Reinforcement Learning Basics

Onur Akman

Complex Social Systems 2025

16/04/2025

onur.akman@uj.edu.pl

Reinforcement Learning: Third Paradigm

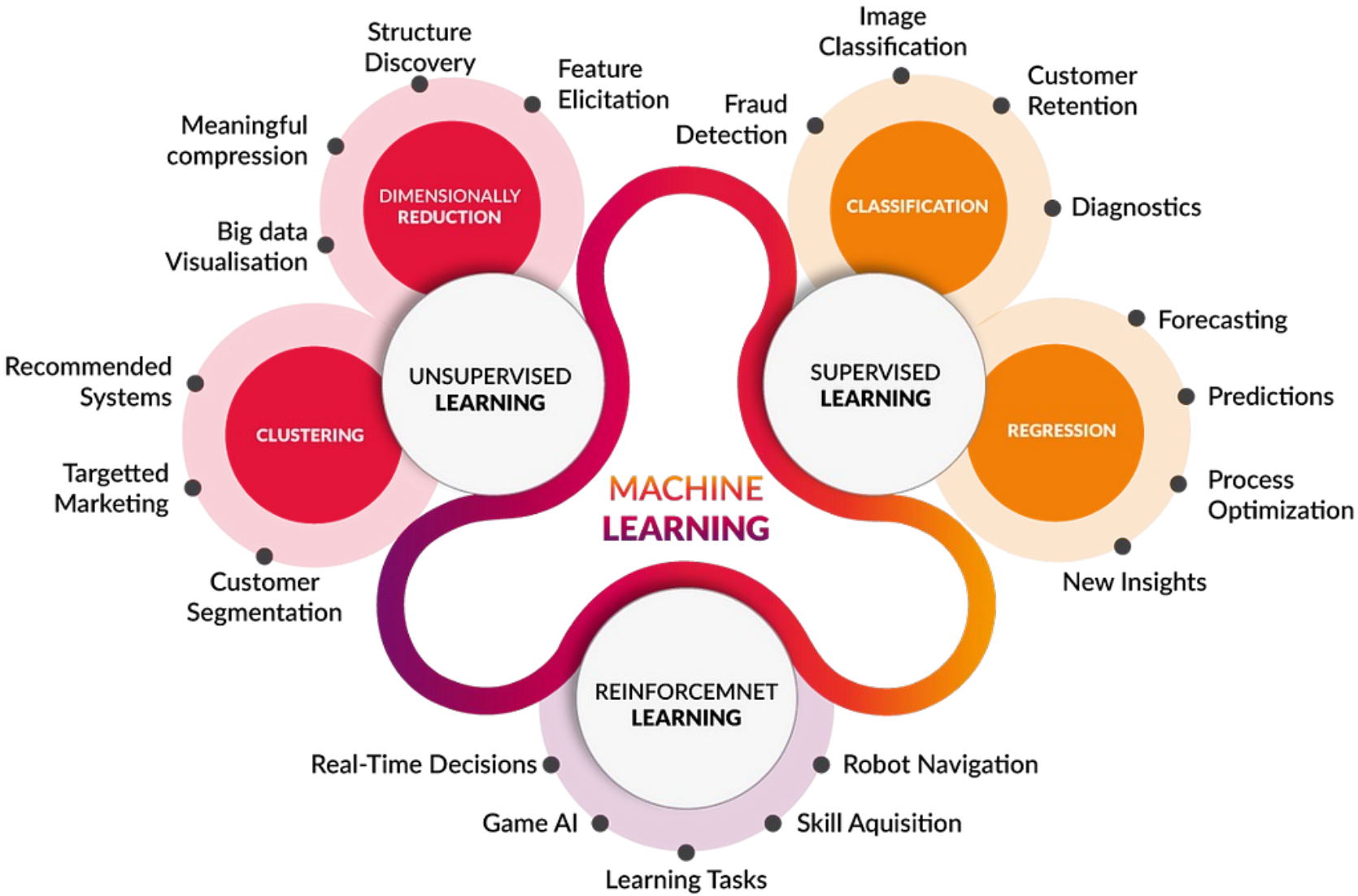
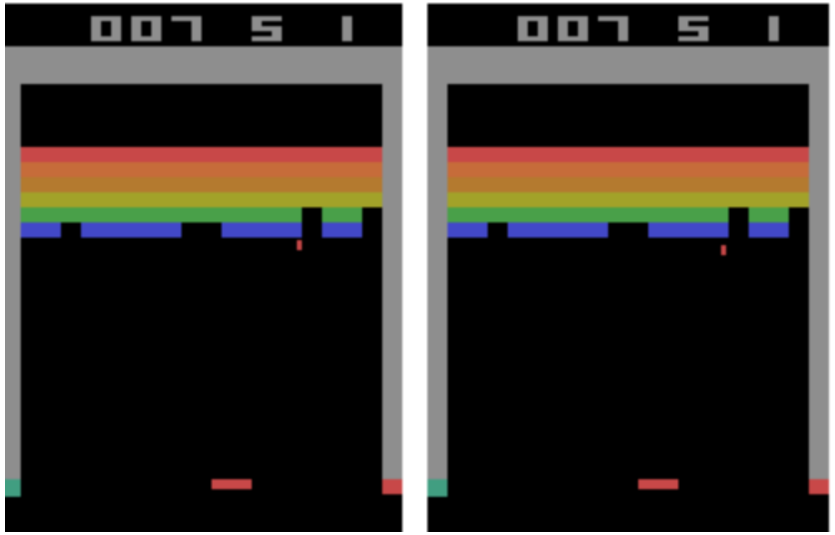


