Deep Reinforcement Learning

1 - Introduction

Prof. Dr. Malte Schilling

Autonomous Intelligent Systems Group

Overview today

1. Lecture information.
2. Overview Machine Learning – towards Reinforcement Learning.
3. Topics of the Lecture.

General Information

The Module consists of

- a lecture,
- exercises,
  - explanation of solution (50 % of problems)
- and an exam
  - brief oral exam

The Lecture

- Takes place on Tuesday, 10:15 AM to 11:45 AM, in M5.
- Presupposes preliminary knowledge in mathematics (linear algebra, probabilities, (multi-dimensional) differentiation).
- Slides for the lecture will be made available (as html repository and PDF).

Exercises

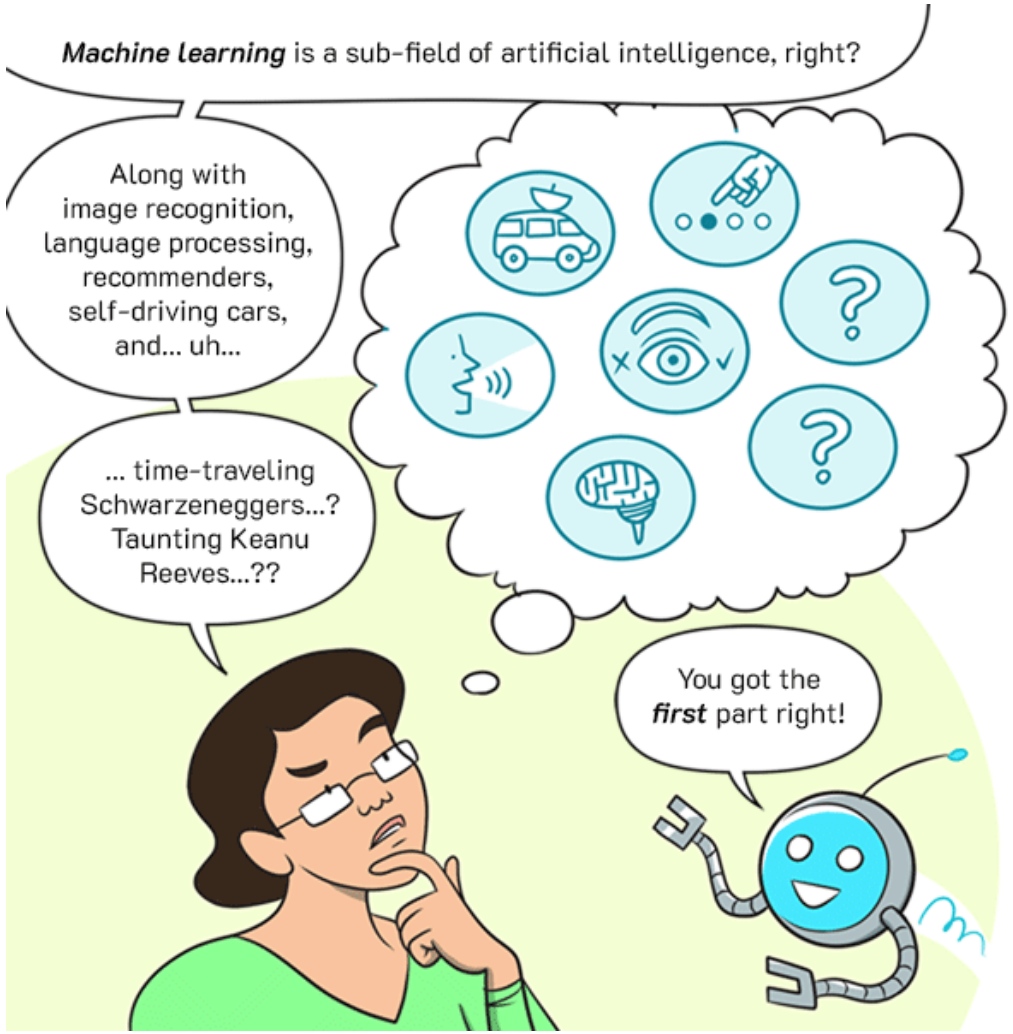
- Exercise slot is Friday morning, 10:15 AM to 11:45, M5.
- Solutions (for programming exercises in python) should be submitted by small teams of two or three persons. Everybody has to be able to present the solution during meetings.
- There will be a new exercise sheet every two weeks. Discussion and presentation of the solution will be during the exercise slot as well every second week.
- For the weeks in between: will be additional time for introduction of concepts, discussion, questions …
- First exercise slot: 21.10.2021, brief introduction to python.
- Honor code: Do collaborate and discuss together, but write up and code independently. Do not show anyone else your writeup or code or post it online (unless specified).

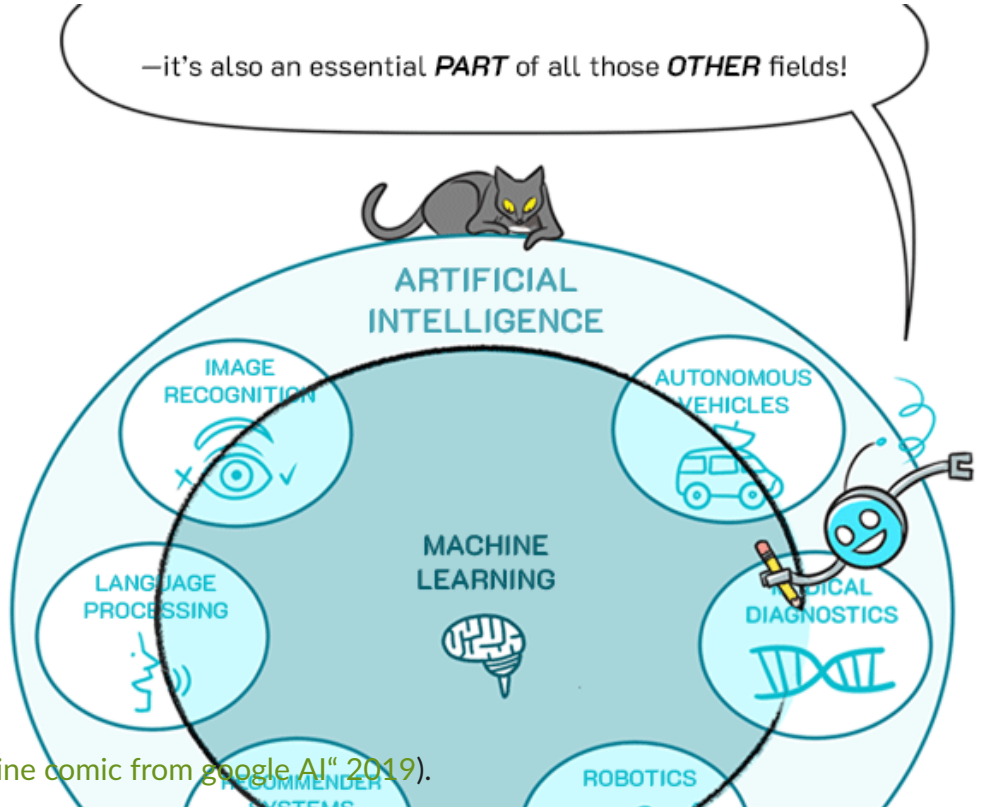
Reading List

- On Machine Learning and Artificial Intelligence in general:
  - (Bishop 2006)
  - Stanford Course on AI, CS-221 (Liang 2018)
- On Deep Neural Networks:
  - Deep Neural Networks book (Goodfellow, Bengio, und Courville 2016), online available.
- On (Deep) Reinforcement Learning:
  - Sutton and Barto, book (Sutton und Barto 2018), also online available.
  - D. Silver online course on Deep Reinforcement learning at UCL.

Machine Learning

Machine Learning is …

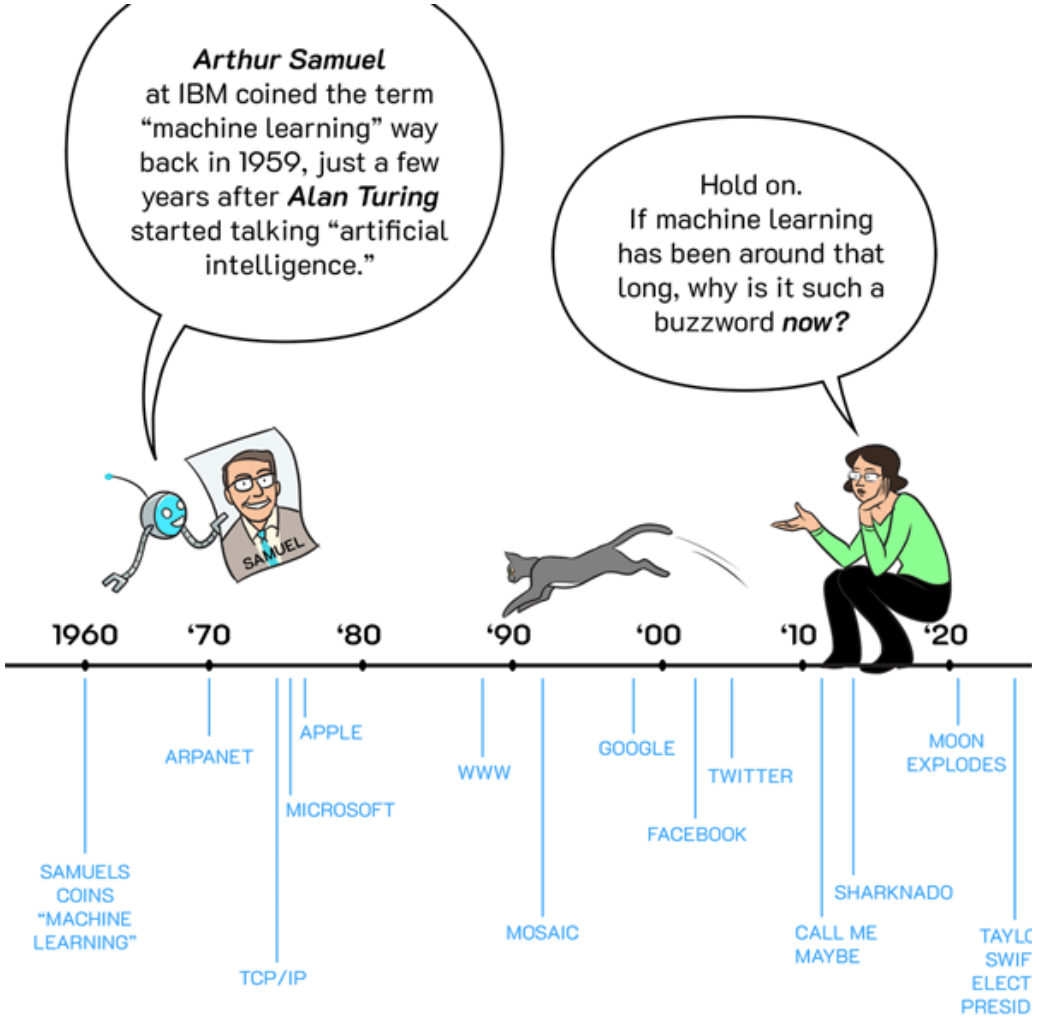




Images from („Machine Learning - an online comic from google AI“ 2019).

