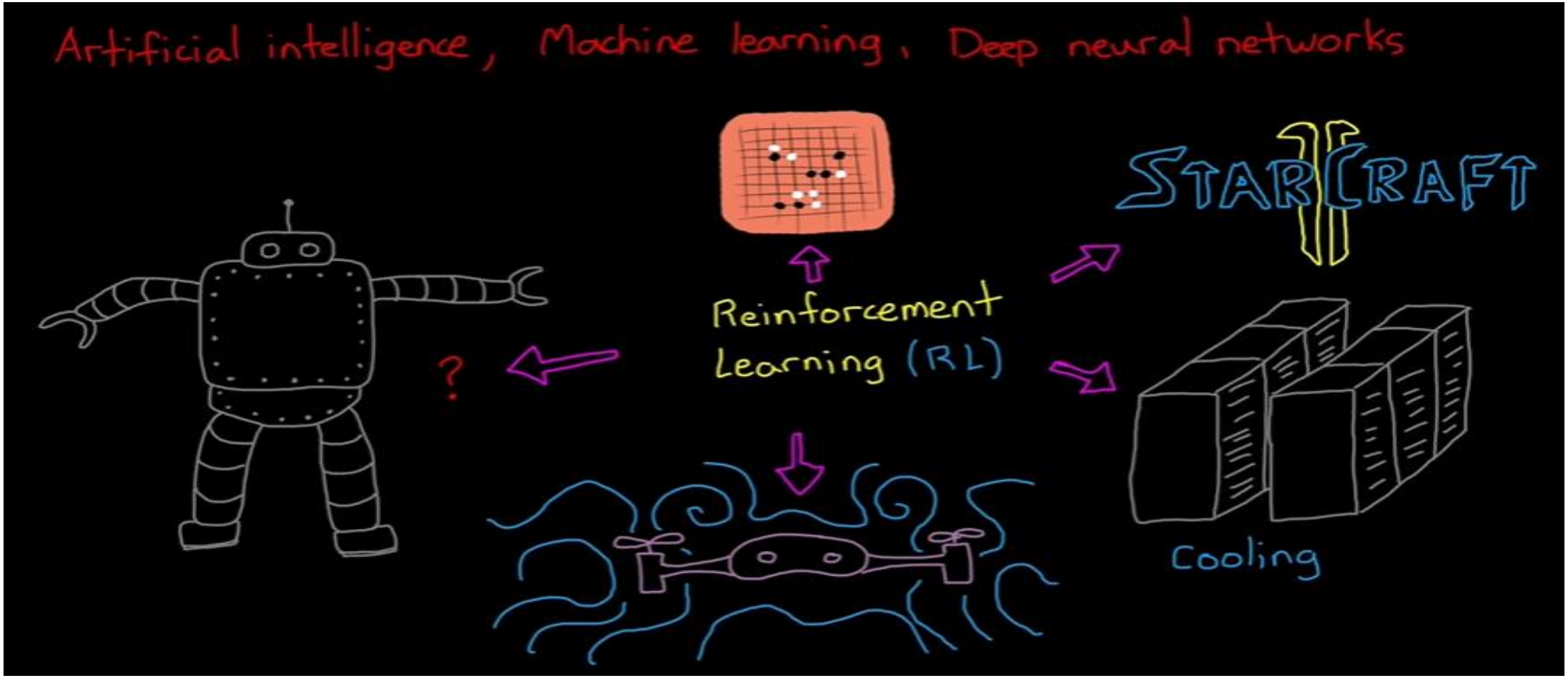
Reinforcement Learning

Reinforcement Learning: Origin

- RL came into existence due to hard and soft control problems in dynamic environment.

➢ For control the robots, cooling systems, best player for alpha go games, stabilized high dimensional drone and star craft.