Figure credit: Different machine learning models, Deepika Yadav, medium.com

Decision Making

- RL is mainly concerned with decision making.
  - **How to make decisions for optimal results?**
- This question addresses tasks like:

Self-flying helicopters:
Learning to maneuver



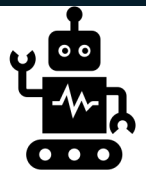
Playing atari games:
Learning to control

Robotic tasks:
Learning to rotate joints

Decision Making Process (2)

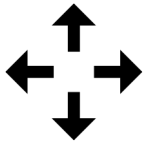
- In contrast to supervised learning, **RL does not rely on labeled data**.
- Task is formulated as a **decision-making process**, and learning is through **interactions**.
- **Designer doesn’t necessarily know the solution, but knows the required outcome**.

Markov Decision Process

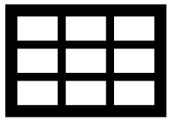


- **Agent**: A decision-maker, learns to „decide better”.

- **Environment**: A structure which receives actions, simulates transitions, emits feedback.



- **Actions**: Agent’s decisions to manipulate the environment.



- **States**: Summary of the current status of environment.

- **Reward**: Feedback signal, representing the agent’s objective, reflects agent’s success rate.

Markov Decision Process (2)

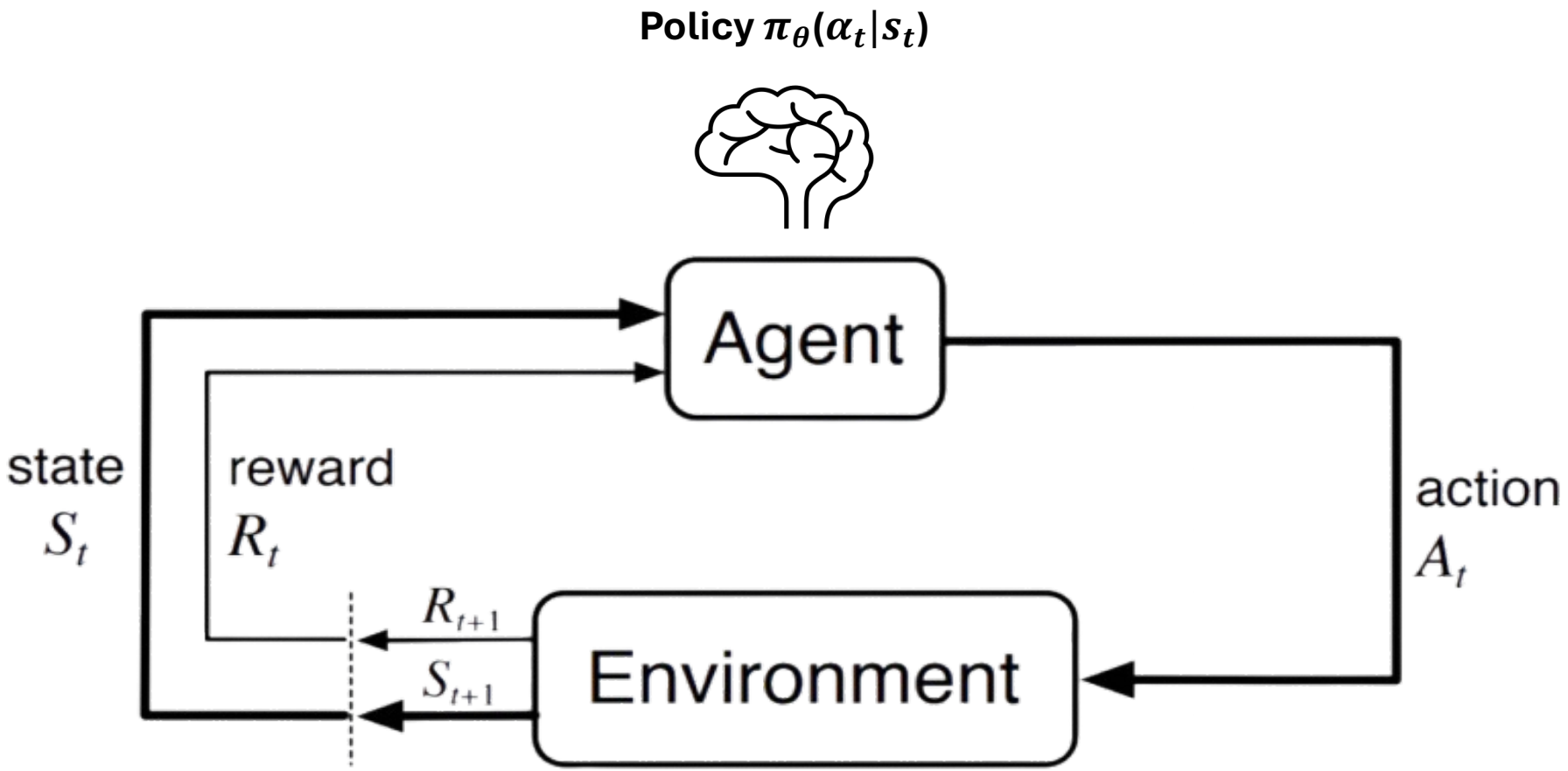
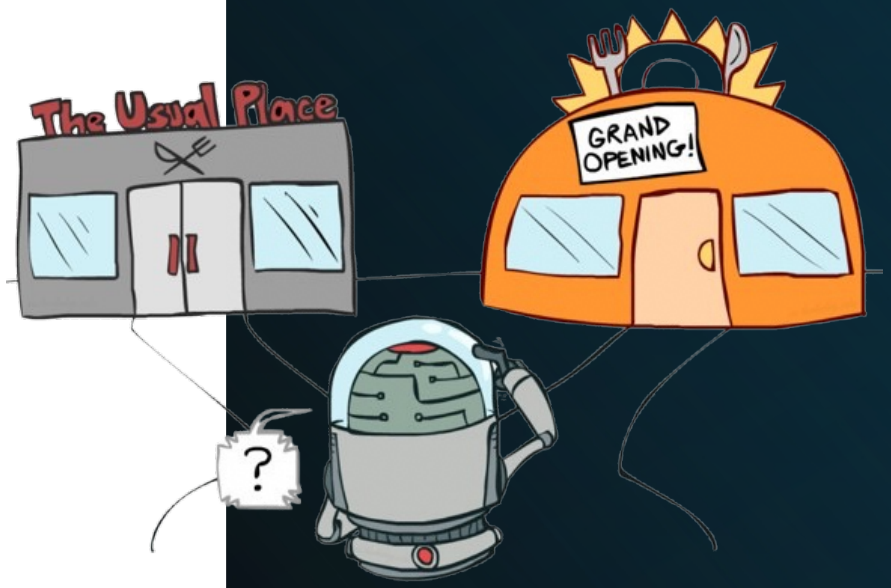


