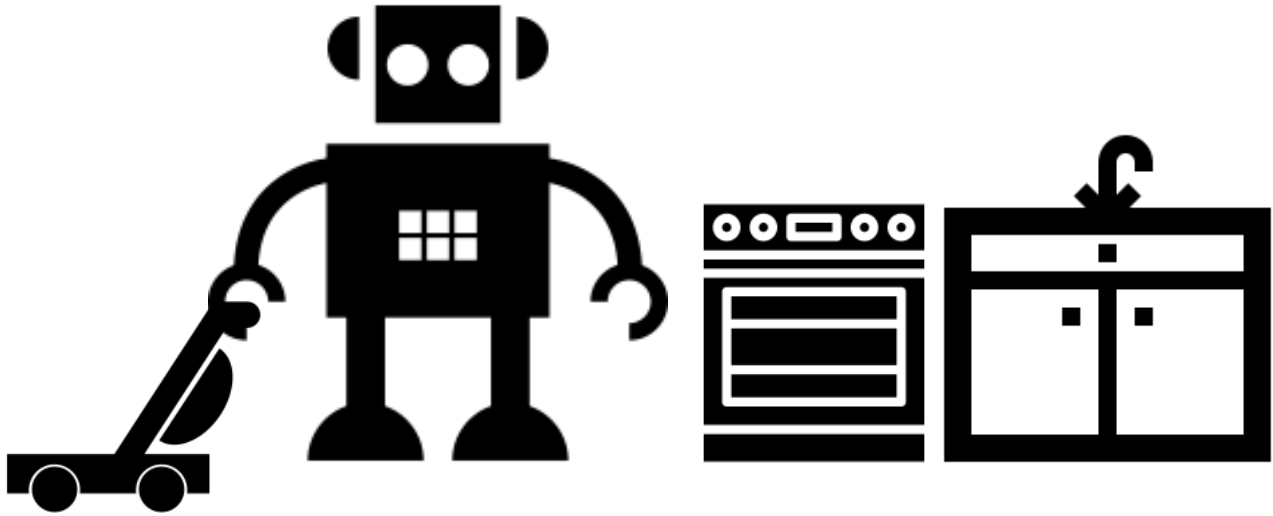
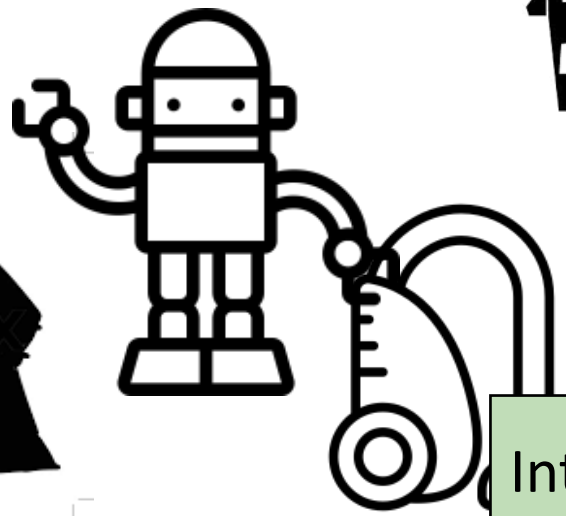
Bridging intelligent robotics and cognitive science

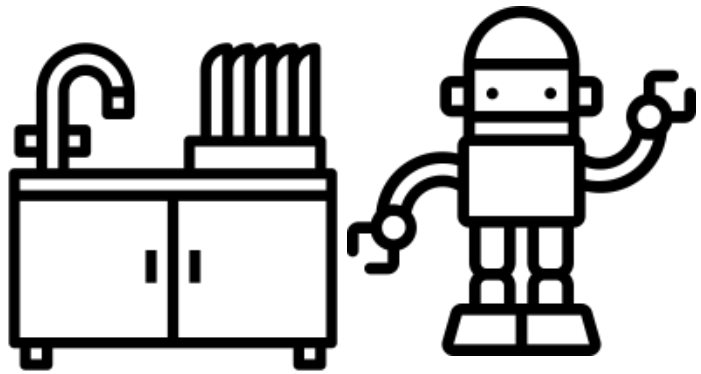
Leslie Pack Kaelbling
MIT CSAIL

My research goal:
understand the computational mechanisms
necessary to make
a general-purpose intelligent robot



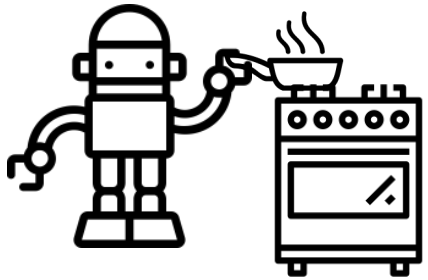
Definition of intelligence: Do any job in any house! ;-)