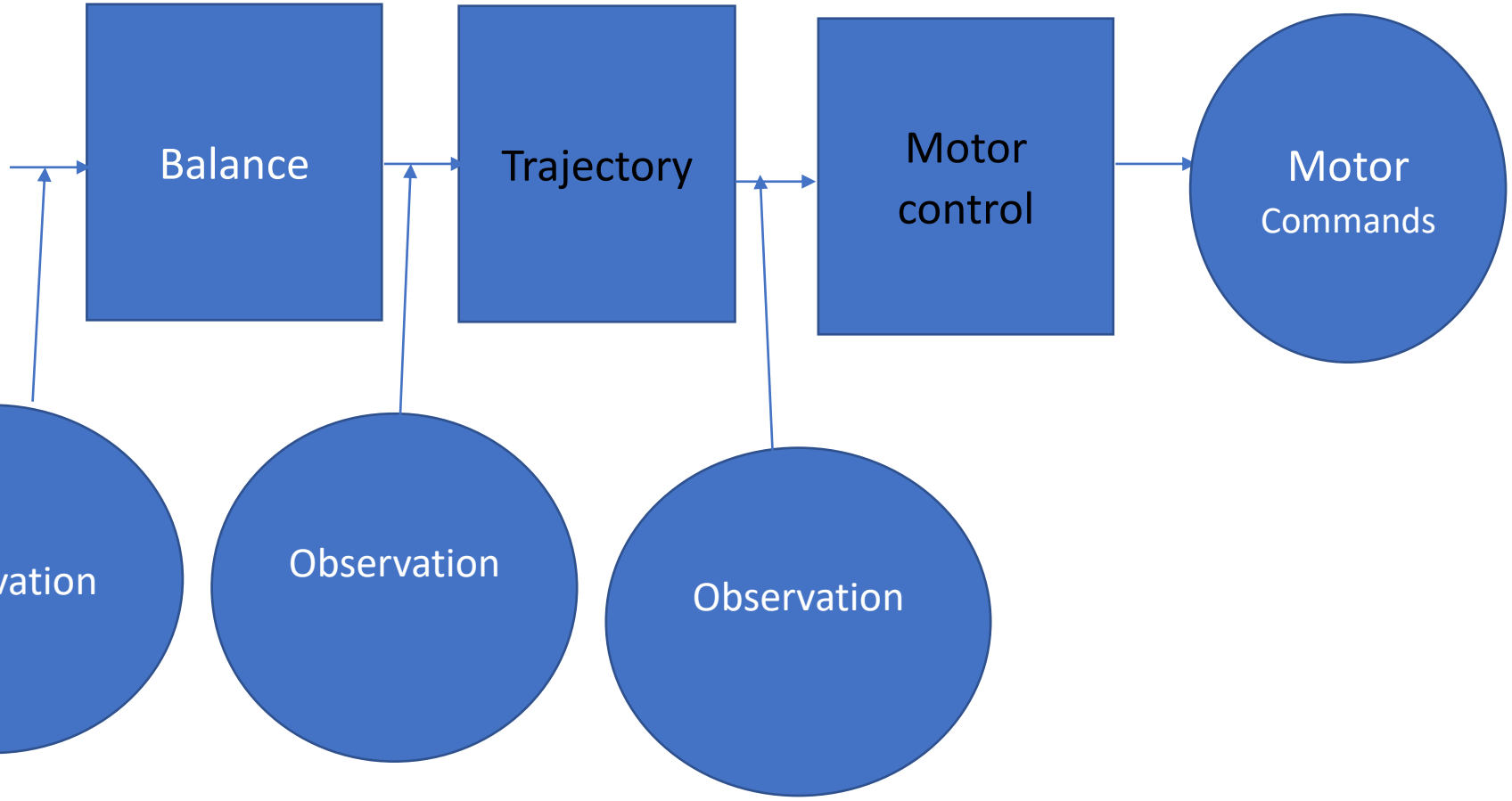
Reinforcement Learning: Introduction (based on control theorem using state (observations) and actions)

- Reinforcement learning is a type of machine learning, it has ability to soft and hard control the problems.