Figure credit: RS Sutton, AG Barto, Reinforcement learning: An Introduction, 2nd edition, MIT Press

What can go wrong?

- **Exploration vs Exploitation**
  - Agents need to try new things, while also sufficiently exploiting known solutions.
- **Credit Assignment**
  - Agent needs to learn which particular action in an action sequence contributes to success/failure.
- **Reward formulation**
  - Reward formulation should accurately reflect the desired outcome.



Reinforcement Learning

On-policy vs off-policy

a) Learn from your interactions in real time.

b) Learn from collected experiences in an offline setting.

Model-based vs Model-free

a) Learn an environment model, calculate a solution.

b) Learn a solution.

Value-based vs policy-based

a) Learn a value function, and act according to your expectations.

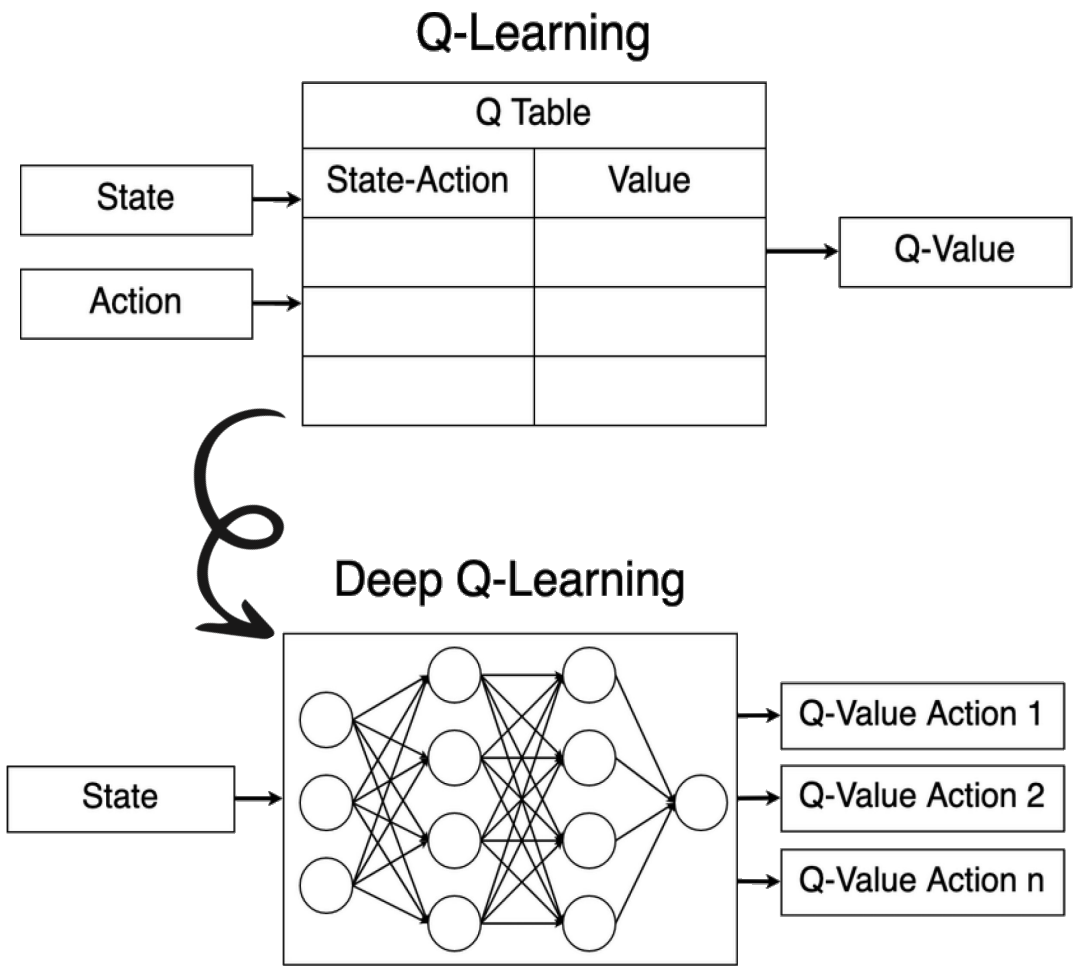
b) Learn which action is the best for a given state.

