Learning Objectives

At the end of the class you should be able to:

- characterize simplifying assumptions made in building AI systems
- determine what simplifying assumptions particular AI systems are making
- suggest what assumptions to lift to build a more intelligent system than an existing one

Dimensions

- Research proceeds by making simplifying assumptions, and gradually reducing them.
- Each simplifying assumption gives a dimension of complexity
  - multiple values in a dimension: from simple to complex
  - simplifying assumptions can be relaxed in various combinations

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
- Perfect rationality or bounded rationality

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to a see the Mona Lisa in Paris.

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to a see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical
- A flat description is either continuous or discrete. Hierarchical reasoning is often a hybrid of continuous and discrete.

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.
- Individuals and relations
  - There is a feature for each relationship on each tuple of individuals.
  - Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

Planning horizon

...how far the agent looks into the future when deciding what to do.

- **Static:** world does not change
- **Finite stage:** agent reasons about a fixed finite number of time steps
- **Indefinite stage:** agent reasons about a finite, but not predetermined, number of time steps
- **Infinite stage:** the agent plans for going on forever (process oriented)

Uncertainty

There are two dimensions for uncertainty. In each dimension an agent can have

- **No uncertainty:** the agent knows which world is true
- **Disjunctive uncertainty:** there is a set of worlds that are possible
- **Probabilistic uncertainty:** a probability distribution over the worlds.

Why Probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - disjunctions: be careful or you will be run over
  - point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Uncertain dynamics

If an agent knew the initial state and its action, could it predict the resulting state?

The dynamics can be:

- Deterministic : the resulting state is determined from the action and the state
- Stochastic : there is uncertainty about the resulting state.

Sensing Uncertainty

Whether an agent can determine the state from its observations:

- Fully-observable: the agent can observe the state of the world.
- Partially-observable: there can be a number states that are possible given the agent's observations.

Goals or complex preferences

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters
  - cardinal absolute values also matter
- Examples: coffee delivery robot, medical doctor

Single agent or multiple agents

- Single agent reasoning is where an agent assumes that any other agents are part of the environment.
- Multiple agent reasoning is when an agent reasons strategically about the reasoning of other agents.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

Perfect rationality or bounded rationality

- **Perfect rationality:** the agent can determine the best course of action, without taking into account its limited computational resources.
- **Bounded rationality:** the agent must make good decisions based on its perceptual, computational and memory limitations.

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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- perfect rationality or bounded rationality

The Dimensions Interact in Complex Ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:
    - Hierarchical reasoning
    - Individuals and relations
    - Bounded rationality