

北京城市实验室 ( B C L ) 2 0 1 8 年 会  
DATA + FUTURE : 新数据与未来城市

# 人工智能城市的门槛与途径

李栋

2018年06月



# 几个关键问题

- Q1 人工智能是什么
- Q2 人工智能的能和不能
- Q3 城市能否运行于人工智能之上
  - Q3-1 城市规划与人工智能
- Q4 人工智能的未来走向
- Q5 提高城市智能的几种方式

# Q1 人工智能是什么

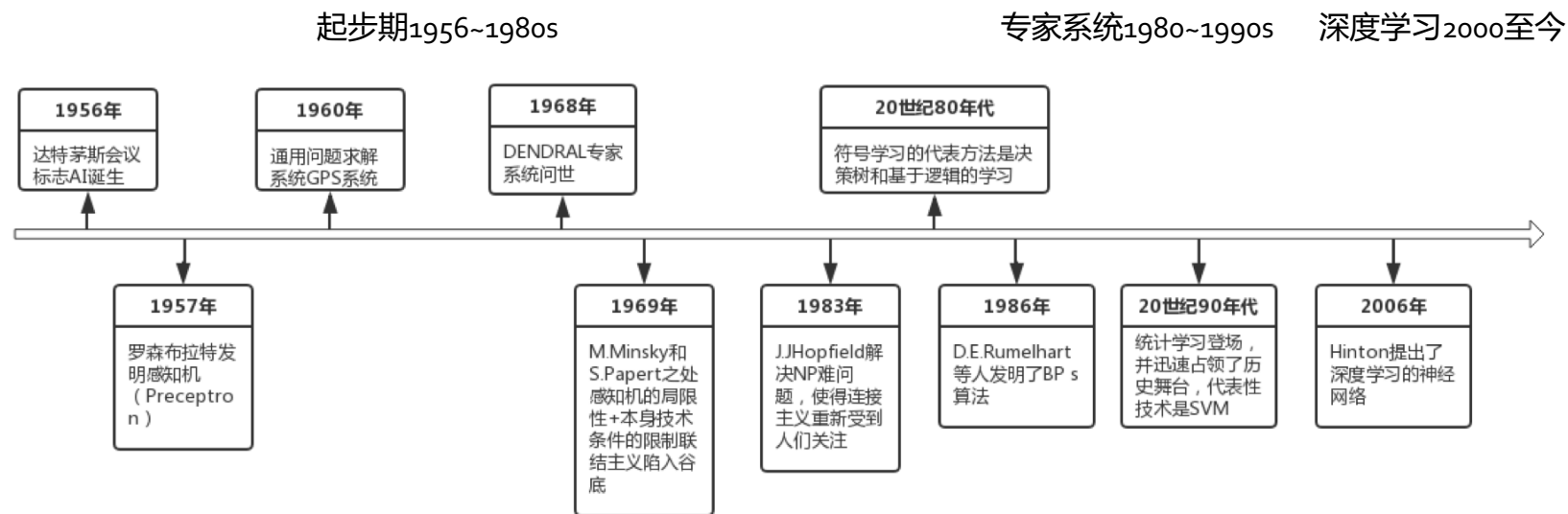
- 人工智能的目标：
  - 通过智能的机器，延伸和增强人类在改造自然、治理社会各项任务中的**能力和效率**，最终实现一个人与机器和谐共生共存的社会
  - 智能机器：可以是一个虚拟的或者物理的机器人（软件或设备）
  - 智能机器有自主的感知、认知、决策、学习、执行和社会协作能力，符合人类情感、伦理与道德观念。

# Q1 人工智能是什么

- 人工智能的多种划分，可被大致归纳为：
  - 像人一样思考、像人一样行动、理性的思考、理性的行动.....
  - 弱人工智能、强人工智能（通用人工智能）
  - 历史脉络：符号主义、专家系统、机器学习与深度学习相继登场
  - 应用分支：计算机视觉、自然语言、认知推理、机器人、伦理等