Artificial Intelligence

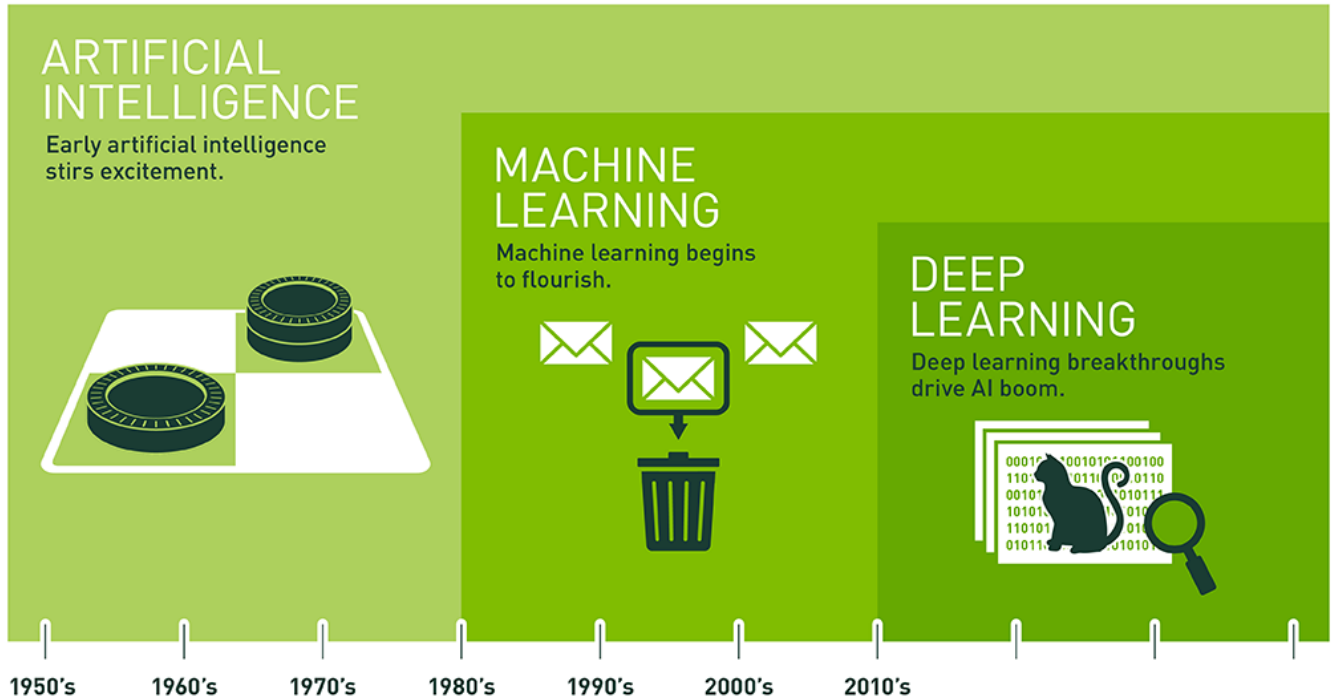
— Artificial Intelligence is the new electricity.
(Andrew Ng)



Artificial Intelligence

- Understanding human intelligence has been a research question for long times.
- Artificial intelligence has turned towards how machines could become intelligent.
- Initially, approaches focussed on specific mechanism and logic descriptions. This allowed to solve –for human– difficult problems based on formal methods and application of rules.
- It became obvious that the true challenge for Artificial Intelligence are problems which are easy to perform, but hard to formalize.

> AI is “the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable … (McCarthy, 1956)



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

from nvidia

A spectrum of intelligence

Motor intelligence

Search problems
Markov Decision Proc.
Adversarial Games

Constraint Satisfaction
Bayesian Networks

Reflex-like **State-based** **Variables, relational**

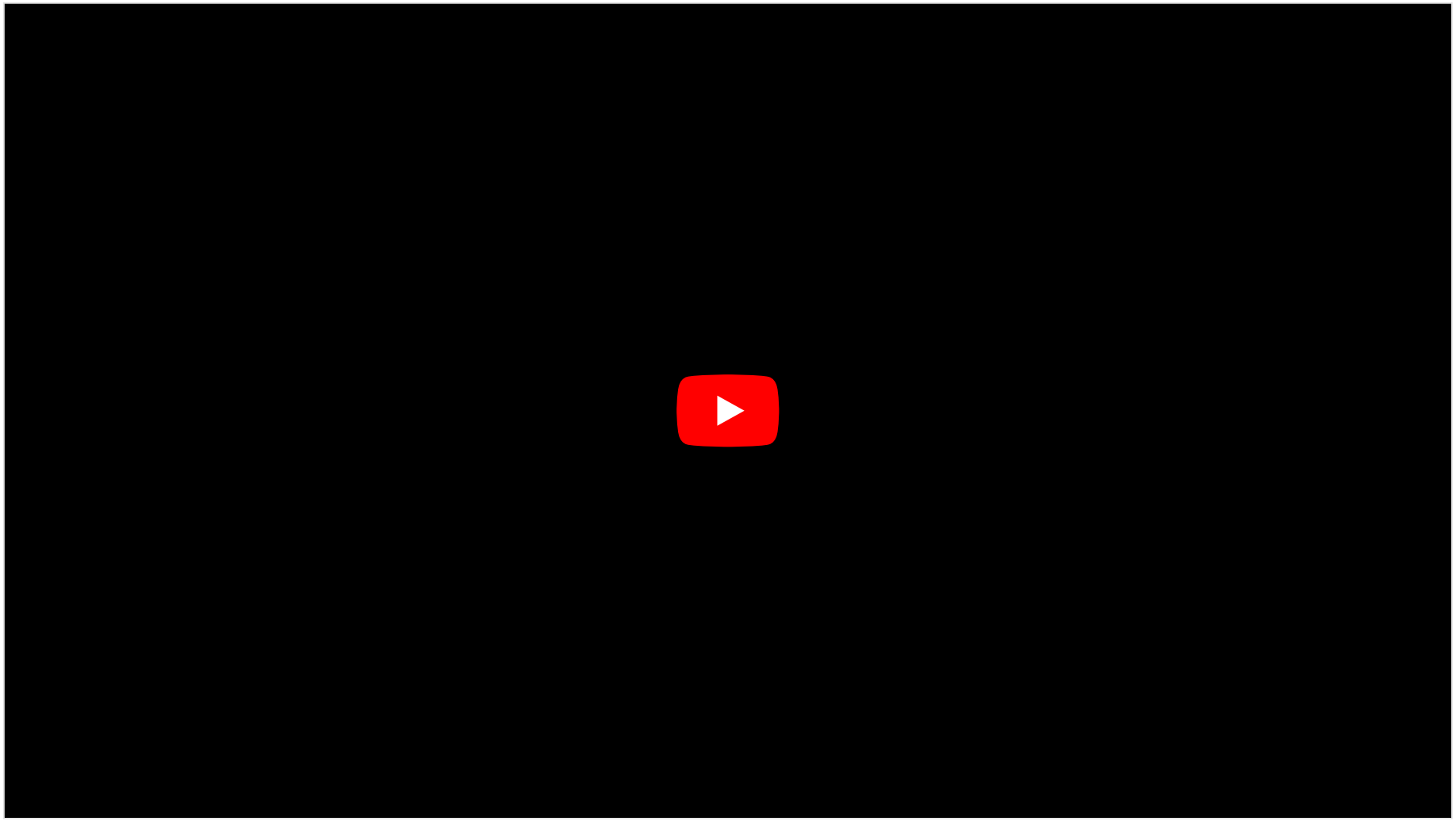
"low-level intelligence" "high-level intelligence"

Spectrum of different "levels" of intelligence – for Machine Learning (following (Liang 2018)).

Promise of Learning

- Many seemingly simple tasks are hard to describe in a formal way.
- Nonetheless, humans are quite good to solve such problems and manage such tasks.
- Crucial is the ability to learn which is a characteristic and prerequisite for intelligent behavior.
- Machine Learning in general aims at learning how to solve a task through training instead of relying on a formal description.

Learning from demonstration



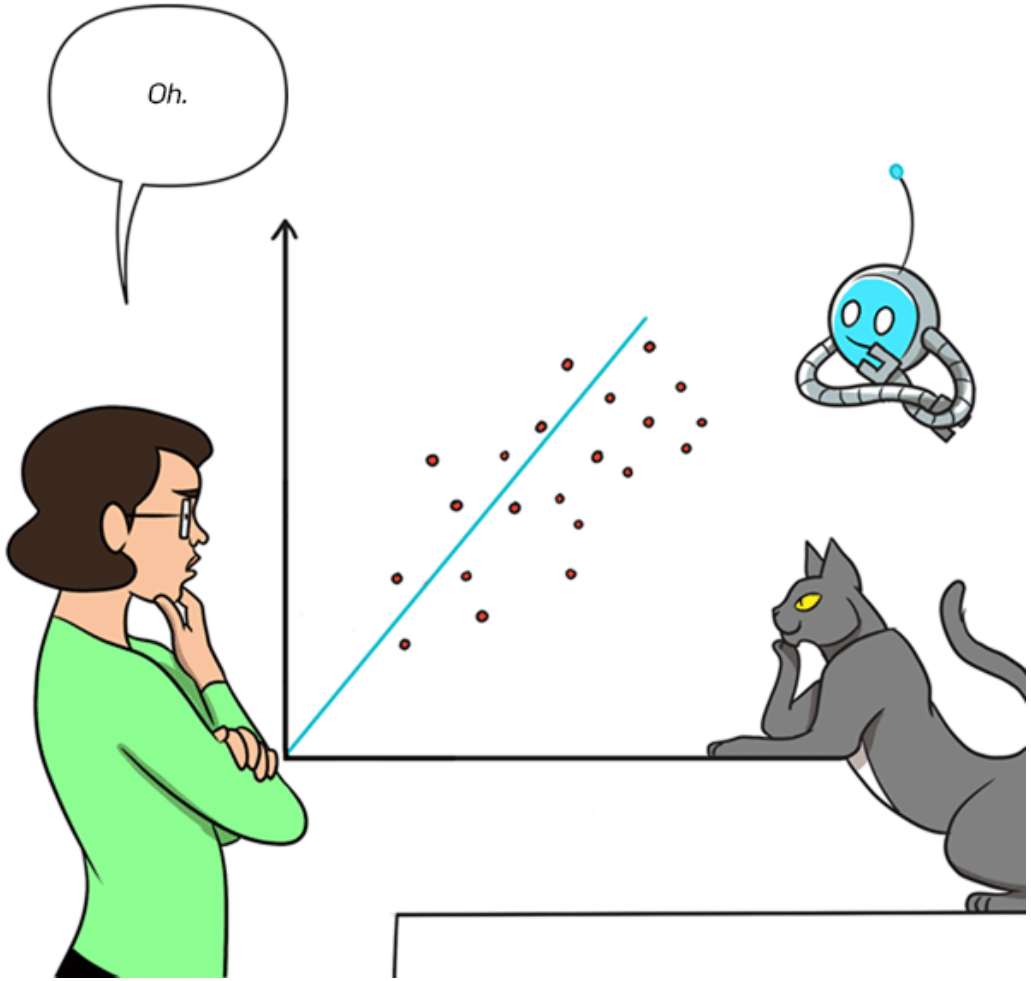
Toyota Research Institute built a household robot that users can train using a virtual reality interface. The robot learns a new behavior based on a single instance of VR guidance. Then it responds to voice commands to carry out the behavior in a variety of real-world environments (paper accepted at ICRA 2020).

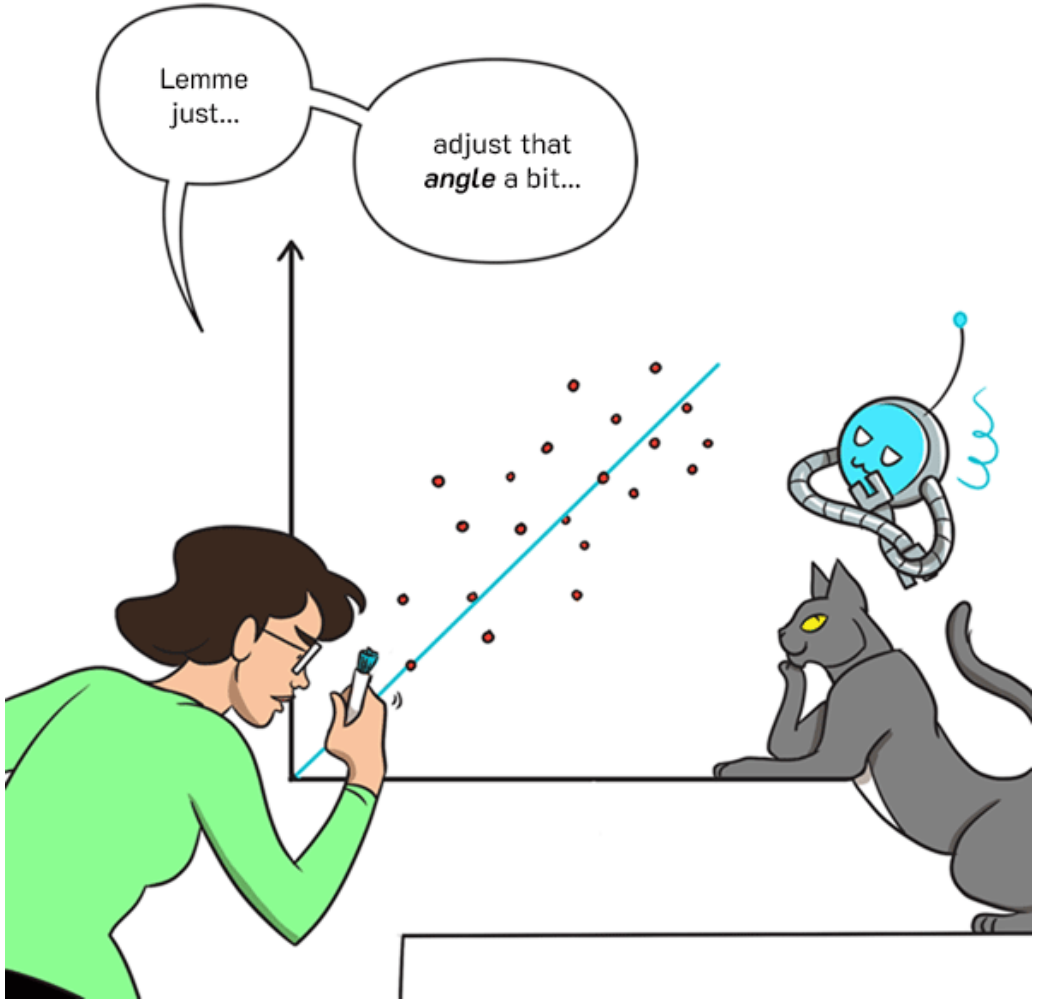
Learning is the answer to a number of important questions:

- How to enhance limited knowledge and skills?
- How to improve performance on a task?
- How to avoid prestructuring everything by hand?
- How to cope with novelty and change?
- How to get around in a world that can only partially be known?