Q-Learning

- Agent may learn a Q-function, which maps a state-action pair to a return expectation.

Initialize $Q(s, a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(\textit{terminal-state}, \cdot) = 0$
Repeat (for each episode):
  Initialize S
  Repeat (for each step of episode):
    Choose A from S using policy derived from Q (e.g., ϵ-greedy)
    Take action A, observe R, S'
    $Q(S, A) \leftarrow Q(S, A) + \alpha\left[R + \gamma \max_a Q(S', a) - Q(S, A)\right]$
    $S \leftarrow S'$;
  until S is terminal

Towards deep-RL: DQN



Fill a tabular memory with the state-action values of each known state-action pair.

Use function approximators (e.g. DNNs) for better generalization.

Taxonomy of Deep-RL

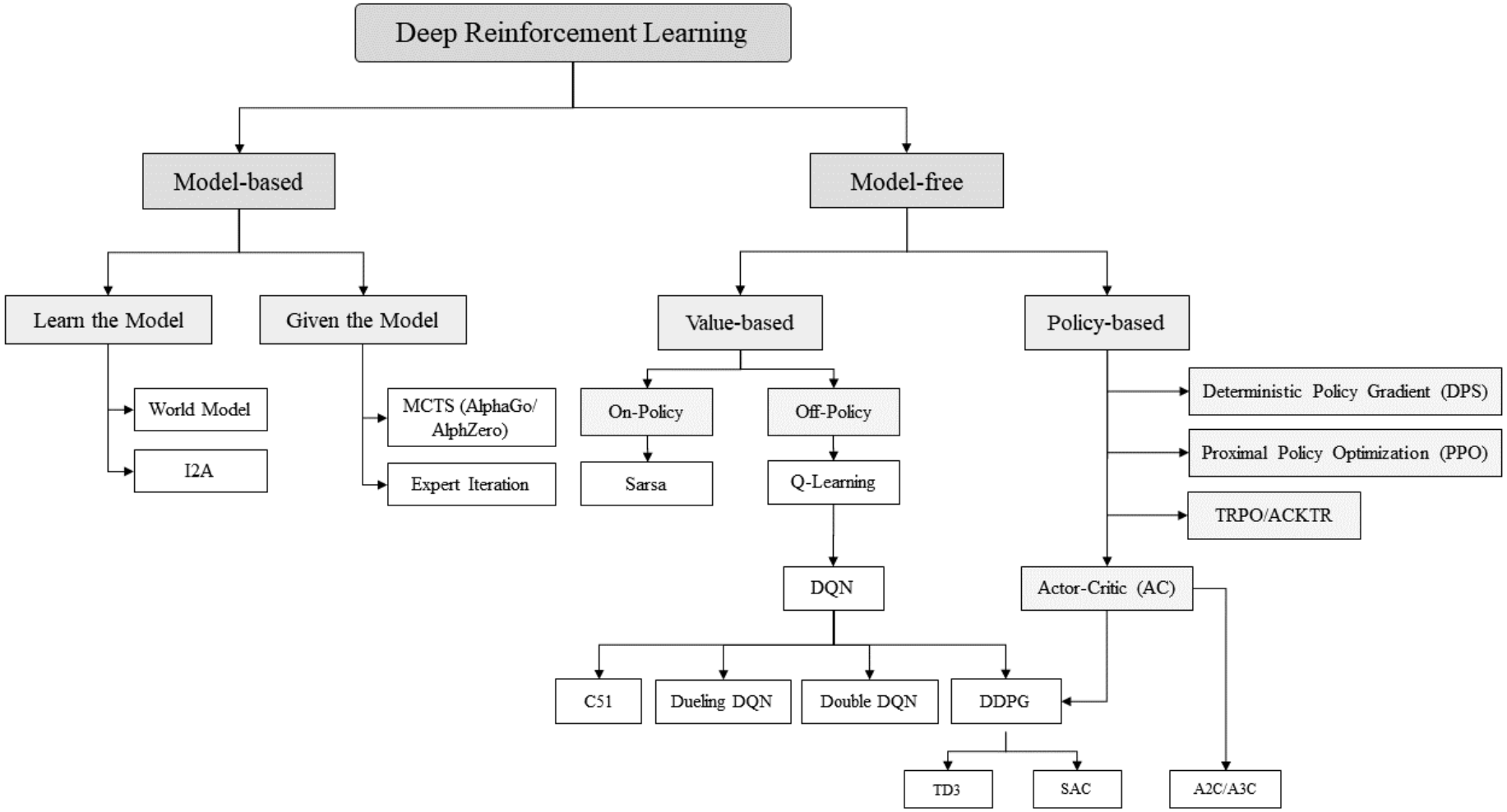
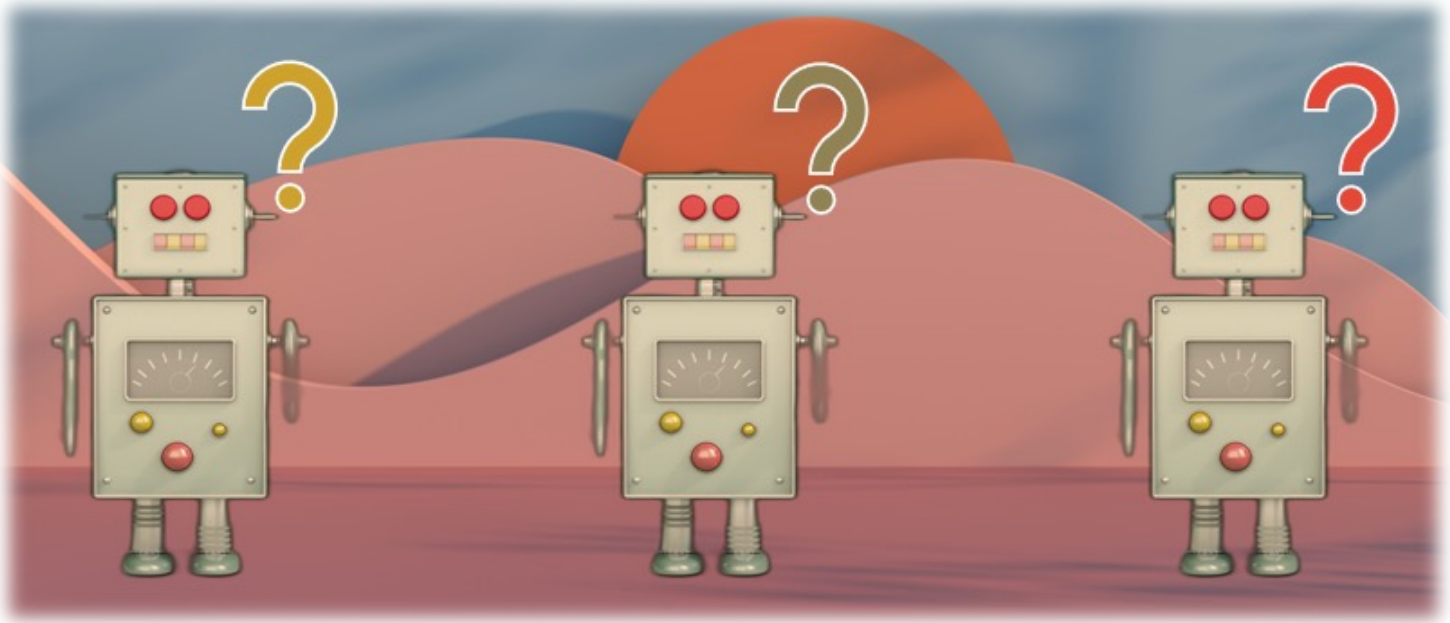
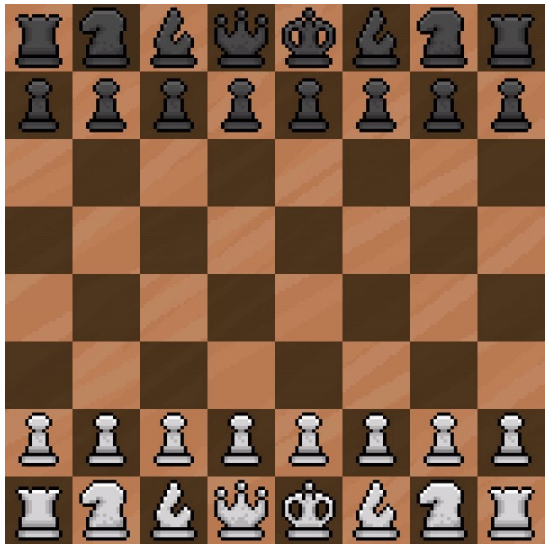