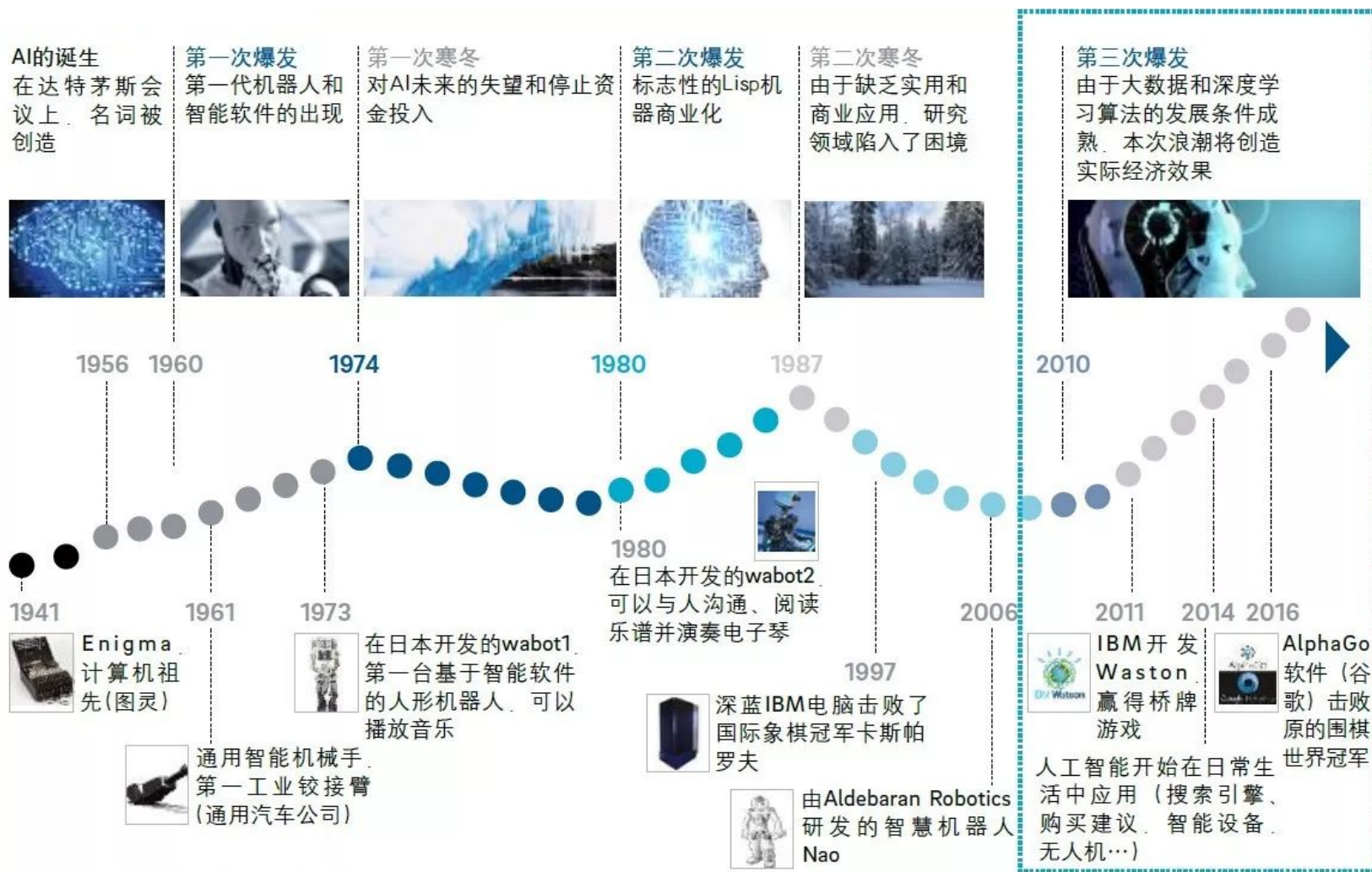
# Q1 人工智能是什么

- 人工智能发展脉络中的起起落落：
  - 前30年以数理逻辑的表达与推理为主、后30年以概率统计的建模、学习和计算为主
  - 由于历史发展的原因，传统人工智能领域研究者多数已经过世或退休，自1980s以来基本抛弃了前30年以逻辑推理与启发式搜索为主的研究方法，取而代之的是概率统计建模、学习的方法



