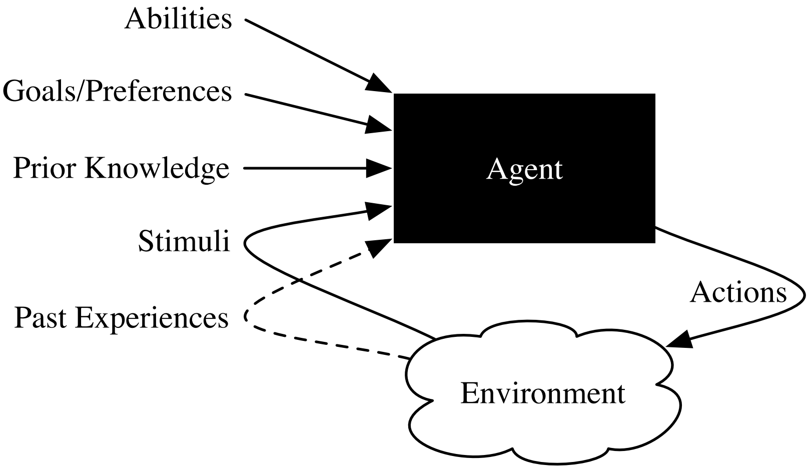
Course Overview

- Agents acting in an environment
- Future and Ethics of AI
- Dimensions of complexity

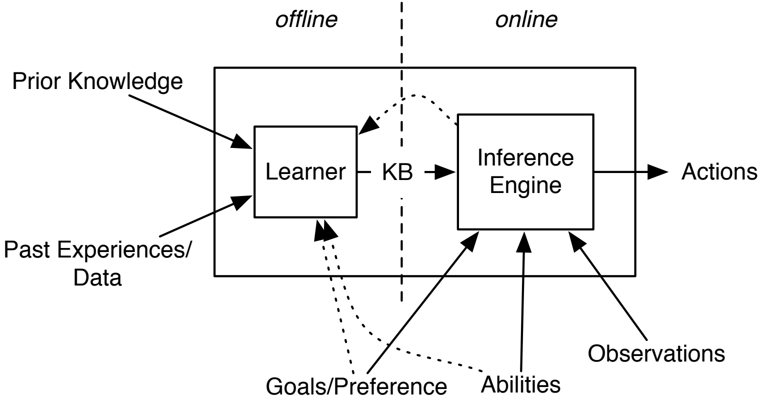
What is Artificial Intelligence?

- Artificial Intelligence is the synthesis and analysis of computational agents that act intelligently.
- An agent is something that acts in an environment.
- An agent acts intelligently if:
  - its actions are appropriate for its goals and circumstances
  - it is flexible to changing environments and goals
  - it learns from experience
  - it makes appropriate choices given perceptual and computational limitations

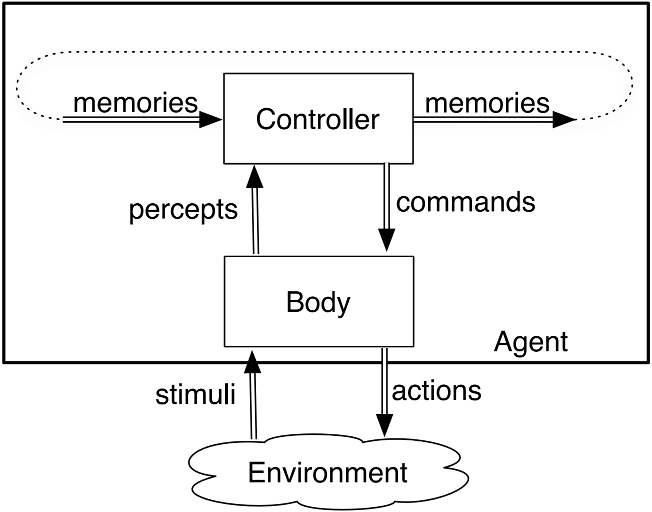
Agents acting in an environment



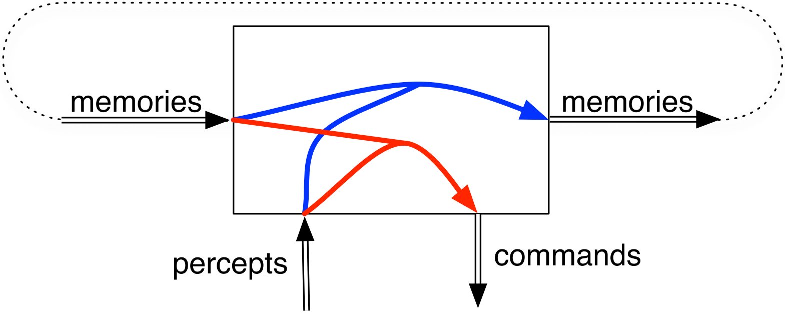
Inside Black Box



Controller


memories
Controller
memories
percepts
commands
Body
Agent
stimuli
actions
Environment

Functions implemented in a controller



For discrete time, a controller implements:

- belief state function returns next belief state / memory.
  What should it remember?
- command function returns commands to body.
  What should it do?

Future and Ethics of AI

- What will super-intelligent AI bring?

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if people are not longer needed to make economy work?

Future and Ethics of AI

- What will super-intelligent AI bring?
    - Automation and unemployment? What if people are not longer needed to make economy work?
    - Smart weapons? Automated terrorists?

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if people are not longer needed to make economy work?
  - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if people are not longer needed to make economy work?
  - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?
  - predict the future
  - optimize (constrained optimization)

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if people are not longer needed to make economy work?
  - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?
  - predict the future
  - optimize (constrained optimization)
- Whose values/goals will they use? (Why?)

Future and Ethics of AI

- What will super-intelligent AI bring?
    - Automation and unemployment? What if people are not longer needed to make economy work?
    - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?
    - predict the future
    - optimize (constrained optimization)
- Whose values/goals will they use? (Why?)
- Will we need a new ethics of AI?

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if people are not longer needed to make economy work?
  - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?
  - predict the future
  - optimize (constrained optimization)
- Whose values/goals will they use? (Why?)
- Will we need a new ethics of AI?
- Is super-human AI inevitable (wait till computers get faster)? (Singularity)

  Is there fundamental research to be done?

  Is it easy because humans are not as intelligent as we like to think?

Dimensions of Complexity

- Flat or modular or hierarchical
- Explicit states or features or individuals and relations
- Static or finite stage or indefinite stage or infinite stage
- Fully observable or partially observable
- Deterministic or stochastic dynamics
- Goals or complex preferences
- Single-agent or multiple agents
- Knowledge is given or knowledge is learned from experience
- Reason offline or reason while interacting with environment
- Perfect rationality or bounded rationality

State-space Search

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Classical Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Decision Networks

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Markov Decision Processes (MDPs)

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Decision-theoretic Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Reinforcement Learning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Relational Reinforcement Learning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Classical Game Theory

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Humans

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- reason offline or reason while interacting with environment
- perfect rationality or bounded rationality

Comparison of Some Representations

	CP	MDPs	IDs	RL	POMDPs	GT
hierarchical	✔					
properties	✔		✔	✔		
relational	✔					
indefinite stage	✔	✔		✔	✔	
stochastic dynamics		✔	✔	✔	✔	✔
partially observable			✔		✔	✔
values		✔	✔	✔	✔	✔
dynamics not given				✔		
multiple agents						✔
bounded rationality						