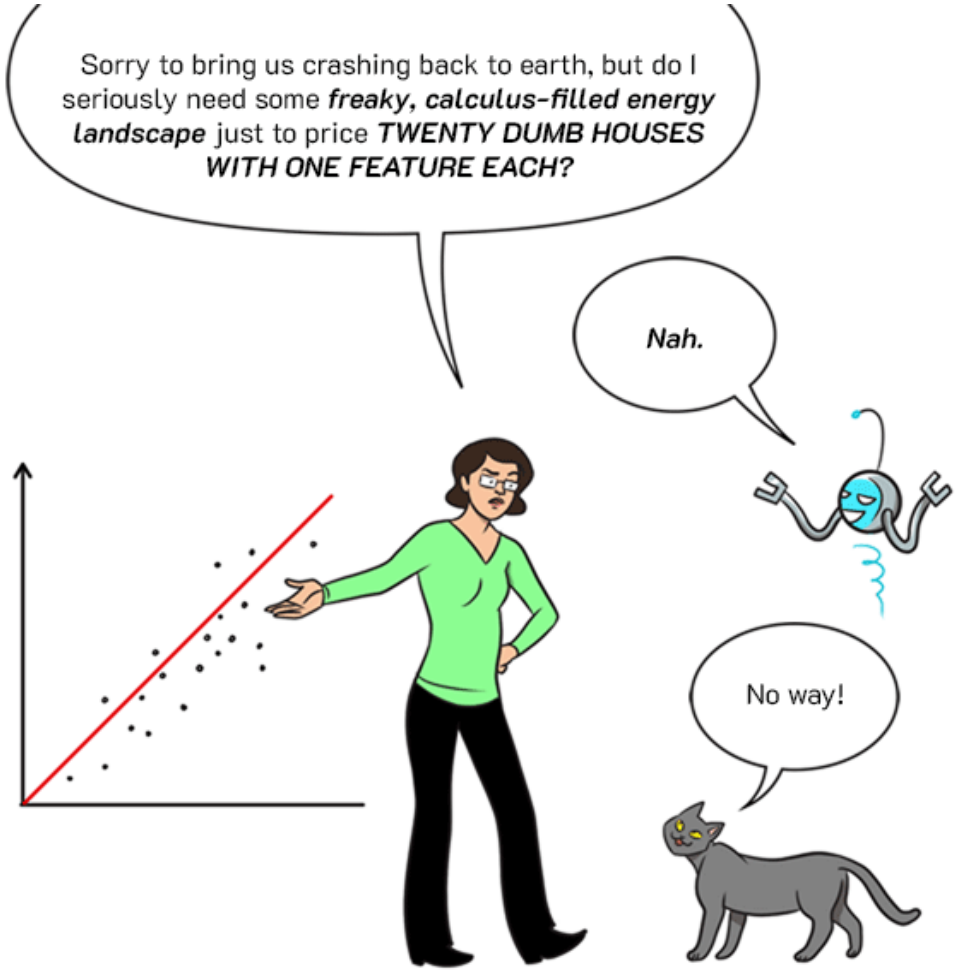
Simple example: Learning from experience

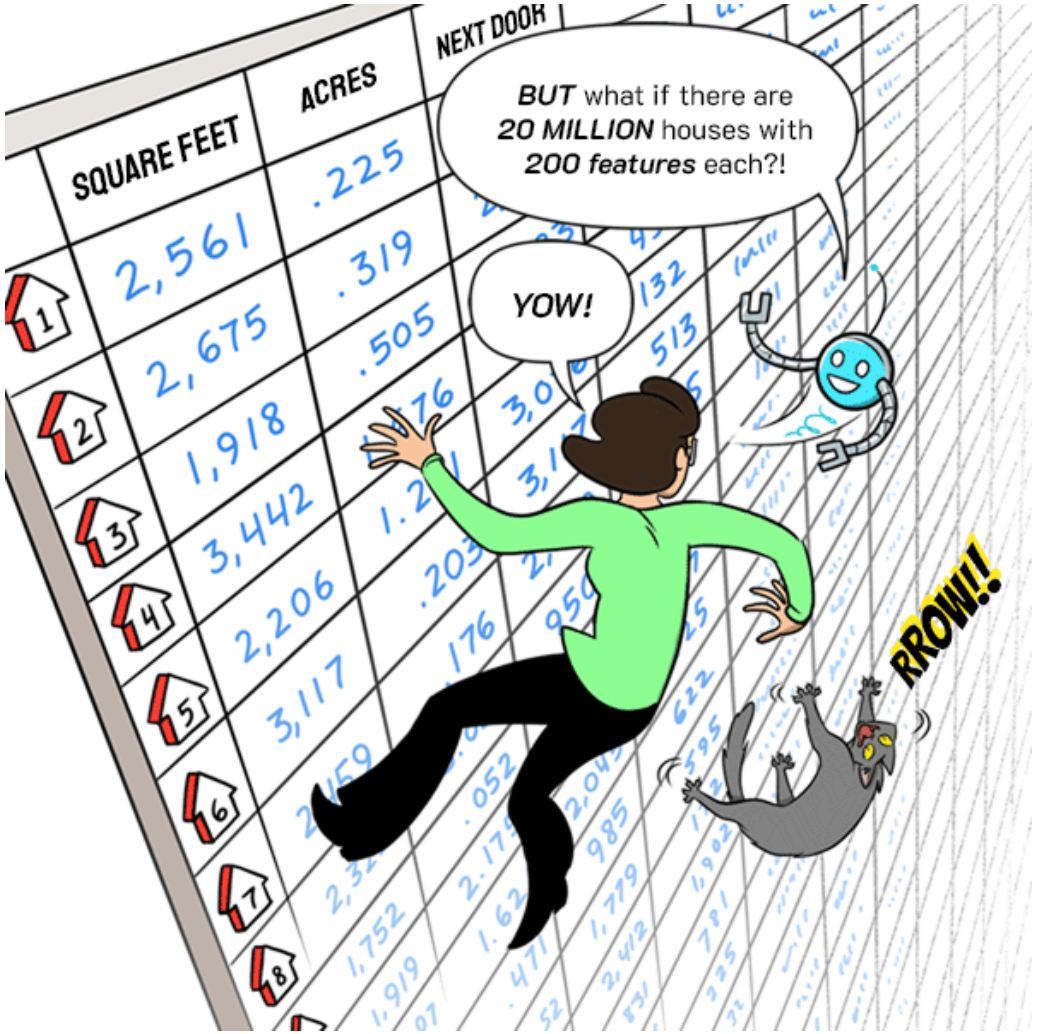
Regression as Learning





Regression for Function Approximation





Task: How to automatically learn a regression line?

Short break – discussion, solve in small groups.

Question: Can you explain how to fit such a simple linear model towards a set of given data points? Either describe an (or multiple) approaches in your own words at first. Or, if possible, try to formalize the process.

- How can we fit a regression line to a set of data points?
- Do you know other possible Machine Learning methods that we could use for this task?

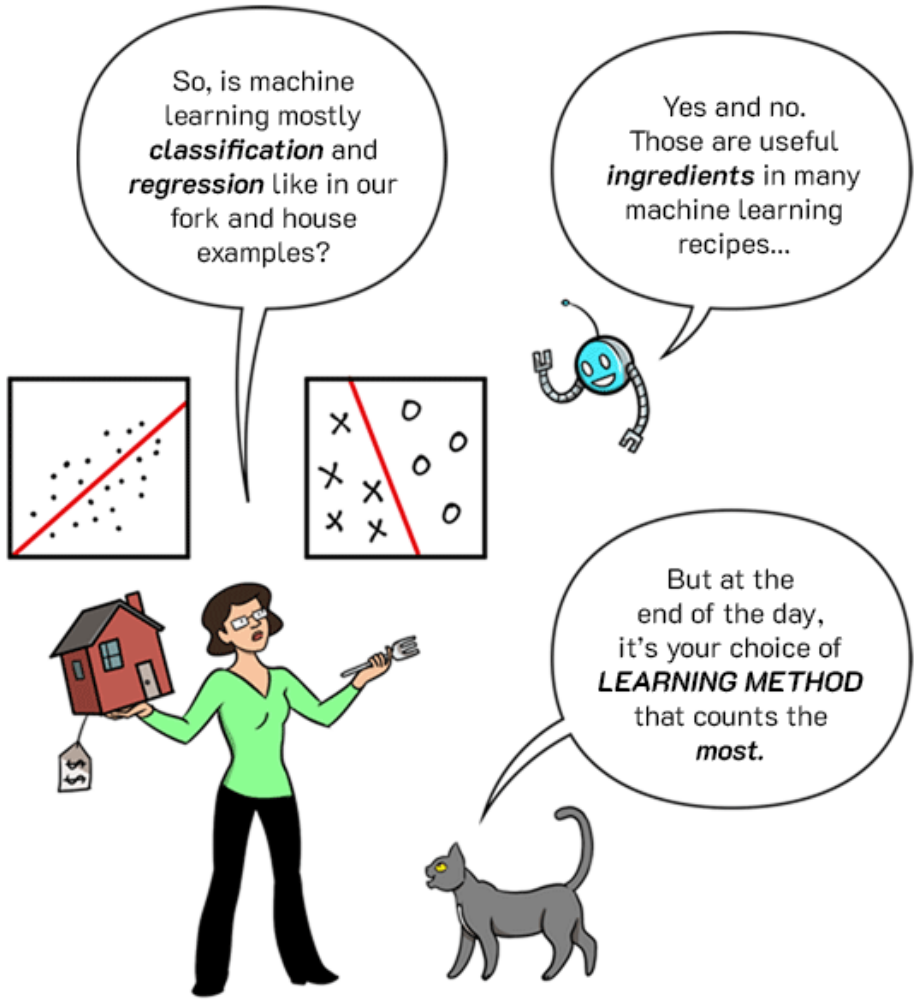
Regression Example

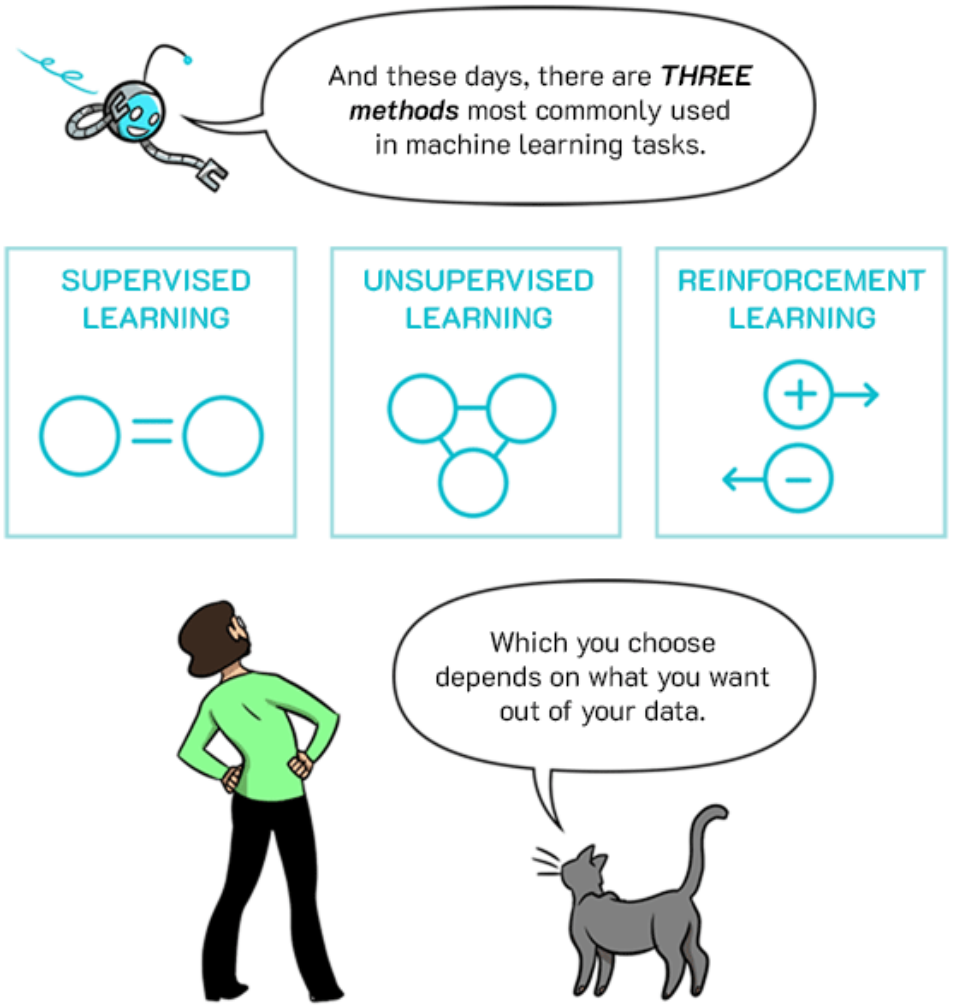
Set of data points, develop how to fit regression line.