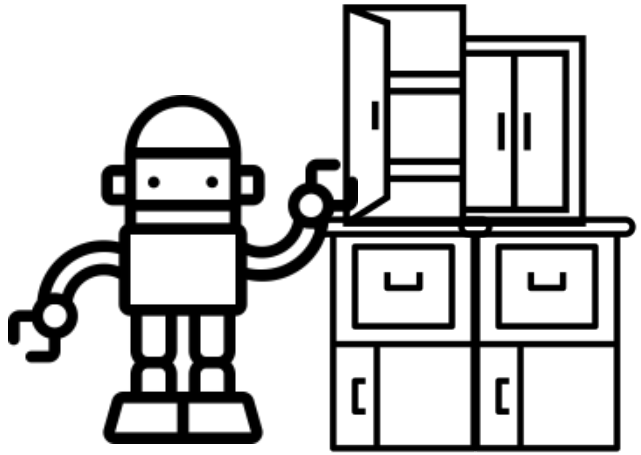




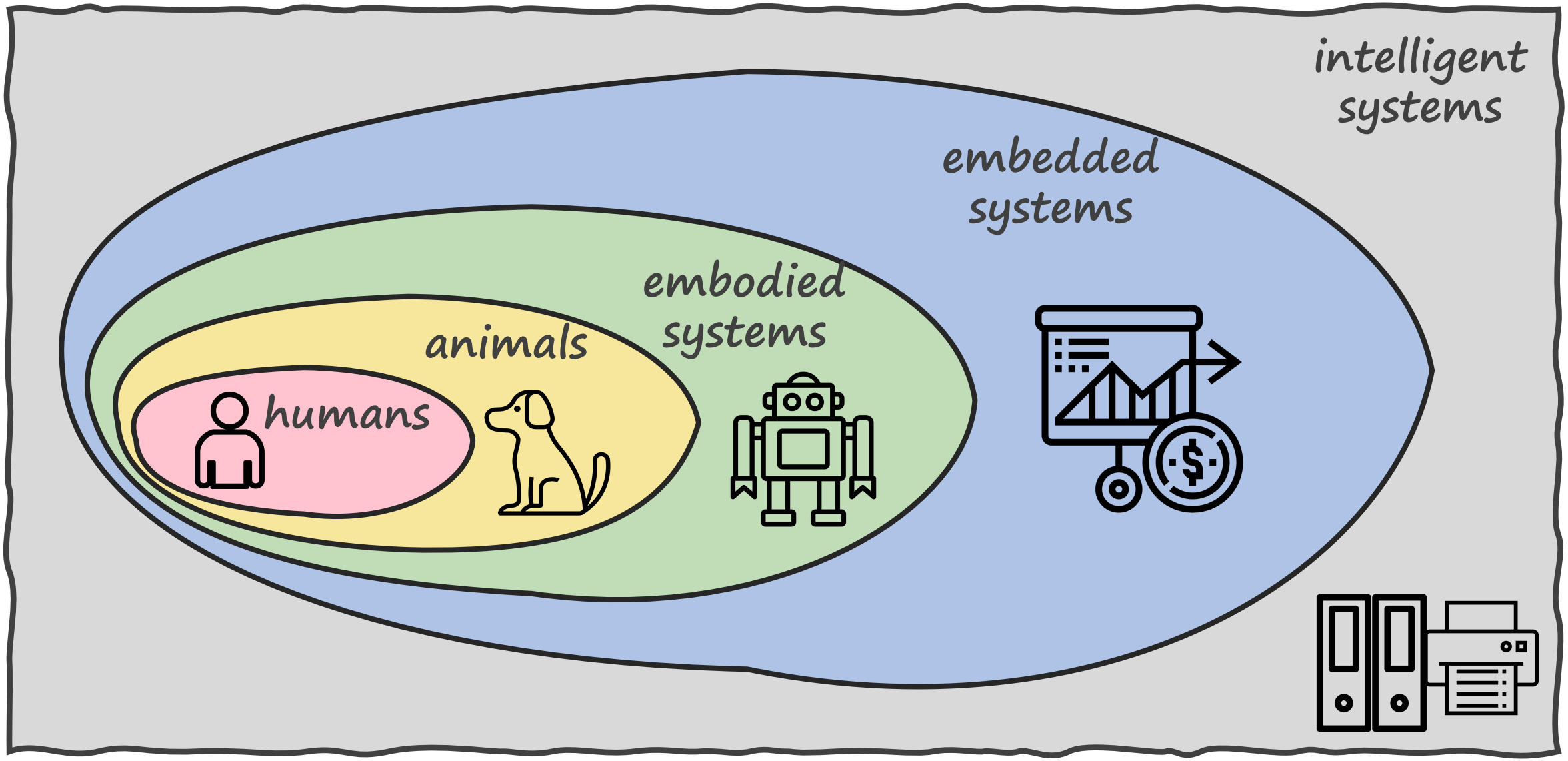
Intelligent behavior:

- flexible
- robust
- purposeful
- adaptable
- long-horizon
- ….





