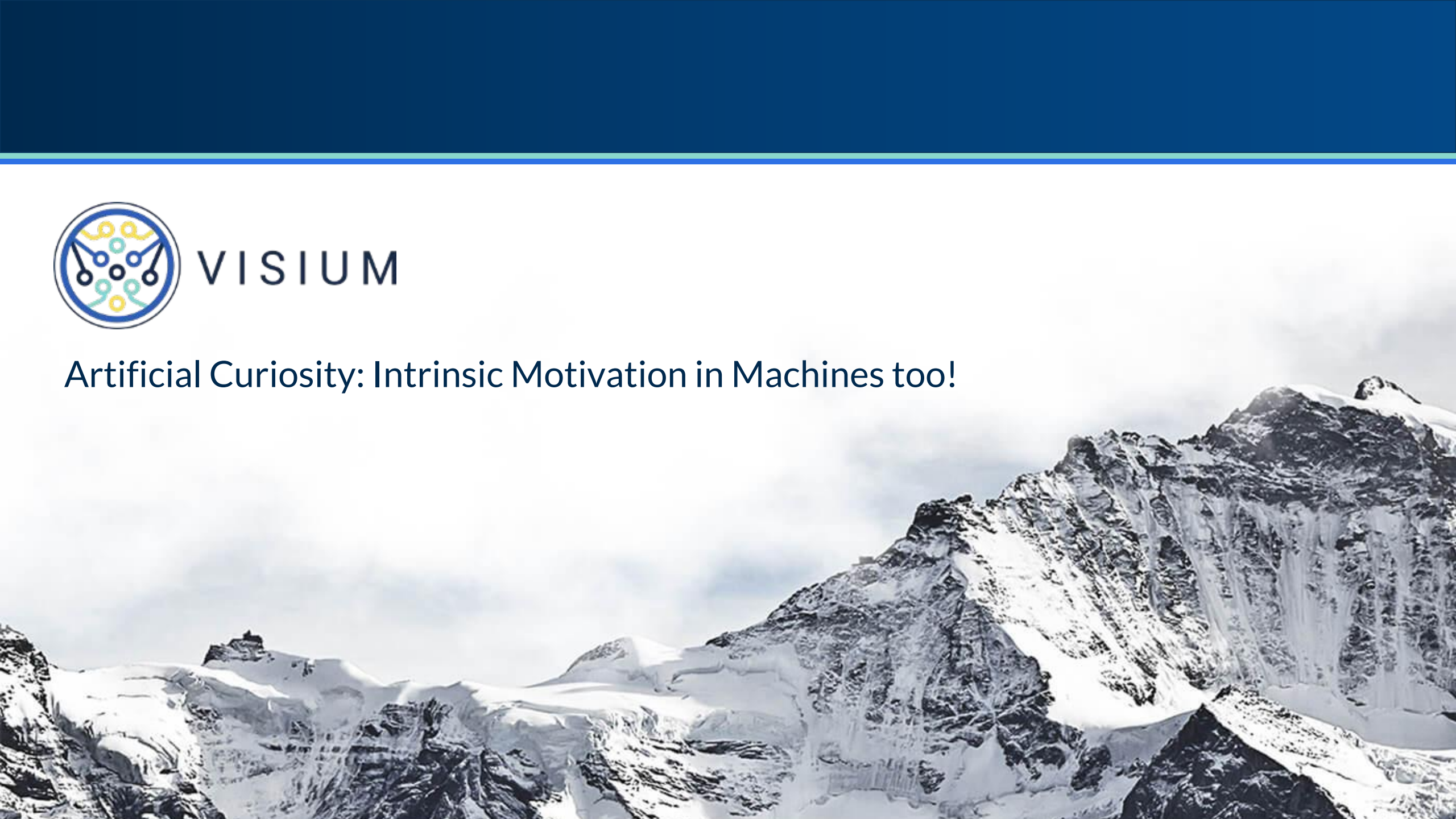


Artificial Curiosity: Intrinsic Motivation in Machines too!

Welcome!