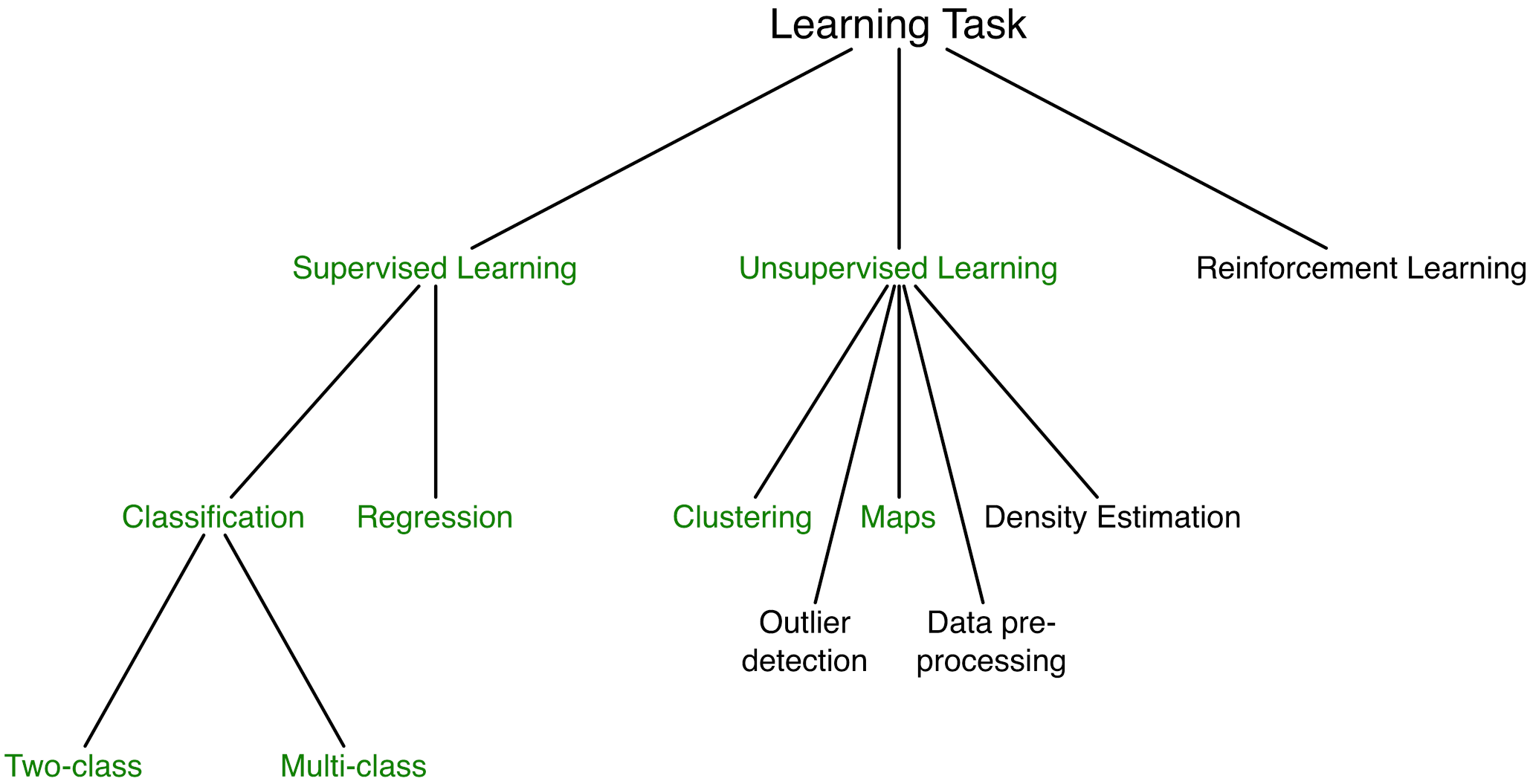
Methods of optimization:

- solve using pseudo-inverse
- gradient descent

Different forms of Learning

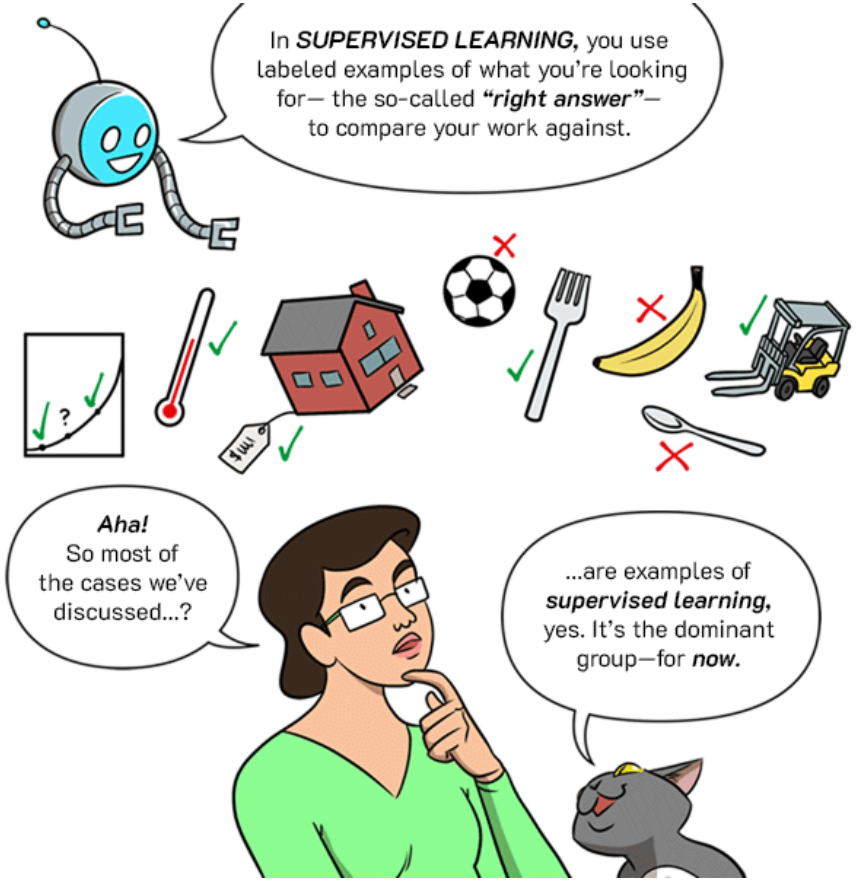


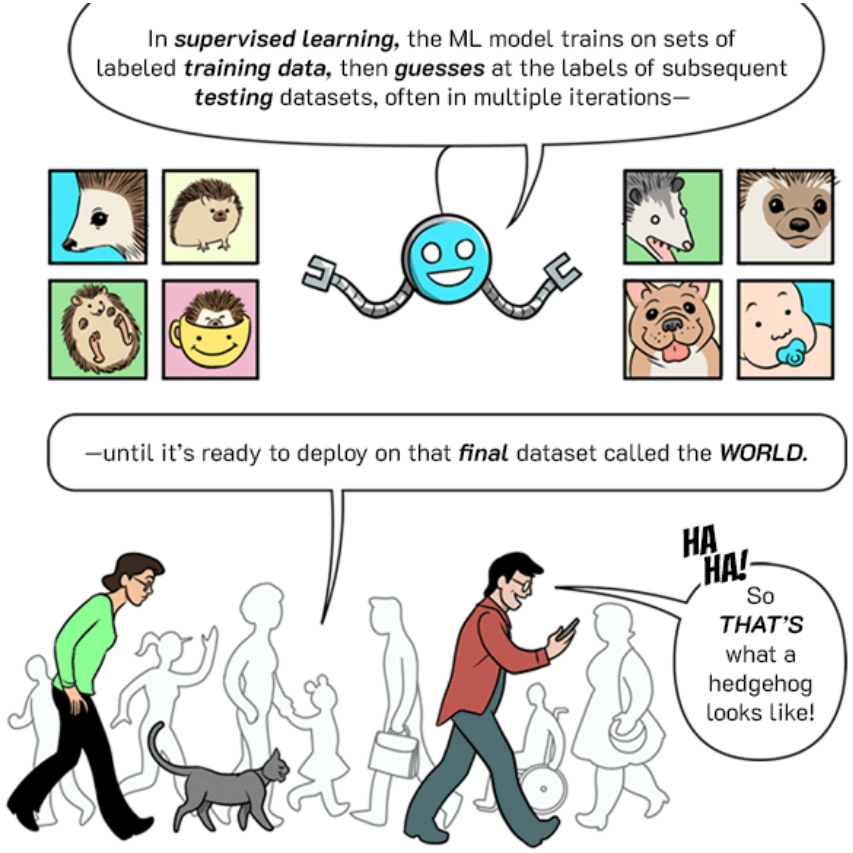




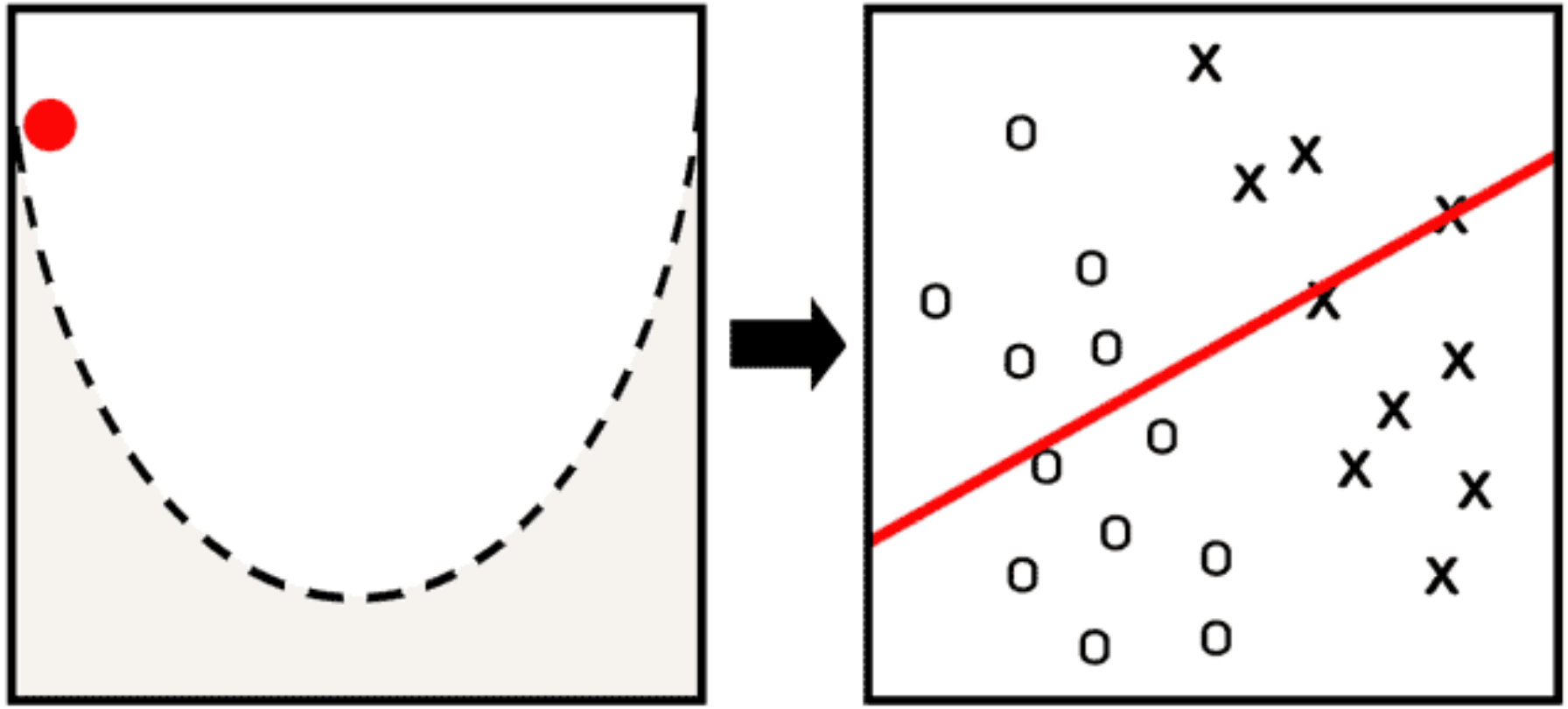
Distinction of different types of learning

Supervised Learning





Supervised learning algorithms experience a dataset containing features, but each example is also associated with a label or target (Goodfellow, Bengio, und Courville 2016).