Intelligent systems: what subset are we studying/building?



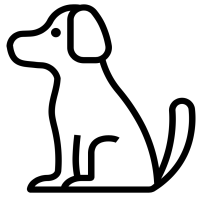
Many possible relationships between artificial intelligence and natural science

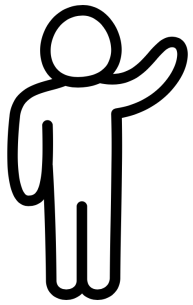
1. We should completely understand brains at a **neuronal level** and then try to replicate their processing in detail in computers
2. We should understand brains at a functional/**algorithmic level** and then try to replicate those algorithms in computers
3. We should understand animal or human behavior at an **input/output level** and then try to replicate that behavior in computers
4. We should **replicate evolution** in simulation and see if what we end up with resembles natural systems
5. We should **engineer** intelligent computer systems and see if what we end up with resembles natural systems

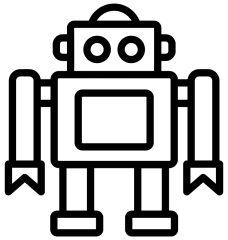
The way I think about building intelligent embodied systems

- I want to really build these systems
- If humans are going to do the engineering, that imposes some constraints on the solution and/or the process by which we arrive at a solution
- Nature solves problems in ways that are beautiful but sometimes very difficult to understand
- We humans might have more success at building embodied intelligent systems in ways that are significantly non-natural
- I am still happy to get any help I can from studying natural systems

A science of intelligence

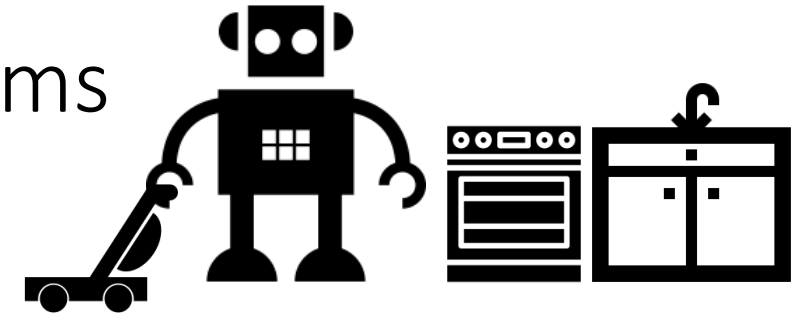






General principles of intelligent information processing and control

A basis of computational mechanisms



Learn

- transition models
- inference rules
- search control

Build in general representation and inference mechanisms:

- feedback control
- convolution in space and time
- kinematics and motion planning
- forward/backward causal inference
- abstraction over objects
- state abstraction/aggregation
- temporal abstraction
- utility maximization

Some things I know that might be useful to natural scientists

- Discrete search (usually) takes time either
  - exponential in the length of the solution
  - linear in the size of the state space
- Local optimization (usually) finds locally optimal solutions
- Generalization (usually) requires an amount of data exponential in some measure of the effective complexity of the hypothesis space
- Closed-loop feedback can (often) make up for approximate reasoning with poor models

Some things I wish I knew about natural intelligences

- What kinds of “knowledge” are innate?
  - Individuals need to learn from their environment with small amounts of data
- What corners can we safely cut?
  - We like to pose inference problems in terms of search or optimization, but optimality is intractable
- What kinds of modularity do we see in brains?
  - Modularity is very important to human engineers
- How do brains encode spatial information:
  - To support short-term obstacle avoidance?
  - To support long-term navigation?
  - To manipulate their limbs to grasp objects?
  - To make judgements about whether an object (or the agent!) will fit in a space?

More things I wish I knew about natural intelligences

- There are surely multiple scales and mechanisms of learning in nature, right?
  - Do any of them show effective extrapolation (rather than interpolation)?
- What mechanisms (mostly) keep (most) animals from fruitlessly repeating unsuccessful actions?
- What are some plausible models of how natural intelligences model other agents?

A not-yet-intelligent robot