Artificial Curiosity: Intrinsic Motivation in Machines too!

https://github.com/thibaultcalvayrac/AMLD_artificial_curiosity

Follow the installation instructions to run the code on your laptop

Today's program:

1. What is reinforcement learning?
2. Let's build an intelligent agent!
3. Let's make it curious!

Welcome!

Machine Learning Engineer

Gaetan

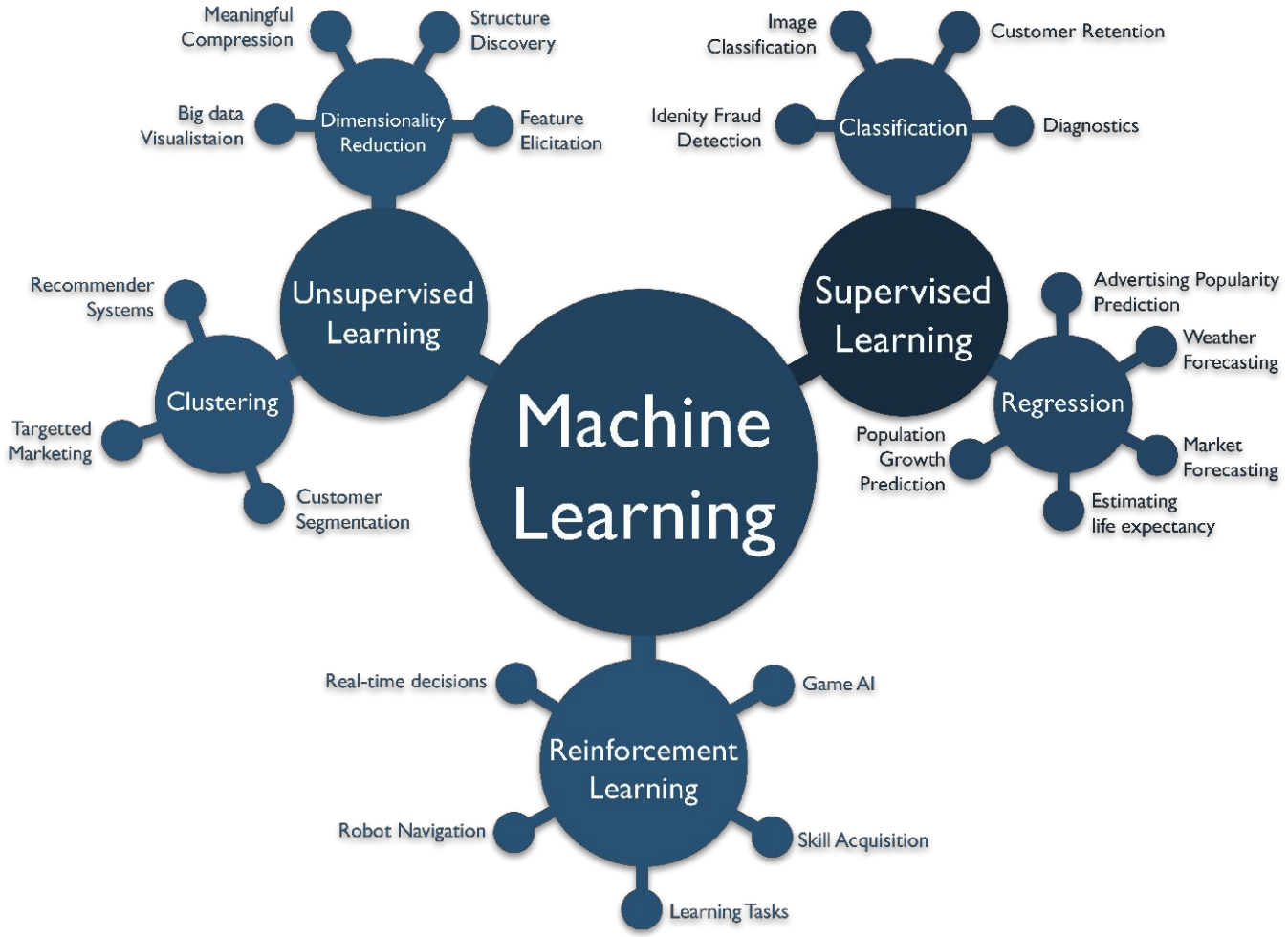
Solution Architect

Axel

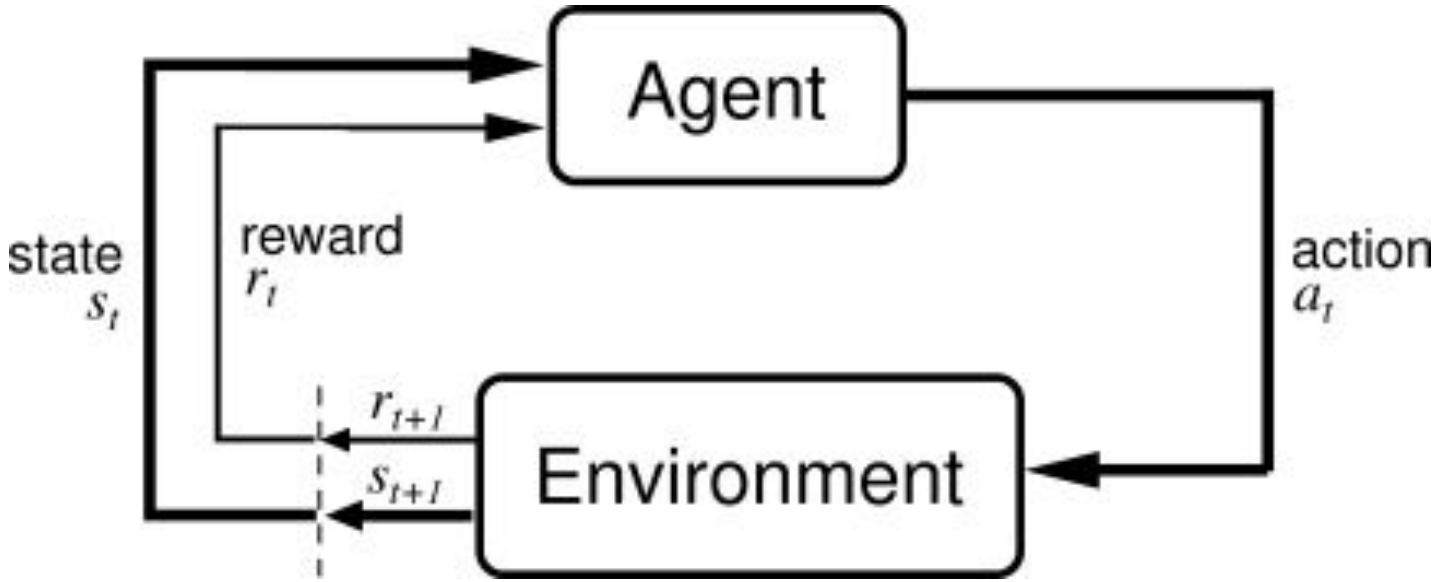
Machine Learning Engineer

Thibault

Three areas of Machine Learning



Reinforcement Learning



Objective: select actions that maximize the cumulated rewards:

$$R_t = r_{t+1} + r_{t+2} + r_{t+3} + \cdots + r_T,$$

Reinforcement Learning

Episode: sequence of states observed, actions taken and rewards received

$\{ s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_3, \dots , s_{T-1}, a_{T-1}, r_T, s_T \}$

Policy: function that maps a state observation to an action to take

$\pi(s, a) = \text{Proba}[a_t=a \mid s_t=s]$

Discounted Return: sum of future rewards received after a step until the episode ends

the contribution of each reward to the sum decreases with time

$R_t = r_{t+1} + \gamma\, r_{t+2} + \gamma^2 r_{t+3} + \dots + \gamma^{T-t-1} r_T = \sum_{k=0,1,\dots,T-t-1} \gamma^k r_{t+k+1}$ with $0 \leq \gamma \leq 1$

Value Function: expected discounted return from a state when following a given policy

$V^{\pi}(s) = E_{\pi}\{R_t \mid s_t=s\}$

It’s a me, Mario!