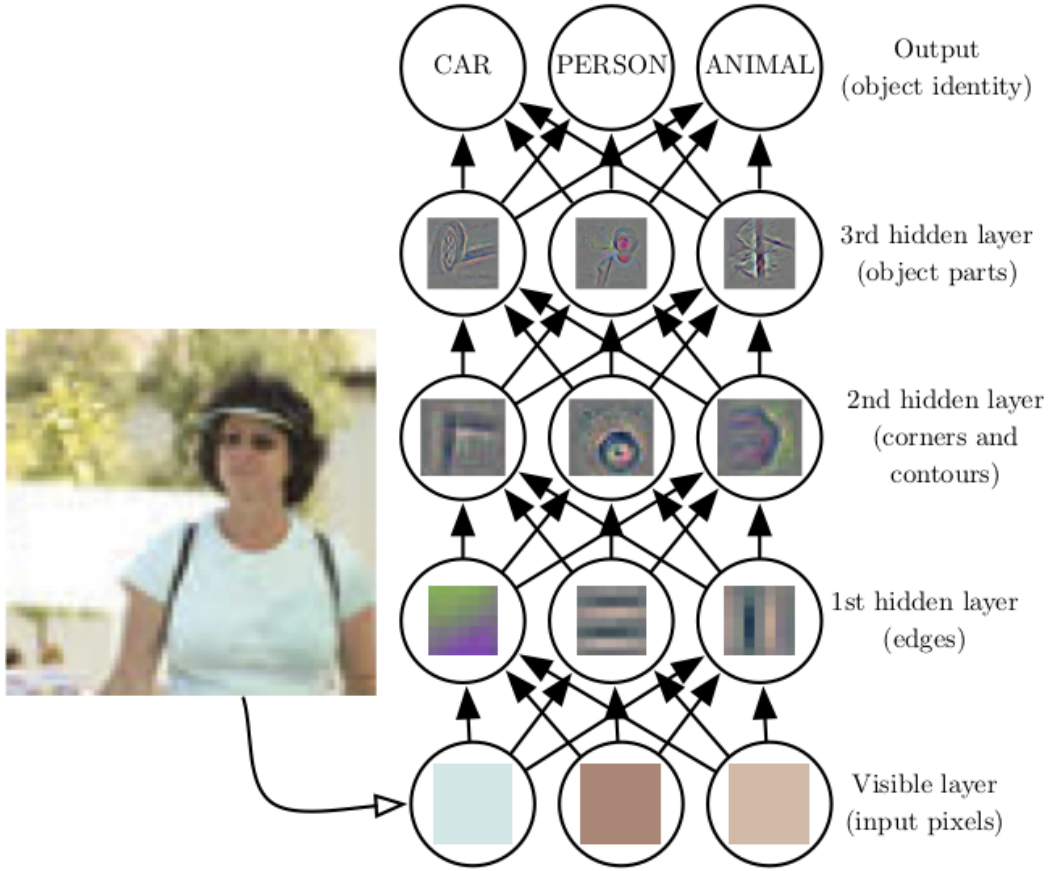
Instant camera translation allows you to see the world in your language by just pointing your camera lens at the foreign text (Gu 2019).

Google translate on a mobile phone.

Classification: Approaches

Classification - Promise of Deep Neural Networks

- Deep Learning deals with high dimensional input data and tries to learn multiple layers of representation.
- The goal is to transform input into gradually higher levels of representation that represent more and more abstract functions of the raw input.
- In Deep Learning all those levels are learnt. As an advantage, features are represented in a distributed fashion and can be shared.



Overview of feature hierarchy in deep neural networks (Goodfellow, Bengio, und Courville 2016).

Unsupervised Learning

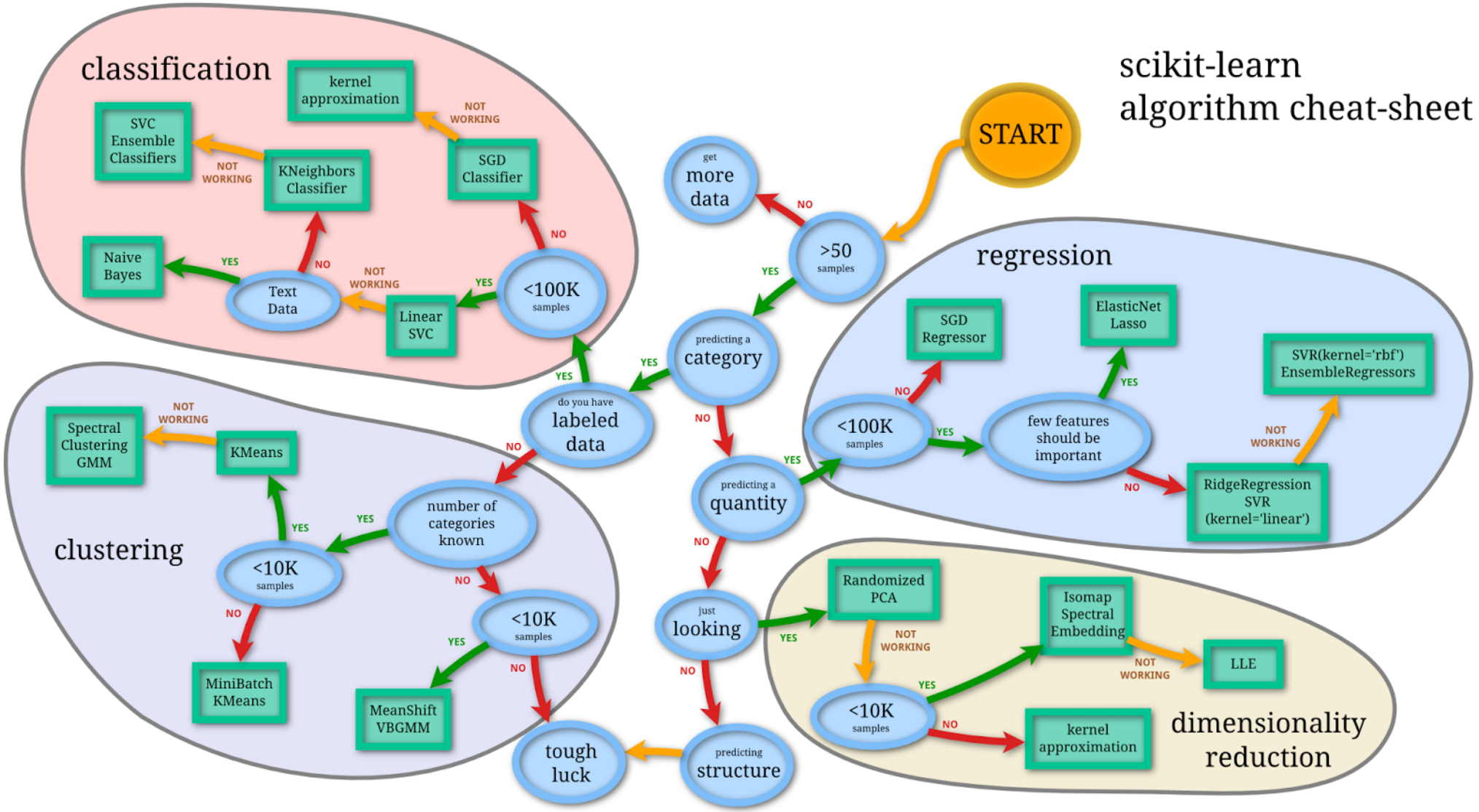
Unsupervised learning algorithms experience a dataset containing many features, then learn useful properties of the structure of this dataset. In the context of deep learning, we

usually want to learn the entire probability distribution that generated a dataset

(Goodfellow, Bengio, und Courville 2016).

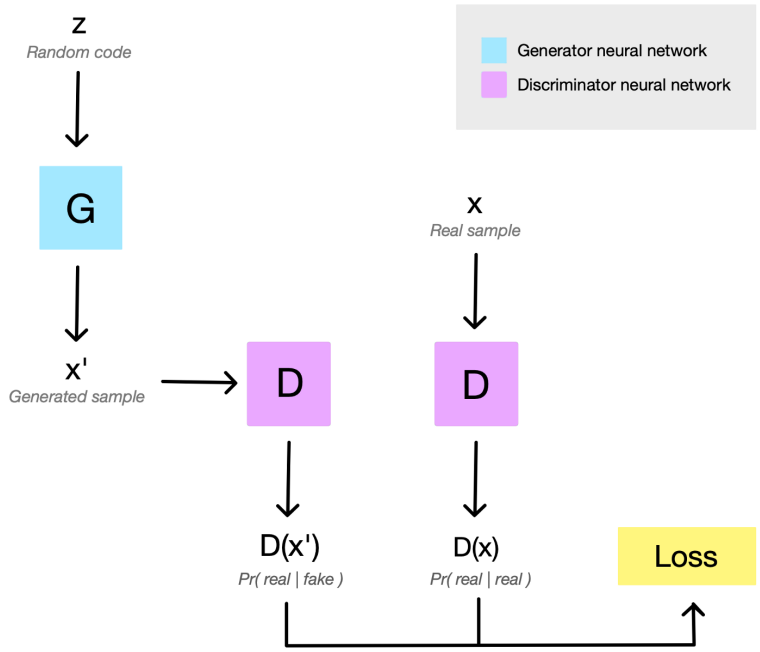
Clustering

What methods do you know?

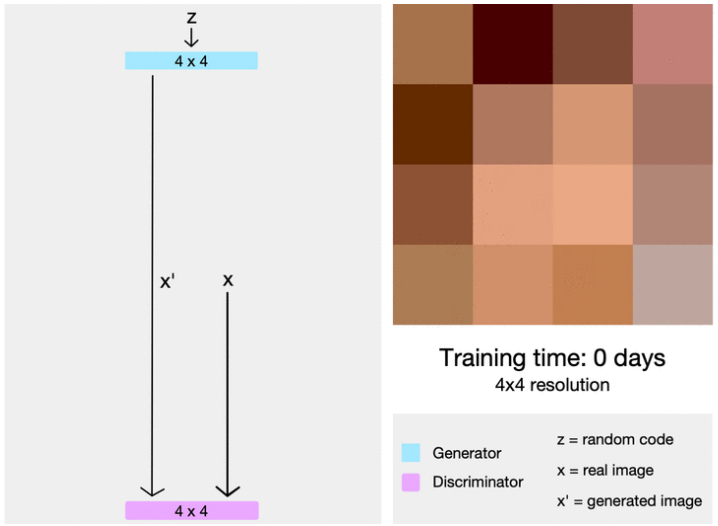


How to decide which machine learning algorithm to choose for a problem („Choosing the right estimator“ 2018).

Example: Generative Adversarial Networks



GAN architecture



ProGAN progressive training

Introduction to ProGAN (Wolf 2018).

