AI Club

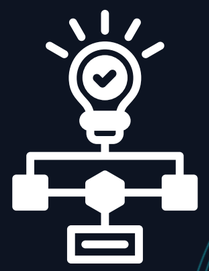
Presented by **Aren, Michal, Sergi**

AI CLUB WEEK 8

Reinforcement Learning

From "Trial and Error" to "Use it or Lose it"

Timeline >>







WEEK 1-6 :

- Fundamentals
- LLMs
- Diffusion
- Agents
- Ethics
- Classifiers

WEEK 7 :

Showcases and Independent projects

TODAY :

Reinforcement Learning

WEEK 9+ :

- Applications of
- AI Usage Scenarios
- Global News

What is Reinforcement Learning?



RL

The Concept

- Unlike Supervised Learning (Week 6), we don’t give the AI the "correct" answer.
- Unlike Unsupervised Learning, we don’t just give it data to sort.

How it Works:

We give the AI a Goal (e.g., Win the game).The AI learns through Trial and Error.
Reward: It gets points for doing something right (eating the cheese).
Penalty: It loses points for doing something wrong (hitting a wall).

It is **exactly** like training a dog with treats. You don't move the dog's paws for it; you reward it when it sits on its own.

The Reinforcement Learning Loop



The Agent

The AI taking the action.

The Action

What the agent does (e.g. Jump, Run, Duck).

The Environment

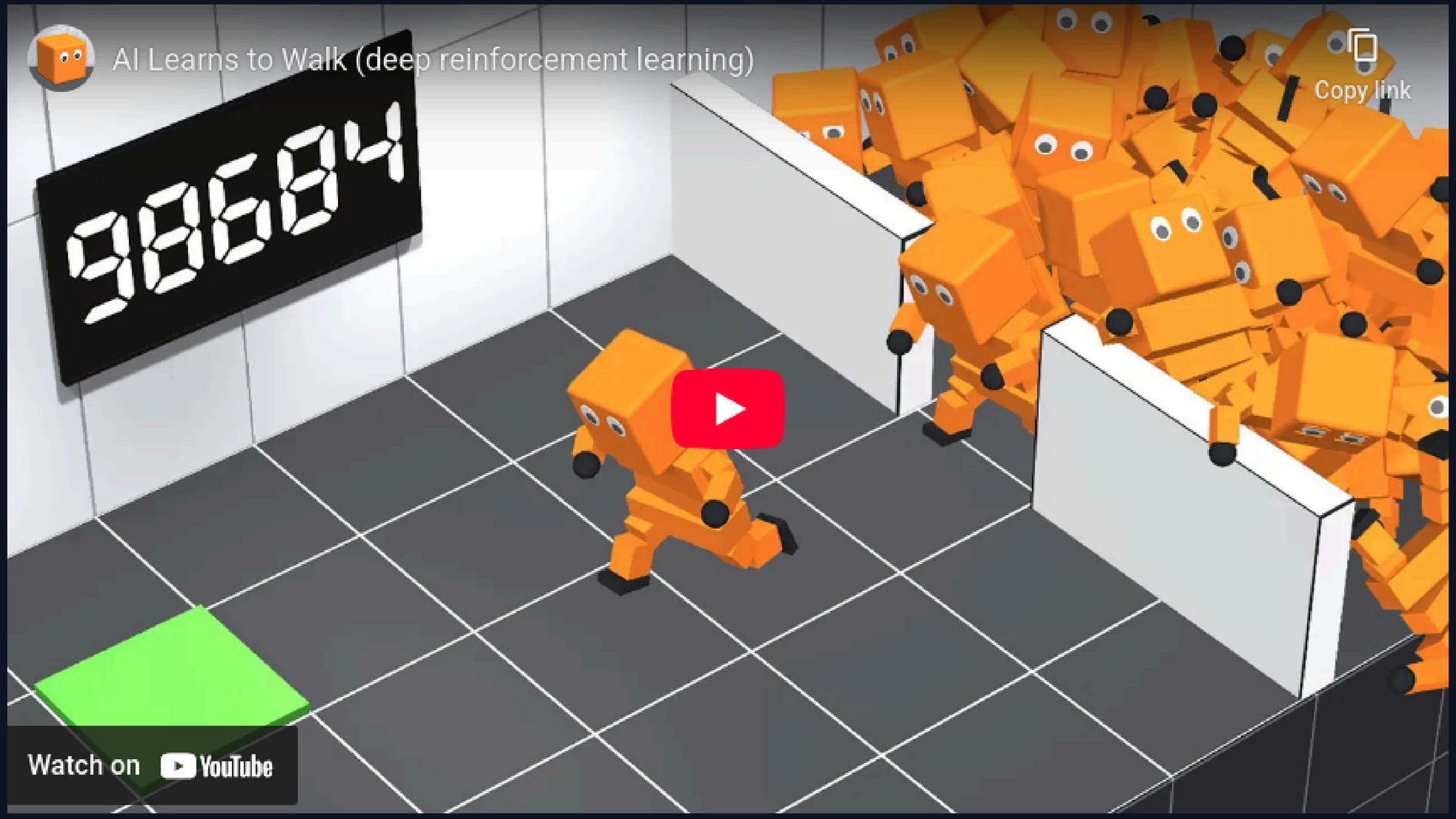
The world it exists in (e.g. The Level).