# Q1 人工智能是什么

## 人工智能发展脉络中的起起落落：



资料来源：案头研究：中国人工智能学会；罗兰贝格分析

# Q1 人工智能是什么

- 当前人工智能涉及领域广泛、分散：
  - 计算机视觉（含模式识别，图像处理等议题）
  - 自然语言理解与交流（含语音识别、合成、对话等）
  - **认知与推理（含各种物理和社会常识等议题）**
    - 什么是常识？常识就是我们在这个世界和社会生存的最基本的知识：（1）使用频率最高；（2）可以举一反三，推导并且帮助获取其它知识
  - 机器人学（机械、控制、设计、运动规划、任务规划等议题）
  - 博弈与伦理（多代理人agents的交互、对抗与合作，机器人与社会融合等议题）
  - 机器学习（各种统计的建模、分析工具和计算的方法等议题）

## Q2 人工智能的能不能

- 目前的人工智能（弱人工智能）只是一种考虑了概率的拟合函数
  - 这个函数可以非常复杂
  - 需要有非常多的参数
  - 需要极大数量训练样本
  - 需要高性能的计算能力CPU、GPU
- 函数拟合存在着局限性
  - 只能尽量拟合训练样本涉及到的函数或者说规则
  - “奥卡姆剃刀”原则：研究者会尽力泛化这些函数或规则
  - “No Free Lunch”原则：针对无限空间中的有限数据，函数不可能无限泛化下去

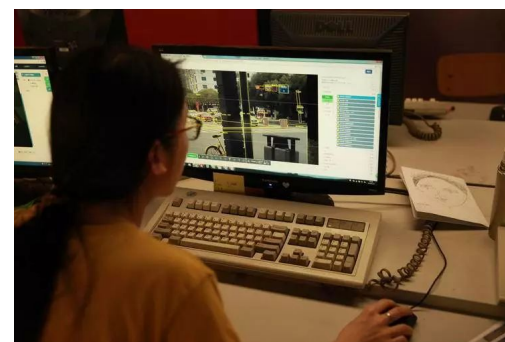
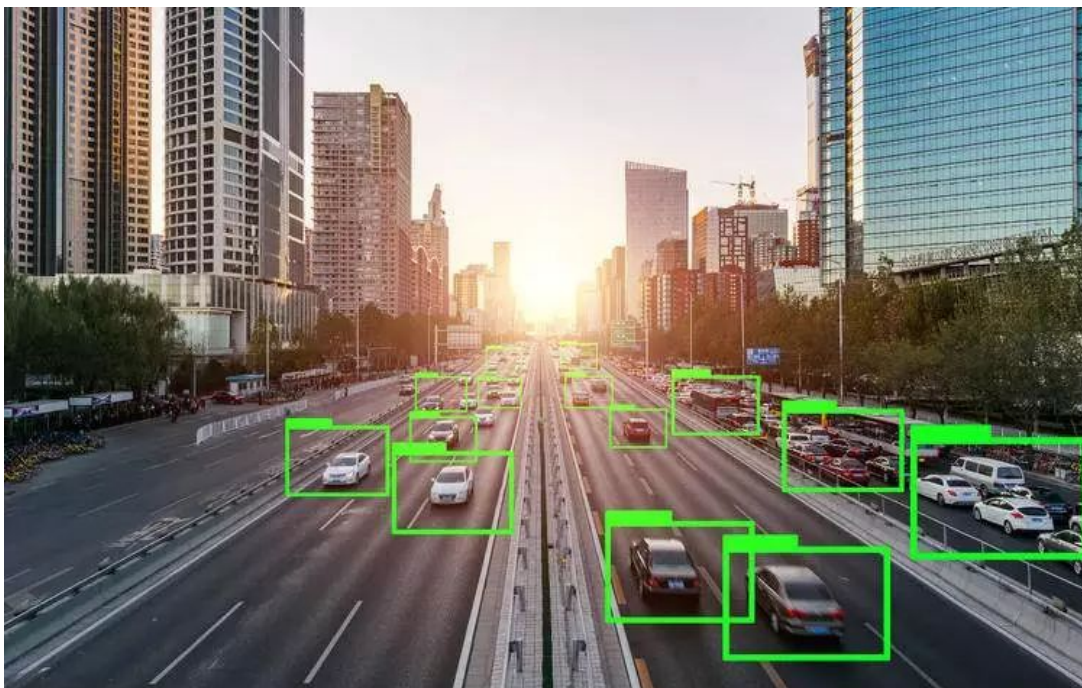