Example: This person does not exist

Example: Generative Models – Diffusion

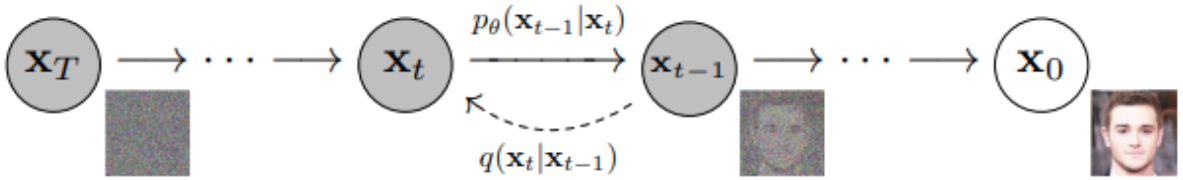
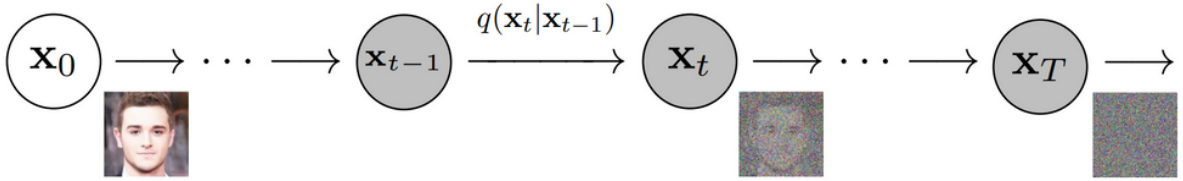
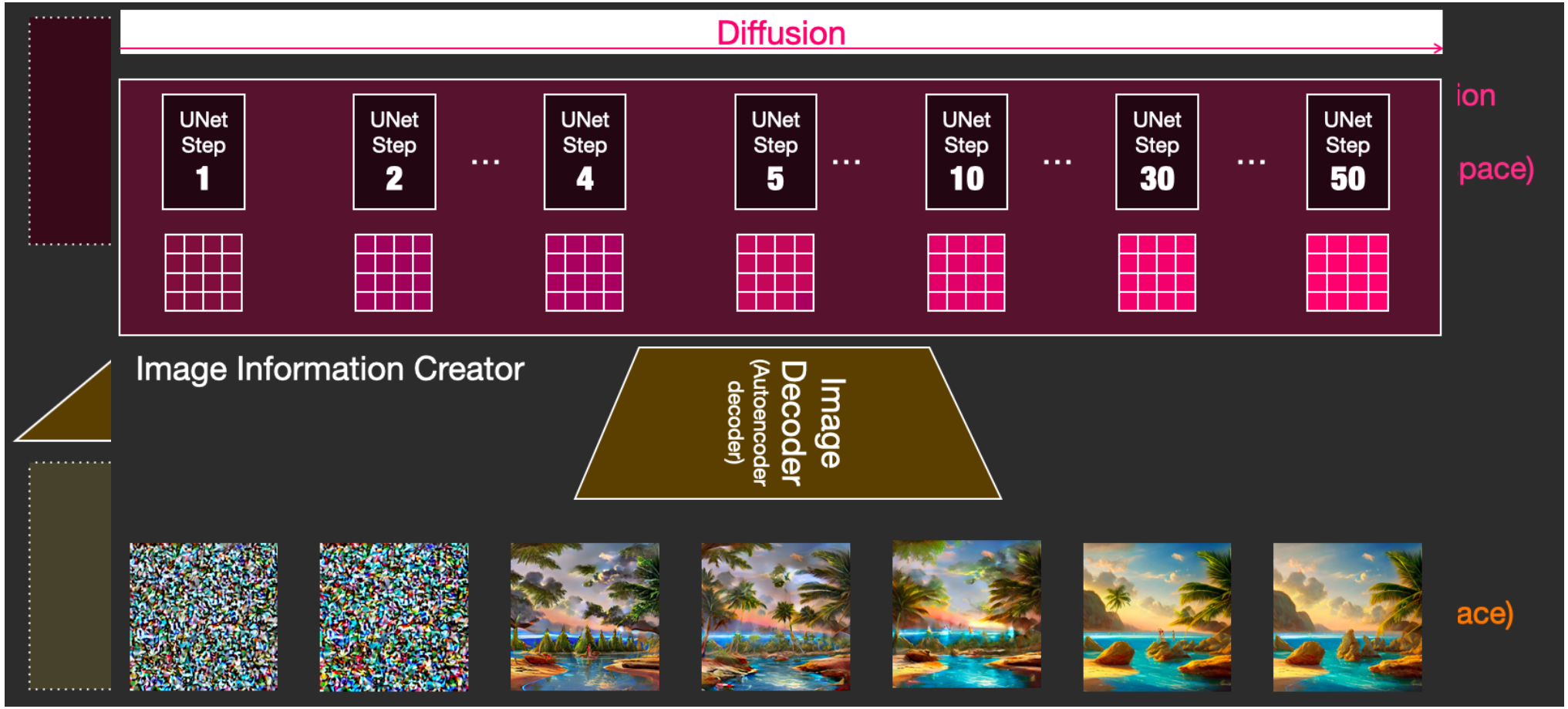


Figure 2: The directed graphical model considered in this work.



(Alammar 2022) and (Ho, Jain, und Abbeel 2020)

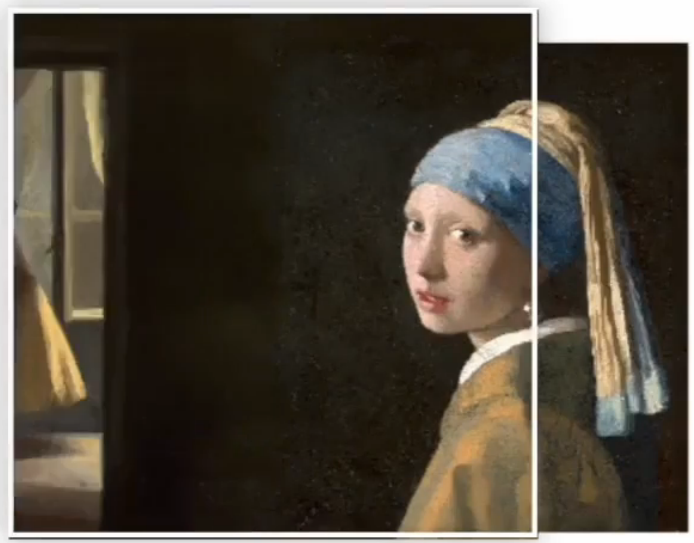
Example: Generative Models – Diffusion 2



(Alammar 2022)

Example: Generative Models – Diffusion 3

(Alammar 2022) and („Dall-E 2“ 2022)

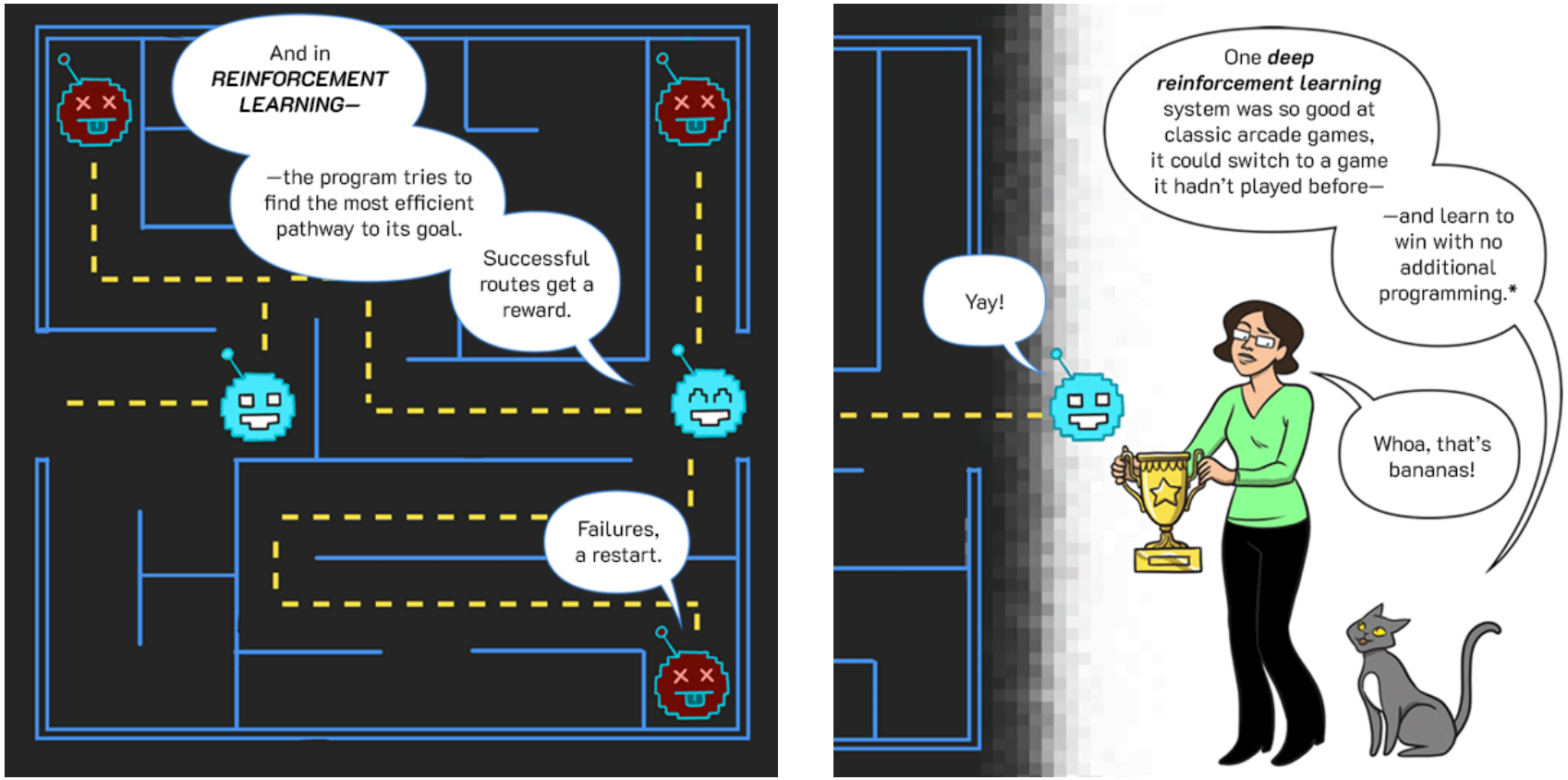


Diffusion Process conditioned on text

(„Dall-E 2“ 2022)

Reinforcement Learning

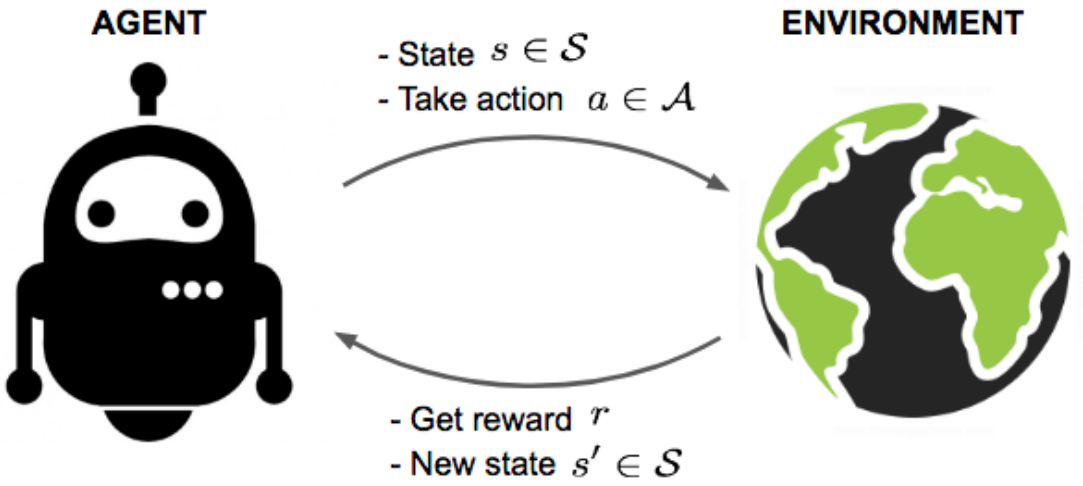
Reinforcement Learning



What is Reinforcement Learning (RL)?

Goal of RL

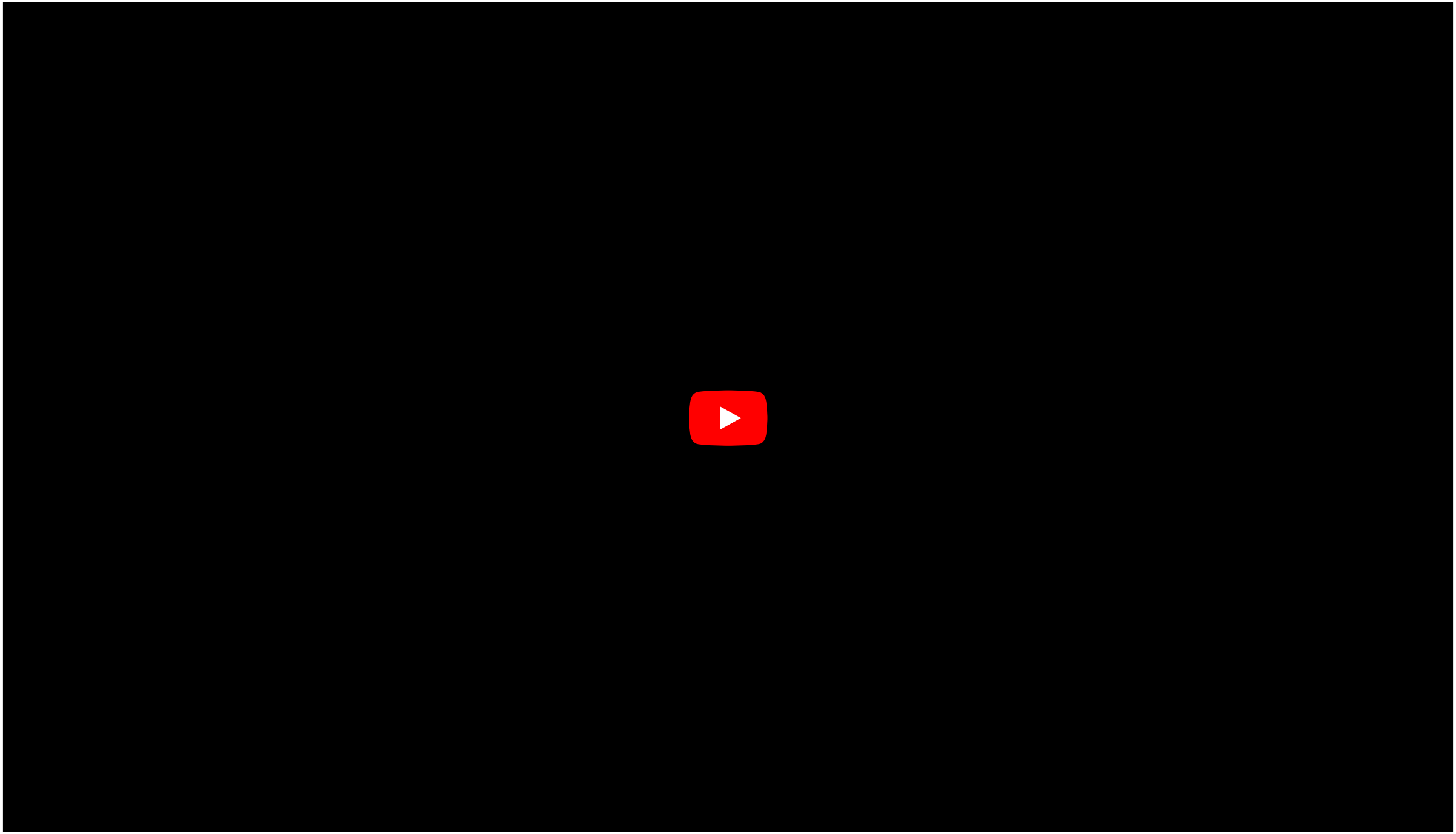
Actively learn a good strategy for an agent from interaction with the environment.