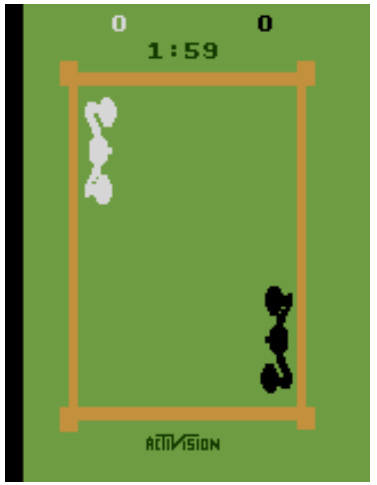
Figure credit: AI Search Algorithms for Smart Mobility by Alaa Khamis and Yinan Wang, at smartmobilityalgorithms.github.io, accessed in 11/24.

Multi-agent Reinforcement Learning

- Multi-agent RL extends single-agent RL with **multiple decision-makers existing in a shared environment**.







Depending on the problem, these agents can:

- Cooperate
- Compete
- Mix of both
- None of the above

What else can go wrong?

- Non-stationarity
  - Environment dynamics become non-stationary **as other agents evolve**, making it harder to find an optimal strategy.
- Large spaces
  - In cooperative tasks, **the joint state-action space becomes larger with the increasing numer of agents**, making it harder to sufficiently explore.
- Multi-agent credit assignment
  - Need to identify each agent’s contribution to success/failure.

Training in MARL

- **Independent learning**
  - Treat other agents as a part of the environment, learn as if it's a single-agent task.
- **Centralized learning**
  - Learn a collective strategy with collective training.
  - Can use a central coordinator or inter-agent communication.
- **Opponent modeling**
  - Learn about your opponent's strategy to come up with the best counter-strategy.