涵盖
 - 基于图形设计与视觉艺术，
 - 感知心理学和认知科学等原理创建有效可视化的技术和算法，
 - 增强对复杂数据的理解。
- ❑ 毕业设计体验。
 - CSE593 应用项目：在一门课中完成一个大项目。CSE593将计入10门课之一。
 - 或者 作品集（Portfolio） 在完成的10门课中选两门课的课程项目组成一个作品集。作品集不计入10课程中