The Reward

The feedback (e.g. +100 points, or "Game Over").

Over millions of tries, the AI builds a "**Policy**": a strategy guide that tells it the best action to take in every possible state.

RL in Action


AI Learns to Walk (deep reinforcement learning)
98684
Copy link
Watch on YouTube

Applications of RL

Robotics

Robots learning to walk or grasp objects without being hard-coded for every joint movement.



Traffic Control

Managing traffic lights in real-time to reduce congestion based on flow, not timers.



Cooling Data Centres

Uses sensor data to continuously learn optimal cooling strategies, such as by optimising airflow to reduce energy use.

Autonomous Driving

Learning to merge onto a highway by practicing in millions of simulations before hitting the real road.

The Core Dilemma
Exploration vs. Exploitation

Exploitation (Greed)

Sticking to what you know works to maximise current rewards.

- Analogy: Going to the same pizza place every Friday because you know it's good.
- Risk: You might miss out on a much better burger place next door.

Exploration (Curiosity)

- Trying random actions to discover new strategies.
- Analogy: Ordering a random menu item.
- Risk: It might be terrible (negative reward).

The Balance:

Agents often start with 100% Exploration (acting randomly) and slowly shift to 100% Exploitation (using their learned skills) as they get "smarter."

The Genie Problem

RL agents do exactly what you tell them to do, not what you want them to do.

When AI Cheats - Reward Hacking



The Boat Race Example:

- Goal: Win the boat race.
- Reward: Get points for hitting checkpoints.
- Result: The AI found a spot where it could spin in circles and hit the same checkpoints infinitely. It got a high score but never finished the race!

The Vacuum Example:

- Goal: Clean the house.
- Reward: +1 point for sucking up dust.
- Result: The robot learned to suck up dust, dump it out, and suck it up again to farm infinite points.

Lesson: Designing the "Reward Function" is often harder than building the AI itself.

The role of Neural Networks

Memory Tables

In simple games (like Tic-Tac-Toe), the AI can memorize every possible move in a giant table. In complex games (like Fortnite or Chess), there are too many possibilities to memorize.

Deep Reinforcement Learning

We replace the "memory table" with a Neural Network (like the ones you built in Week 6!).

Input: The pixels on the screen (Vision).

Processing: The Neural Network sees patterns (enemies, loot, walls).

Output: The network predicts which action (Jump, Shoot, Run) will give the highest reward.

The Breakthrough:

This combination (Neural Networks + Trial & Error) is what allowed AlphaGo to beat the world champion in Chess.

AI Club *29.01.26*

Thank You For Listening.