# Introduction to Artificial Intelligence

Raymond Bian

January 13, 2024

## Contents



### Lecture 1: Intro to AI

**Definition 1.** **Artificial Intelligence** is the science of making machines that think like people, think rationally, act like people, and act rationally.

**Definition 2.** **Machine Learning** is a subfield of AI, where statistical approaches are used. It is also a subfield of Data Science, which has other fields such as visualization and statistics.

### Lecture 2: Intro to AI and Agents

## 1 Agents and Environments

**Definition 3.** An **agent** is an entity that perceives and acts. It perceives the environment through **sensors** and acts on the environment through **effectors**.

**Definition 4.** A **rational agent** selects actions that maximize its utility.

**Definition 5.** Characteristics of the **percepts environment**, and **action space** dictate techniques for selecting actions.

**Definition 6.** An **agent function** is a mathematica description of an agent's behavior that maps sensory perceptions to effector actions:

$$F : P \to A.$$

We should keep things as simple as possible. For example, tabulation itself is very strong and powerful in the right situations. However, this table could be too big for memory, and it could resolve in an infinite loop.

To write an agent program, one must establish:

1. Actions: How the agent changes the environment
2. Sensors: What you can know about the environment
3. Prior Knowledge: Pre-loaded information
4. Objective Function: What the agnet is trying to accomplish
5. Measurement Function: How to tell if we are succeeeding or not

**Definition 7.** Steps 4 and 5 detail **rationality**: "Of all my actions, which ones get me closer to my objective, given what I know right now?"

**Example.** For the bugs in the room, you could award 1 point per bug-free room per time step, and a point for each bug killed. You could remove 1 point for each move action, to prevent the infinite loop. If we don't have a map, we can either explore or exploit. And finally, sensors and effectors could be unreliable.

**Definition 8.** **Reflex agents** choose actions based on the current percept (and maybe memory), and do not consider the future consequences of their actions.

Now, can a reflex agent be rational?

**Example.** If you put a cranberry in front of a frog, it will eat it, then spit it out immediately. Then, it will see the cranberry, and eat it again. This is an example of a reflex agent, and is also how you can "*infinite loop a frog*".

The moral of the story? Do the stupid thing first, it might just work.

**Definition 9.** A **model based reflex agent** can model its actions on the world first, then makes

an action depending on the mmodel.

**Definition 10.** **Planning agents** ask "what if". They make decisions based on hypothesized consequences of actions.

**Definition 11.** **Utility agents** use some utility to evaluate the hypothetical consequences of actions.

**Definition 12.** **Learning agents** get feedback, changes, and improves it knowledge with each action, allowing them to gain knowledge over time.

"If you can't win, just confuse them." Below are some definitions for different environments.

**Definition 13.** **Observability** means that sensors can observe all parts and variables in the environment.

**Definition 14.** **Determinism** means that the world changes exactly as desired. **Stochastic** mean that there is randomness and uncertainty.

**Definition 15.** **Static** means the world doesn't change while agent is deliberating

**Definition 16.** **Discreteness** means that the world is broken up into discrete chunks.

**Definition 17.** **Episodic** means that the story doesn't matter.

**Definition 18.** There many be single or multiple **agents** in an environment.