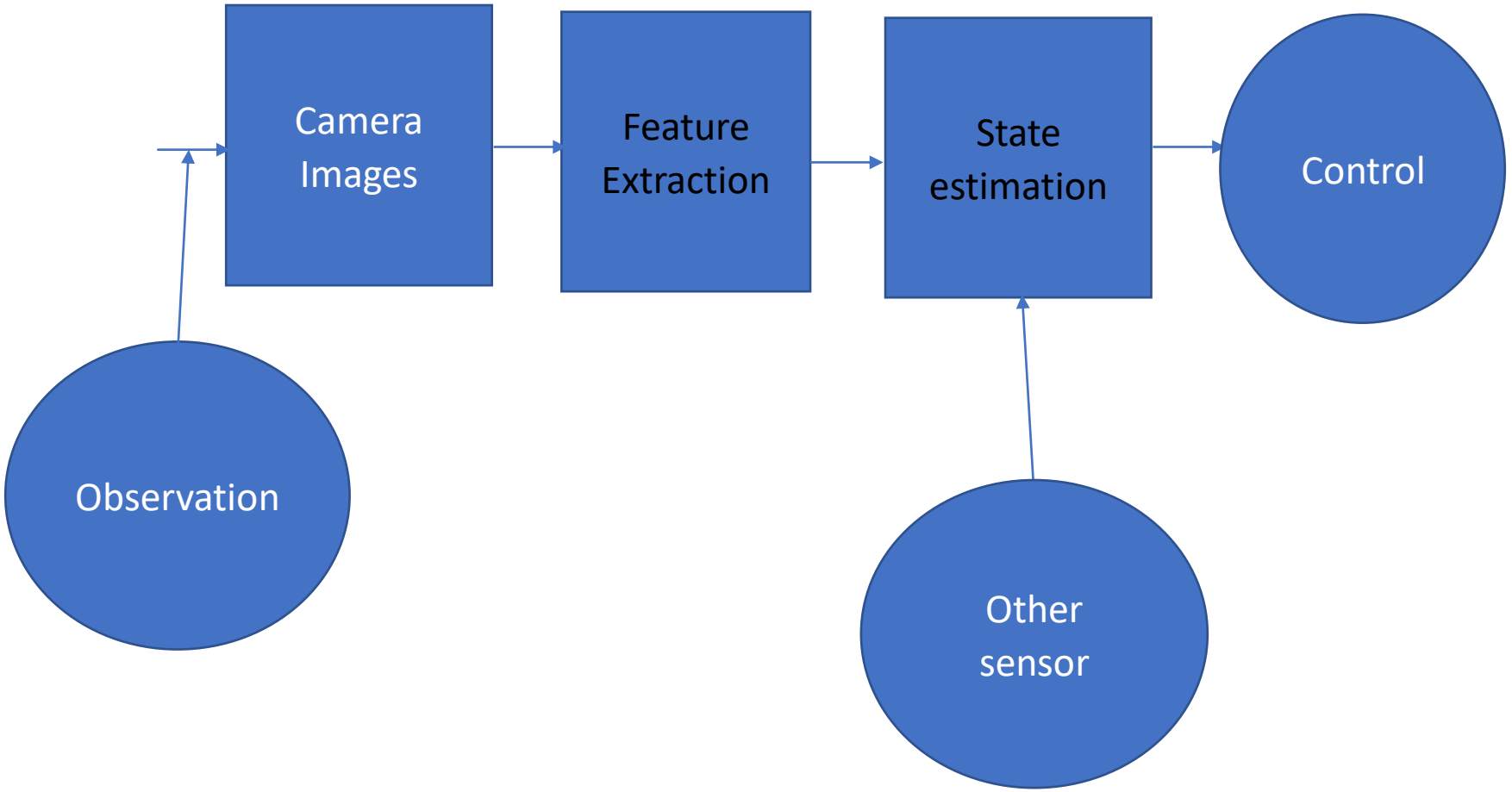
Example- **Walking Robots**

(Multiple challenges related to controlling)

-



Traditional approach



RL Agent:

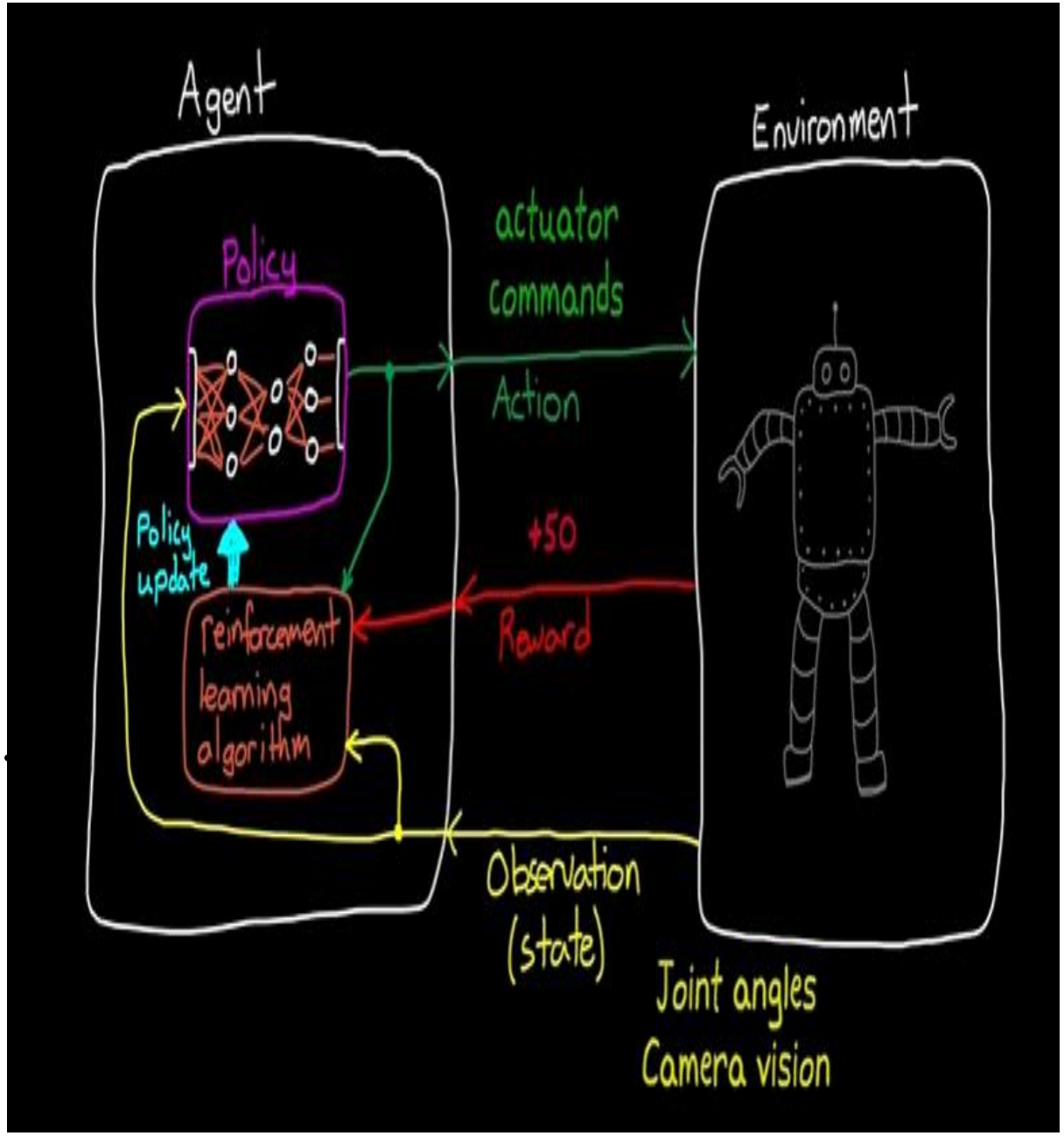
- Ability to map actions to observations.
- Policy mapped observations to actions and find best action.

➢ Reinforcement learning algorithm used to modify policy.

➢**Policy**- how to structure the logic and Parameter.

-In supervised learning used policy deep neural network for feature extraction.

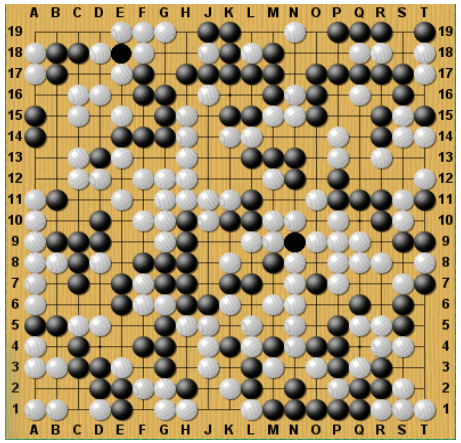
-In walking robot used policy mapped best actions.