More and more complex environments

Not only in games!

Not only in games!

- Inducing app crashes for debugging
- Resources management
- Traffic light control
- Portfolio optimisation
- Warehouse stock optimisation
- & many more...

Exploration vs Exploitation tradeoff

Systems that can take actions can adversely affect future data

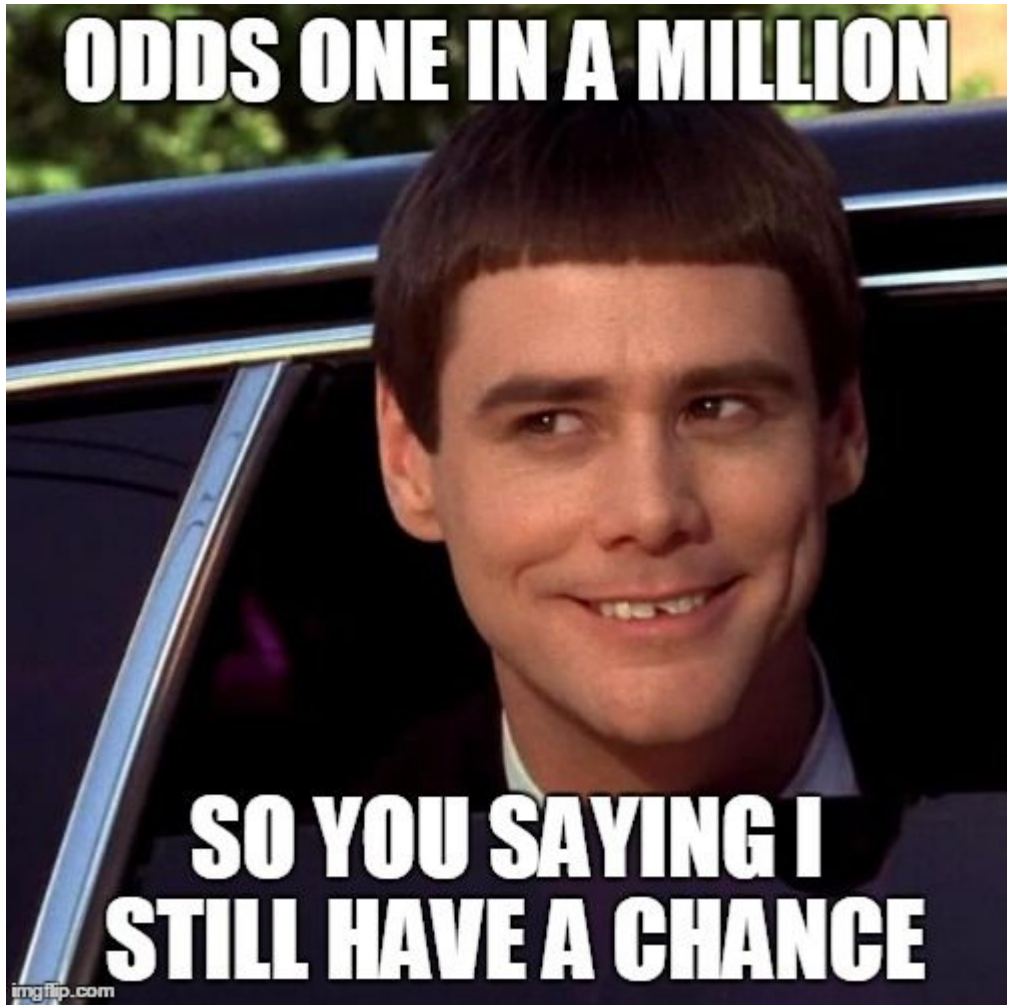
Learn more about what is good and bad

Make the best use of what we believe is good

The problems of exploitation

The problems of exploitation

Sparse reward problem



Reinforcement Learning

Episode: sequence of states observed, actions taken and rewards received

$\{ s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_3, \dots , s_{T-1}, a_{T-1}, r_T, s_T \}$