## Q2 人工智能的能和不能

- 目前人工智能存在着无法解决的问题：泛化问题
  - 不会处理训练样本范围之外的样本
  - 不会处理学习到的规则之外的样本
- 不管问题多复杂
  - 能：规则明确、范围受限，AI就有可能解决；
  - 不能：反之则可能困难重重

## Q2 人工智能的能和不能

- 当前样本的获取：一种基于体力的智能
  - 新兴市场带来了大量劳动力的涌入：在无差别人力劳动这件事上大家是没有门槛的，“原来干淘宝刷单的，现在也能摇身一变做AI数据标注”
  - 如果标注犯错，算法也会跟着犯错：95%以上的标注准确率是理想情况，但从95%提到97%所需花的成本可能是10倍或100倍

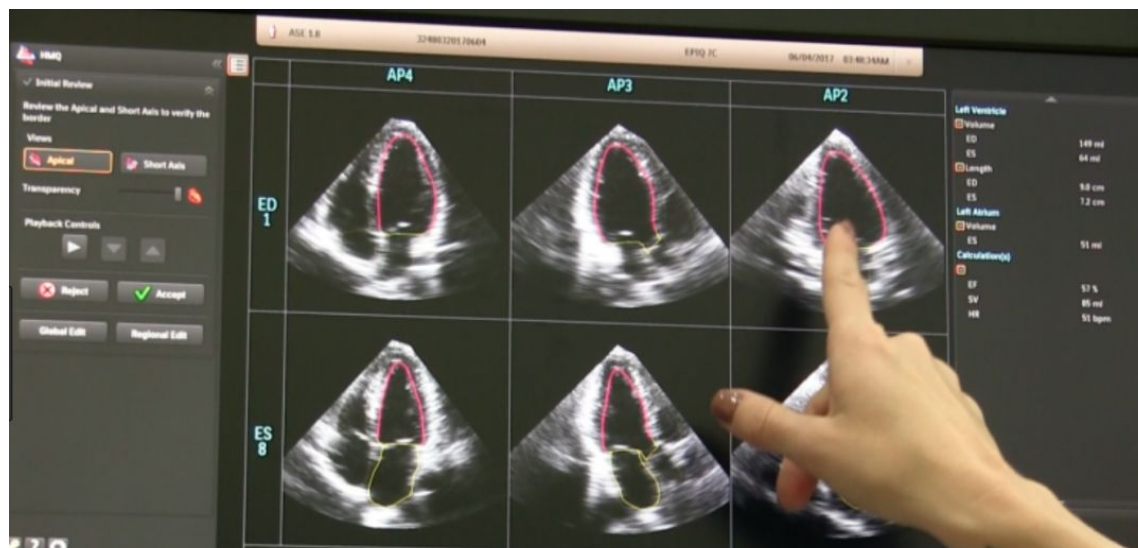


# Q3 城市能否 运行于人工智 能之上

- 城市不是一盘棋：
  - 物理实体
  - 边界开放、机制复杂、快速变化
- “科幻型人工智能城市”的几个门槛：
  - 数据采集：隐私等
  - 数据处理
  - **数据质量**
  - **算法实验：成本等**
  - 法律伦理
  - 标准规则
  - .....