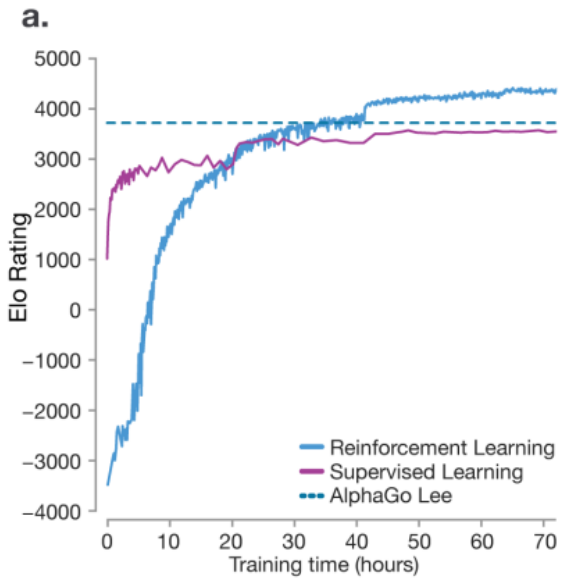
2. Alpha Go -Deep Mind: GO is classical game for artificial intelligence.

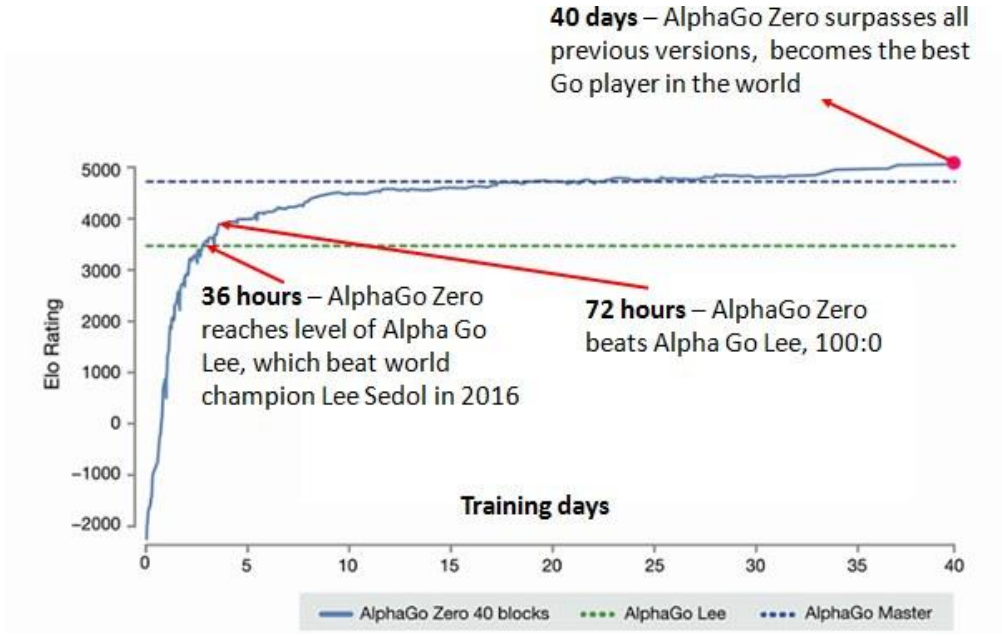
AlphaGo is the first computer program to defeat a professional human Go player.

- Learning how to beat humans at 'hard' games (search space too big)
- Far surpasses (Human) Supervised learning
- Algorithm learned to outplay humans at chess in 24 hours
- Elo rating system is a method for calculating the relative skill levels of players