Policy: function that maps a state observation to an action to take

$\pi(s, a) = \text{Proba}[a_t=a \mid s_t=s]$

Discounted Return: sum of future rewards received after a step until the episode ends

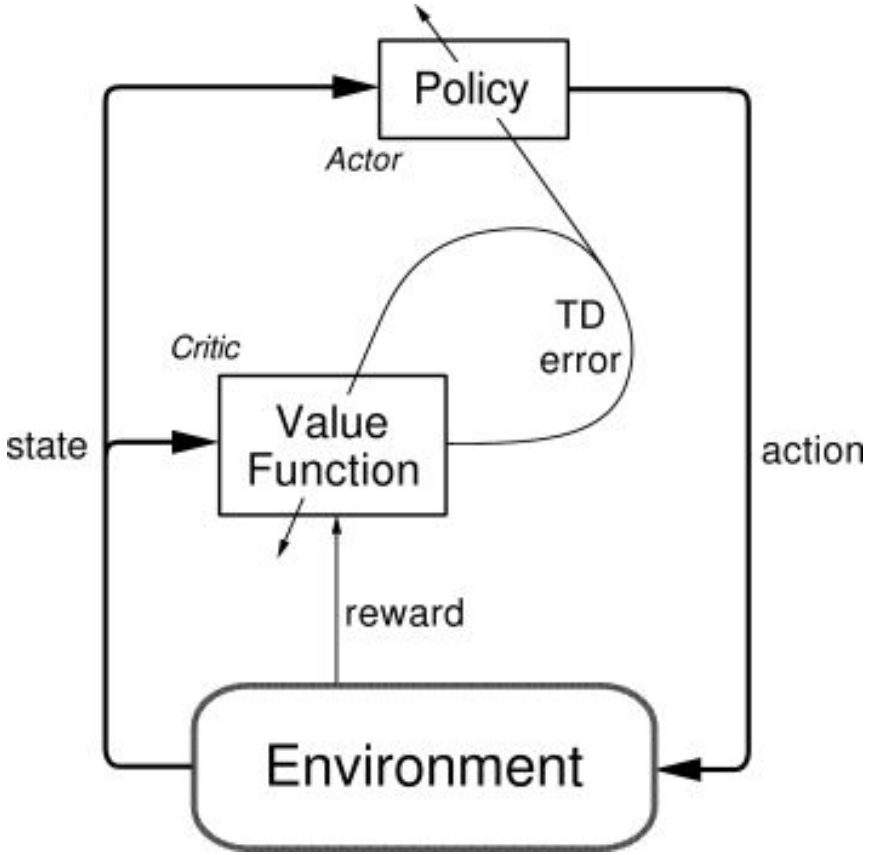
the contribution of each reward to the sum decreases with time

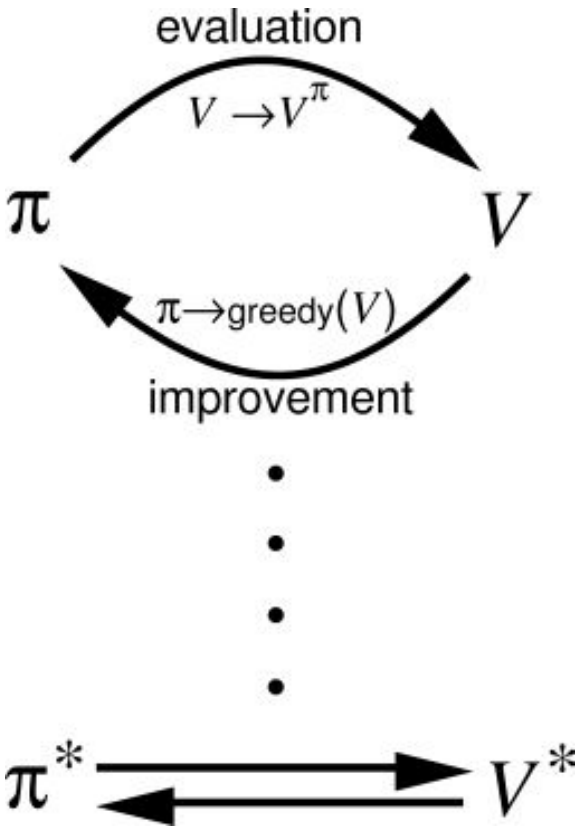
$R_t = r_{t+1} + \gamma\, r_{t+2} + \gamma^2 r_{t+3} + \dots + \gamma^{T-t-1} r_T = \sum_{k=0,1,\dots,T-t-1} \gamma^k r_{t+k+1}$ with $0 \leq \gamma \leq 1$