Reinforcement Learning

Reinforcement learning is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal. The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them (Sutton und Barto 2018).

(Weng 2018)



Video - Learning Flying Stanford Helicopter

Video https://www.youtube.com/embed/0JL04JJjocc

The reward hypothesis

Reward Hypothesis (Def.)

All goals can be represented as maximization of a scalar reward (an expected cumulative reward).

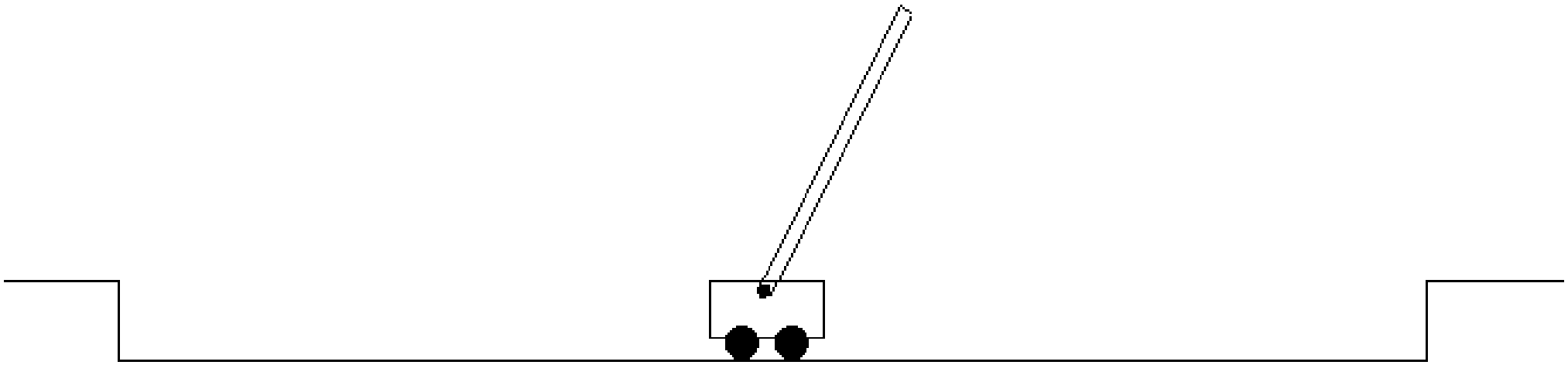
All useful knowledge may be encoded as predictions about rewards, e.g., in the form of a value function.

(Sutton und Barto 2018)

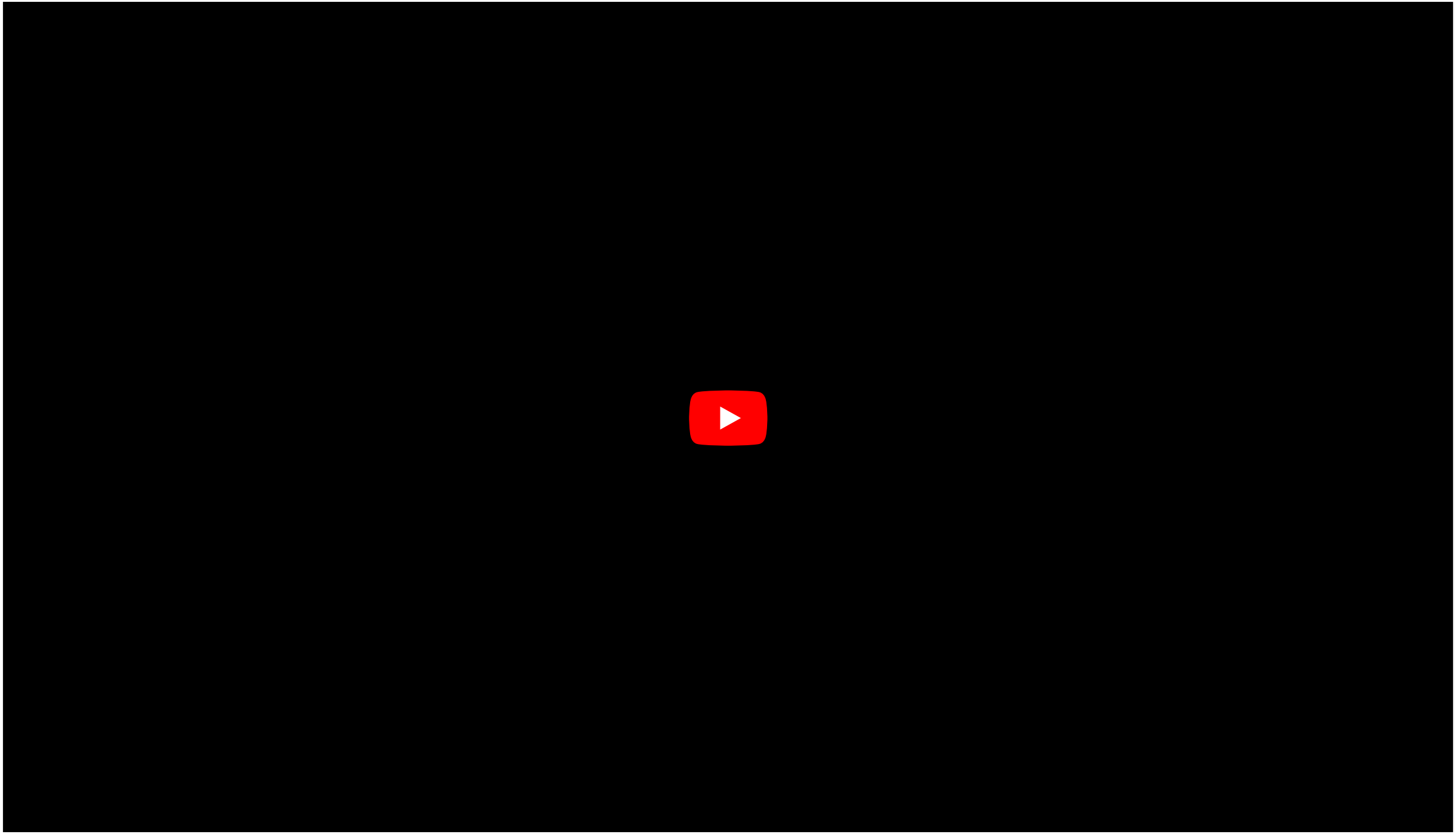
Pole balancing

Training data consists of sequences of situations, actions and an associated reward.

- The task is to learn an (optimal) control strategy that maximizes the reward.
- There is no teacher — only a reward measure.
- There is a tradeoff between exploration and exploitation.



Classical example: A robot that learns to balance a pole.



[redacted]

Video, External Website - Solving Rubik's Cube with a Robot Hand

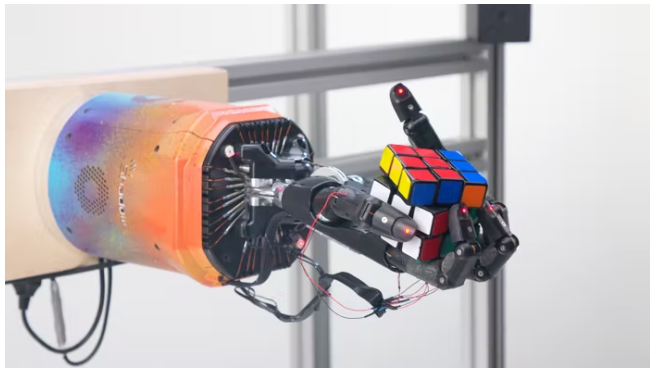
Video https://www.youtube.com/embed/x4O8pojMF0w

Blog entry: https://openai.com/blog/solving-rubiks-cube/

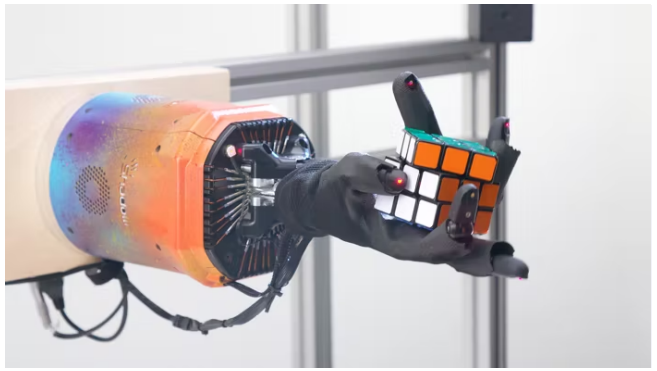
Example: Solving a Rubik's Cube with a Robot Hand

Application of Deep Reinforcement Learning (and many advanced tricks) to learn how to solve a Rubik's Cube with a robotic hand.

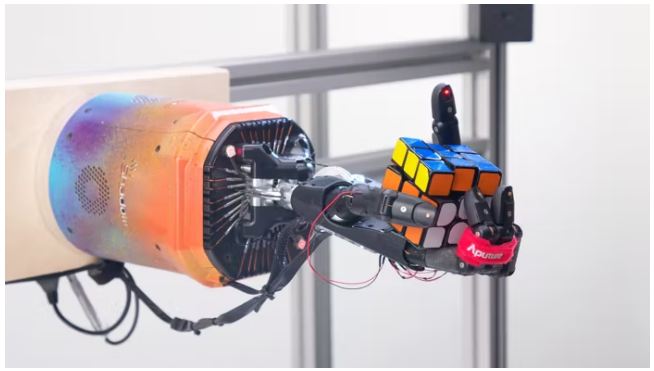
One important ingredient: training in simulation – but allow for enough variation to enforce adaptivity *(automatic domain randomization)*.



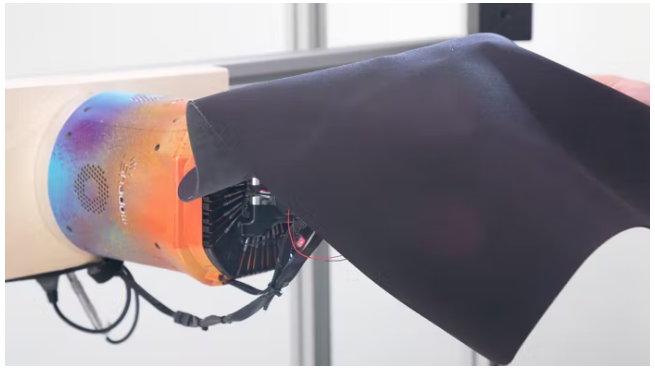
Unperturbed (for reference)



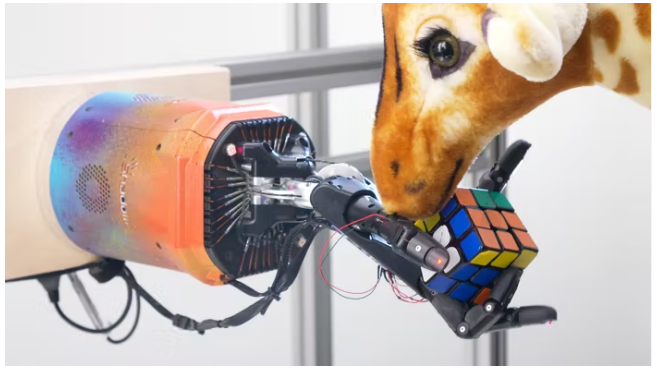
Rubber glove



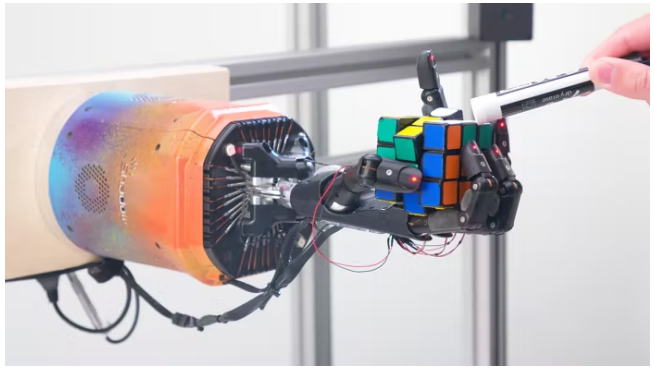
Tied fingers



Blanket occlusion and perturbation



Plush giraffe perturbation



Pen perturbation

Perturbations that we apply to the real robot hand while it solves the Rubik's Cube. All videos play at real-time.