State: Board State
Actions: Valid Moves
Reward: Win or Lose





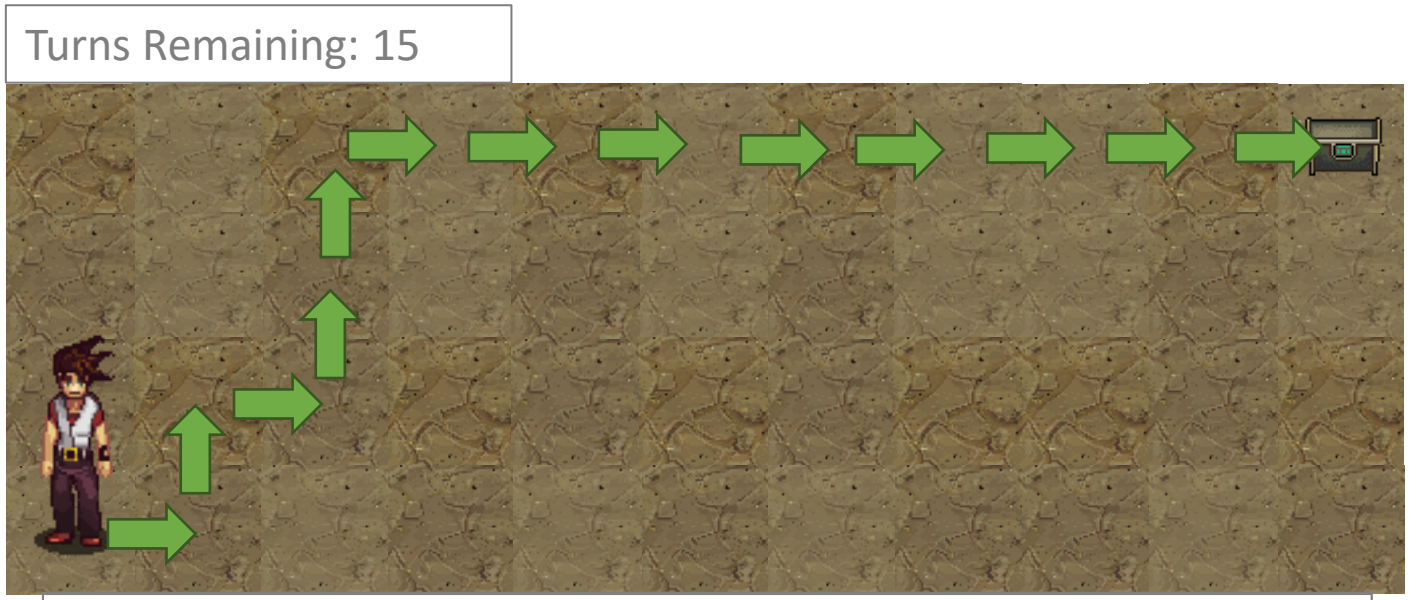
https://deepmind.com/documents/119/agz_unformatted_nature.pdf

How Reinforcement Learning is Different

- Delayed Reward
- Agent chooses training data
- Explore vs Exploit (Life long learning)
- Very different terminology (can be confusing)

Challenges for Reinforcement Learning

- When there are many states and actions
- When the episode can end without reward
- When there is a ‘narrow’ path to reward

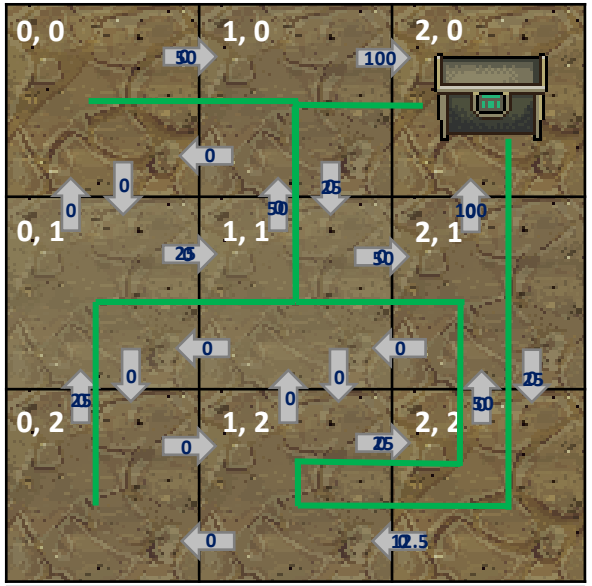


Memory

- Retrain on previous explorations
  - Maintain samples of:

$$P_a(s, s')$$
$$R_a(s, s')$$

- Useful when
  - It is cheaper to use some RAM/CPU than to run more simulations
  - It is hard to get to reward so you want to leverage it for as much as possible when it happens



Some Problems with Q-Learning

- State space is continuous
  - Must approximate $\hat{Q}$ by discretizing

- Treats states as identities
  - No knowledge of how states relate
  - Requires many iterations to fill in $\hat{Q}$

- Converging $\hat{Q}$ can be difficult with randomized transitions/rewards

RL interacting with environment –

- **Exploration-**explore new area in the environment

- **Exploitation-** collect the most reward

Application of Reinforcement Learning

- Recent use in complex research field such as environment generalization, complex chemical compound rectification, risk analysis in share markets, self decision making in railways, banking, airports, and others.
- Secondly, specially in the field of self learning for robotics research efforts.
- Thirdly, its future scope such as the helping to generalize the environment for independent learning, robotics in the form of fire fighter and automation of different sectors.
- Nowadays, flight control, robot-army with self decision making become a national concern.