Value Function: expected discounted return from a state when following a given policy

$V^{\pi}(s) = E_{\pi}\{R_t \mid s_t=s\}$

Actor Critic Architecture

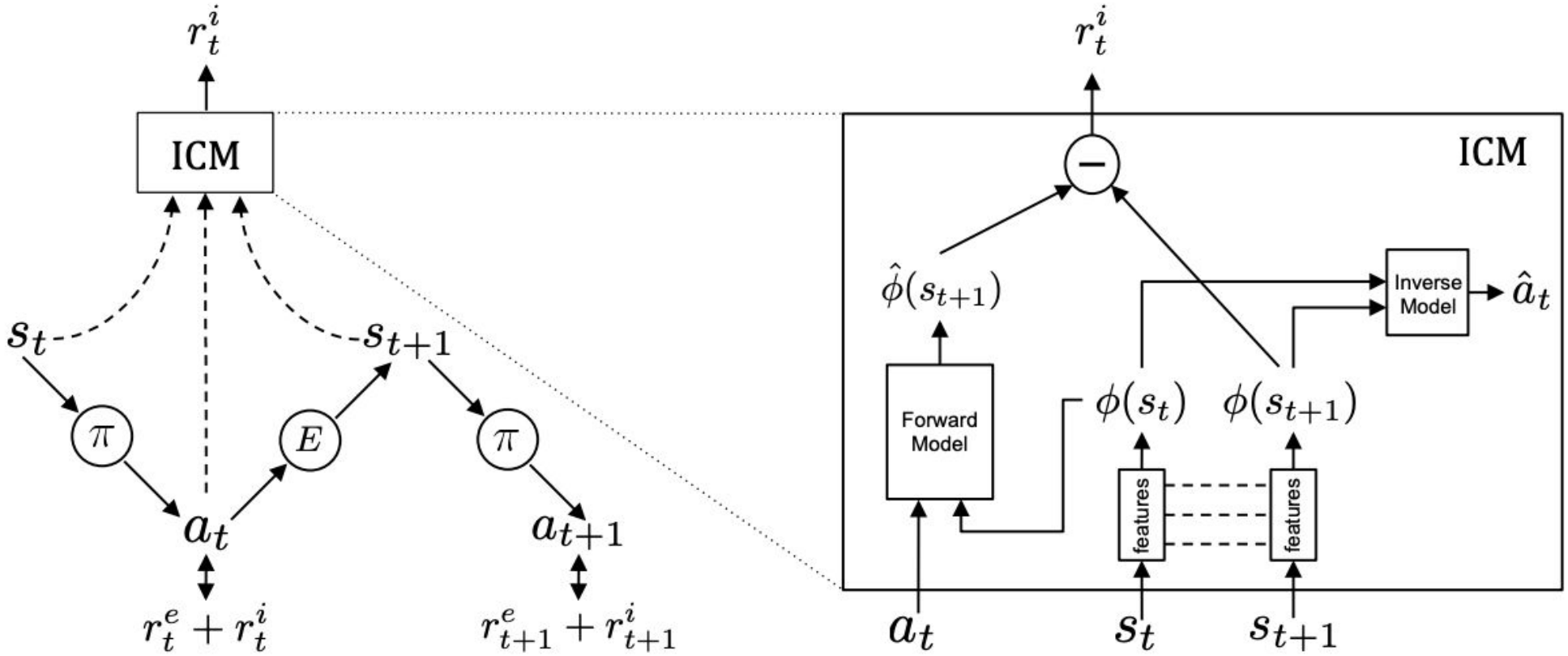




Extrinsic vs Intrinsic Motivation

Surprise as a proxy for Curiosity

Curious Reinforcement Learning



Pathak et al., “Curiosity-Driven Exploration by Self-Supervised Prediction” (2017)

Discussion time!