# Q3 城市能否 运行于人工智 能之上

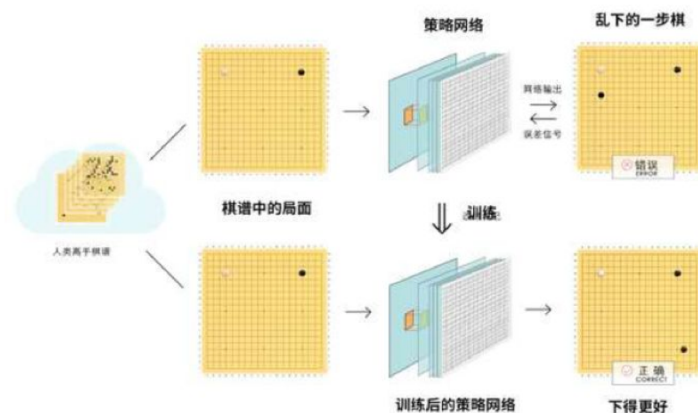
- 数据质量的门槛：
  - 新机器导致唐氏综合症诊断不断上升，由于基础数据采集质量的局限，实验室中精细调教的算法在现实应用中会暴露出大量的问题，给诊断带来直接风险



- “虽然人工智能大工程的积木块已经有了，但是把这些积木块组合在一起的原则还没有出现，**所以这些积木的组合仅仅以特例的方式存在**”

# Q3 城市能否 运行于人工智 能之上

- 算法实验的门槛：
  - 机器学习：基于训练数据（标签）改进算法
    - 一些试图摆脱“标签”输入的无监督学习技术，PCA主成分分析、k-means聚类、DBSCAN聚类、SOM自组织映射、GAN生成式对抗网络.....
  - 深度学习：用深度神经网络完成机器学习（如CNN）
    - 从死记硬背到领悟棋谱的转变，具备了给定规则（策略）的泛化能力
  - 强化学习：在条件/环境下学会正确决策，建立长远目标
    - 以试错的机制与环境进行交互，通过最大化累积奖赏来学习最优策略





# Q3 城市能否 运行于人工智 能之上

- 算法实验的门槛：

- AlphaGo：深度强化学习+蒙特卡罗树搜索（创新点）
  - 深度学习：**策略网络<sub>1</sub>**选择落子位置降低搜索宽度，**价值网络<sub>2</sub>**评估局面以减小搜索深度，使搜索效率得到了大幅提升，胜率估算也更加精确
  - 强化学习：通过**自我博弈**调整改善策略网络性能，使用自我对弈和快速走子结合形成的棋谱数据（新数据）进一步训练价值网络，实战中驱动**走字网络<sub>3</sub>**
- AlphaGo Zero：数个神经网络综合，共享一个网络
  - 神经网络权值完全随机初始化，无需先验知识：不再学习棋谱，但还需要棋盘（**确定规则和边界**）
  - 神经网络优化：结构复杂性降低、引入残差结构（更深的神经网络）等

# Q3 城市能否 运行于人工智能之上

- 算法实验的门槛：
  - 需要明确规则和边界
    - 城市、中国城市发展的规则和边界在哪？
  - 强化学习算法需要大量的试错样例才能学会给定的规则
    - 现实中城市是否能够承担试错成本：自动驾驶、新区开发.....
    - 不但需要城市发展好的案例库（ positive sample ），更需要坏的案例库（ negative sample ）
  - 海量的训练过程
    - 现实中城市无法平行计算、无法加速时间进行实验