- What are the benefits of this approach?
- What are the problems of this approach?

  Hint: could a curious agent “hack” the intrinsic reward?

- How would you solve these problems?

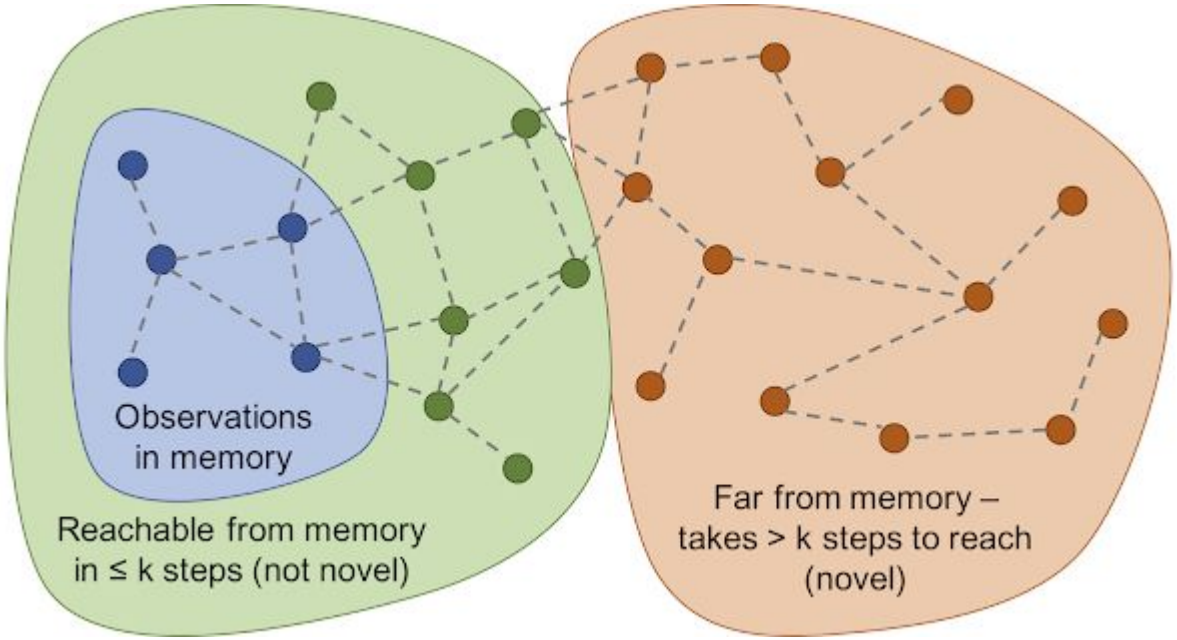
Curious Reinforcement Learning

- No need to implement a reward
- Use a non-sparse reward function
- Learn the exploration strategy automatically
- Learn a model of the environment
- Transfer to other environments

Couch Potato Problem

Burda et al., “Large Scale Study of Curiosity-Driven Learning” (ICLR 2019)

An Active Research Question



Savinov et al., “Episodic Curiosity through Reachability” (2018)

Take Aways

- Reinforcement Learning has a certain aspect of *creativity* and can beat humans by finding new optimal paths
- **But** it requires rewards and a reward function
- Curious RL can be a solution to this problem
- But beware of the couch potato problem…

Thank you!

Visium Technologies SA

EPFL Innovation Park, Lausanne, Switzerland

www.visium.ch

+41 21 691 55 30

au@visium.ch

VISIUM