# Q3-1 城市规划与人工智能

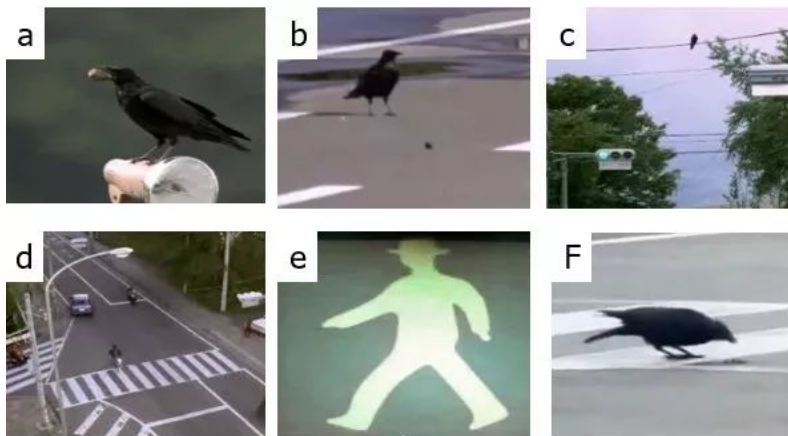
- 城市规划与人工智能：对过去
  - 缺乏城市发展路径的全方面量化
  - 尚未真正把握城市发展的规律
    - 缺乏常识：普遍的城市常识、地方化的城市常识
- 城市规划与人工智能：对未来
  - 价值观的问题：城市规划中的“价值网络”尚未建立
  - 规则的问题：城市规划方案的实施逻辑、实施顺序与价值目标的关系还没有清晰的机制，缺乏“策略网络”
  - 建设主体的问题：规划中的价值和策略不一定能够驱动规划“落子”的主体
  - **群体协调策略难以制定**：多代理人agents的交互、对抗与合作



# Q4 人工智能 的未来走向

- 乌鸦

- 通过观察行人与红绿灯的关系，自主学习交通安全规则，获取车轮碾压之后的坚果
- 完全自主的智能。感知、认知、推理、学习和执行
  - 通过少量数据学习获得
  - 不到人脑1%大小的脑容量



- 鹦鹉

- 有很强的语言模仿能力，类似于当前的由数据驱动的聊天机器人

# Q4 人工智能 的未来走向

- 向动物学习，建立新的认知构架
  - 物理环境客观的现实与因果链条。在不同环境条件下，智能的形式会是不一样的，任何智能的机器必须理解物理世界及其因果链条，适应这个世界
  - 动物的行为都是被各种任务驱动。任务代表了价值观和决策函数，这些价值函数很多在进化过程中就已经形成了，包括人脑中发现的各种化学成分奖惩调制，如多巴胺（快乐）、血清素（痛苦）、乙酰胆碱（焦虑、不确定性）、去甲肾上腺素（新奇、兴奋）等。
    - 动物与生俱来的任务与价值链条。这个任务是一个生物进化的“刚需”。如个体的生存，要解决吃饭和安全问题，而物种的传承需要交配和社会活动。这些基本任务会衍生出大量的其它的“任务”。有了物理环境的因果链和智能物种的任务与价值链，那么一切都是可以推导出来的。
- 构造一个智能系统
  - 定义好身体的基本行动的功能，
  - 再定一个模型的空间（包括价值函数）。
  - 降临在某个环境和社会群体之中，就应该自主地生存，就像乌鸦那样找到一条活路，认识世界、利用世界、改造世界。

## Q4 人工智能 的未来走向

- 外来的数据

- 外部世界通过各种感知信号，传递到人脑，塑造我们的模型。  
数据来源于观察（ observation ）和实践（ experimentation ）
  - 观察的数据一般用于学习各种统计模型，这种模型就是某种时间和空间的联合分布，也就是统计的关联与相关性。
  - 实践的数据用于学习各种因果模型，将行为与结果联系在一起。  
因果与统计相关是不同的概念。

- 主体内在的任务

- 这就是由内在的价值函数驱动的行为、以期达到某种目的。  
我们的价值函数是在生物进化过程中形成的。因为任务的不同，我们往往对环境中有些变量非常敏感，而对其它一些变量不关心，由此形成不同的模型

## Q4 人工智能 的未来走向

- 任务驱动的因果推理与学习
  - 《史记》“货殖列传”：“天下熙熙，皆为利来；天下攘攘，皆为利往。”
- 什么是任务：任务就是在改变场景中某些物体的状态，连续的改变形成“流态” fluent
  - 重在**改变**，而不仅是认识
  - 番茄酱与瓶子的空间位置关系是一个流态，可以被挤出来；
  - 社会关系：从一般人，到朋友、再到密友等
  - 人类和动物忙忙碌碌，都是在改变各种流态，以提高我们的价值函数（利益）

## Q4 人工智能 的未来走向

- 现有认知构架：大数据、小任务（ big data for small task ）
  - 当前的深度学习：针对某个特定的任务，如人脸识别和物体识别，设计一个简单的价值函数Loss function，用大量数据训练特定的模型。这种方法在某些问题上也很有效。
  - 难以泛化和解释
- 建立新的认知构架：小数据、大任务（ small data for big tasks ）
  - 用大量任务、而不是大量数据来塑造智能系统和模型
  - 任务塑造智能

## Q4 人工智能 的未来走向

- 从相关分析到因果推理
  - 人工智能陷入概率关联（probabilistic association）泥潭，在重复上一代机器已经能做的事情：从大量数据中发现隐藏规律
- 因果推理：提供人类水平的智能
  - 因果推理（causal reasoning）取代关联推理（reasoning by association）
  - 不仅是将发烧和疟疾联系起来，还需要具备推理出是疟疾导致发烧的能力
    - 具备因果推理的机制
    - 进一步可提出虚拟的问题——追问在有干扰的情况下，因果关系将如何变化
- 环境模型 + 主体模型：贝叶斯网络

# Q5 提高城市智能的几种方式

- 从IT技术到DT技术

- IT ( Information Technology ) 信息技术，是一项基于计算机和互联网用来提升人们信息传播能力的技术
  - 延伸人类信息传递的器官（触觉、视觉、听觉、肢体、语言、声音等），提升人类沟通效率和沟通能力
- DT ( Data Technology ) 数据技术，数据技术的本质是数据的加工技术
  - 数据反映大千世界事物和活动，数据技术让我们能够借助计算机能力认知这个世界、思考背后规律和逻辑、选择应对措施，所以数据技术的本质也是“认知”的技术、“思考”的技术和“决策”的技术，最终形成“人工智能”

## 信息技术（IT）的本质是信息传播方式



## 大数据的本质是信息加工方式



# Q5 提高城市智能的几种方式

- 在数据科学的指引下开展城市研究：明确的任务
  - 人口分布预测：
    - 清华同衡：长短期记忆网络（LSTM）进行城市人口时空分布的短期预估
    - 中规院-百度：深度残差网络（ResNet）预测未来人流量（短时）
  - 交通出行：
    - 周学松-北京交通发展研究院：
      - 建立北京交通计算图
      - 传统的BP神经网络：第一层交通生成层，第二层交通链路层，每一层都有实际的交通意义，对应不同的交通数据
      - 掌握交通需求的复杂结构性与层次性，对交通现象的产生可基于大数据从本质上进行解释、估计与预测结果



# Q5 提高城市智能的几种方式

- 向“动物智能”学习的城市智慧
- 探索城市发展的“任务库”
  - 环境变化
  - 细颗粒度的正负面城市案例
- 探索城市居民的“需求库”
  - 适者生存的价值函数
  - 群体决策协调
- 探索城市发展因果链条的“算法库”
  - 物理环境客观的现实与因果链条

感谢！