To test the limits of our method, we experiment with a variety of perturbations

(OpenAI u. a. 2019)

Remarks on State-of-the-Art for DRL in Robotics

Caveats

- in testing: robot dropped cube 8 out of 10 trials
- it required 10.000 years of simulated training
- dexterity is very specific to cube
- can adapt to very specific disturbances

One of the parameters we randomize is the size of the Rubik's Cube (above). ADR begins with a fixed size of the Rubik's Cube and gradually increases the randomization range as training progresses. We apply the same technique to all other parameters, such as the mass of the cube, the friction of the robot fingers, and the visual surface materials of the hand. The neural network thus has to learn to solve the Rubik's Cube under all of those increasingly more difficult conditions.

Automatic vs. manual domain randomization

Domain randomization required us to manually specify randomization ranges, which is difficult since too much randomization makes learning difficult but too little randomization hinders transfer to the real robot. ADR solves this by automatically expanding randomization ranges over time with no human intervention. ADR removes the need for domain knowledge and makes it simpler to apply our methods to new tasks. In

None-the-less, this is a remarkable demonstration, but shows how much has to be done to reach human-like capabilities.

(OpenAI u. a. 2019)

My own research

- Interest in understanding and modelling autonomous intelligent systems.
- Understanding how agents can perform intelligently in an environment – meaning: interacting with an environment and dealing with unstructured environments. Towards interacting with other agents in the same space and cooperating with these.
- Locomotion and Manipulation as two prime examples for dealing with an unpredictable environment in an adaptive way.

Cowalski

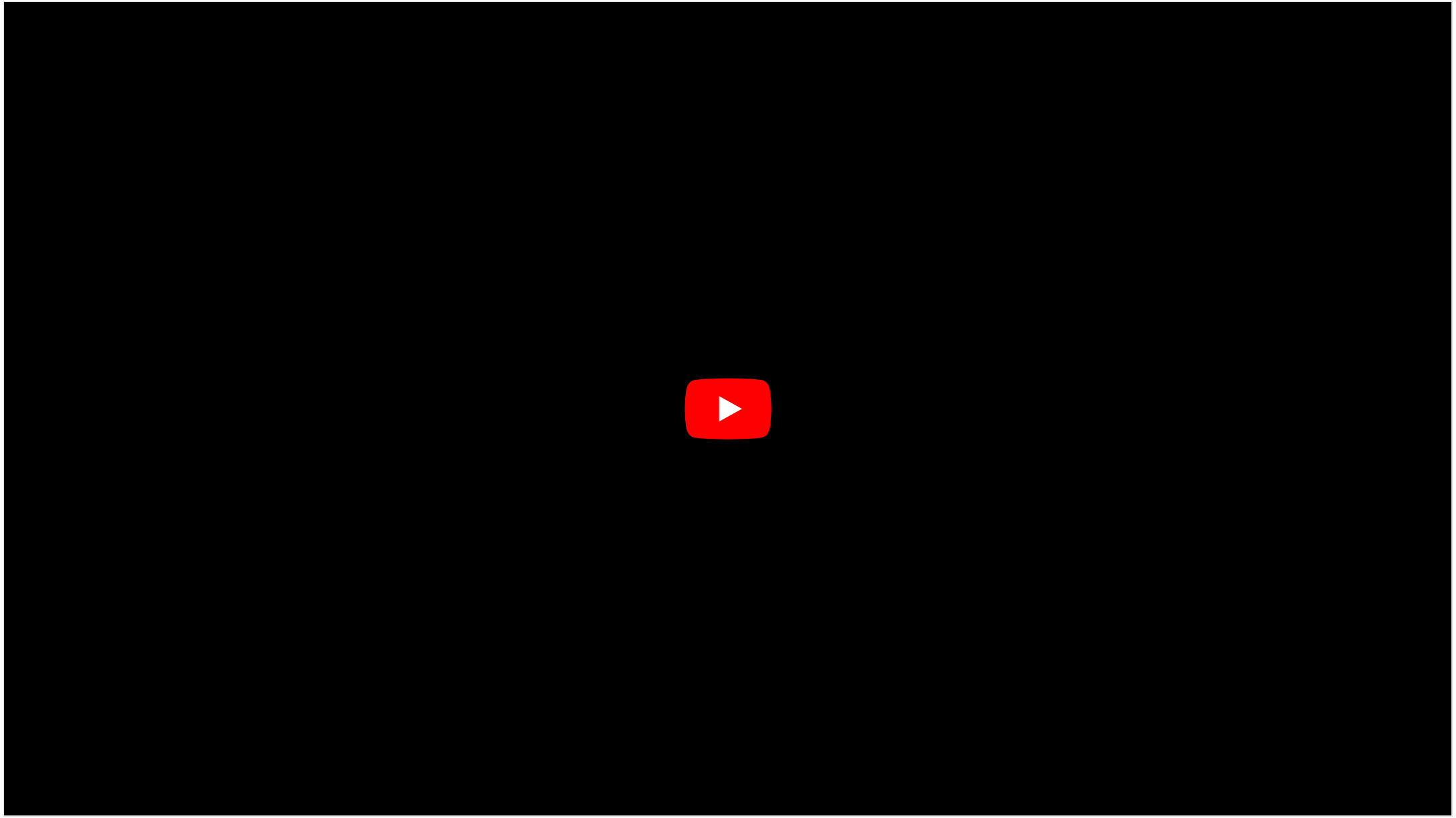
My own research 2

UNIVERSITÄT BIELEFELD
Faculty of Technology

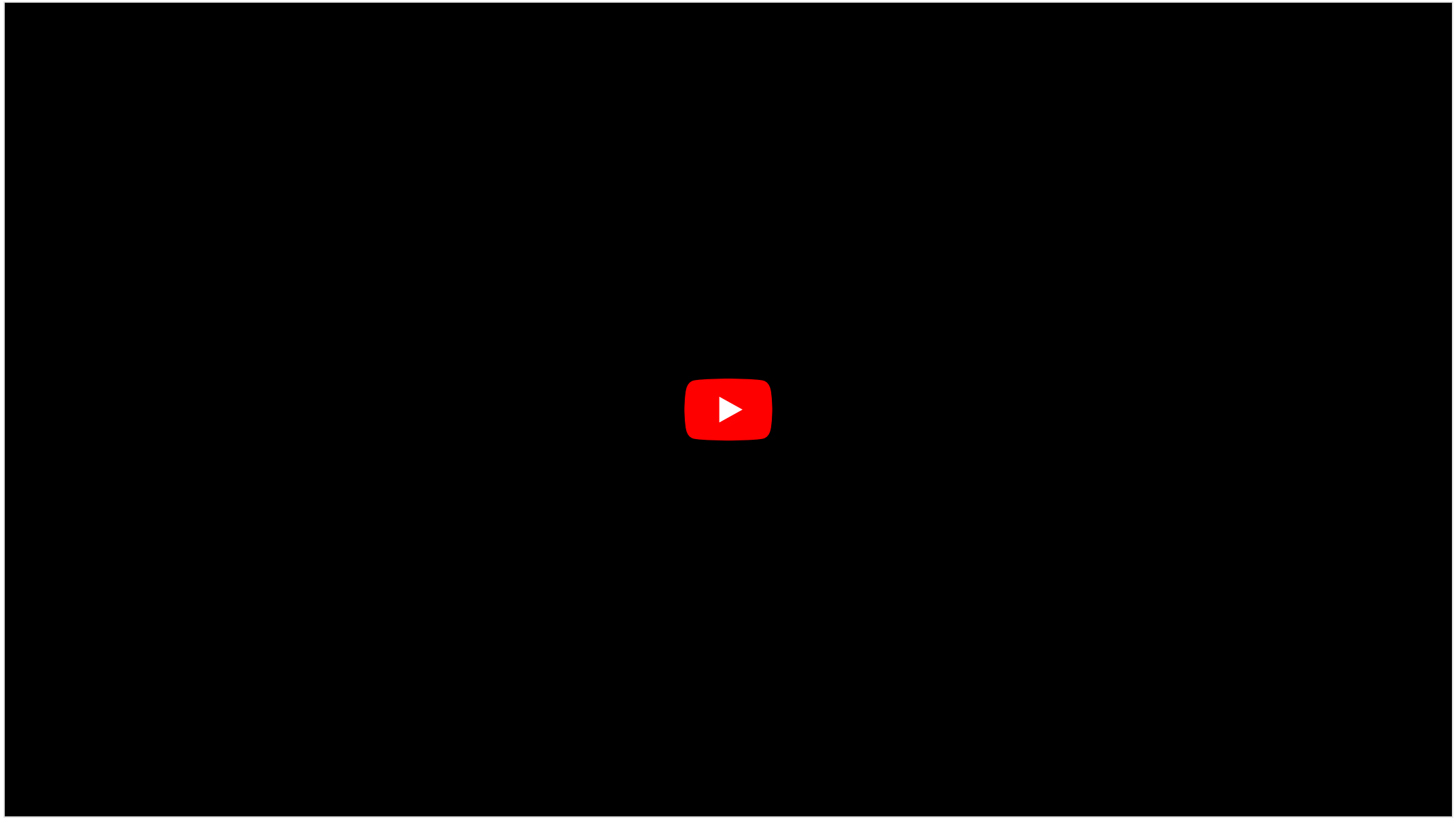
Decentralized Control and Local Information for Robust and Adaptive Decentralized Deep Reinforcement Learning

M. Schilling, A. Melnik, F.W. Ohl, H. Ritter, B. Hammer
Bielefeld University

(Schilling u. a. 2022)



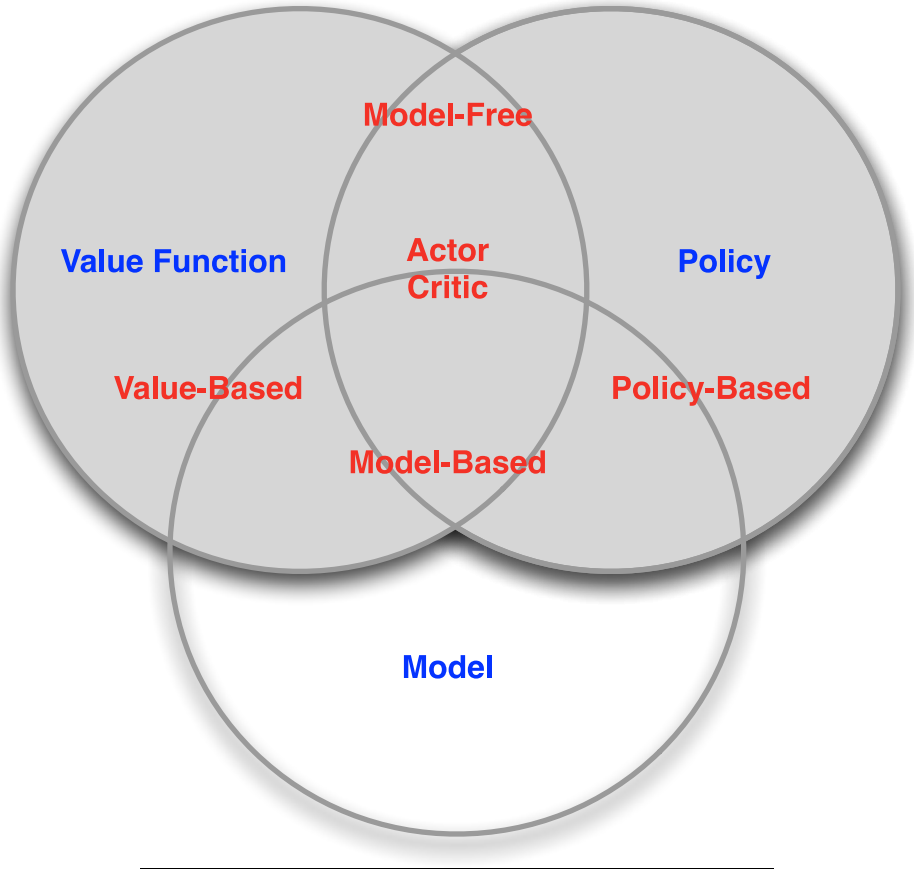
Robotic Systems Lab, ETH Zürich, (Miki u. a. 2022).



from (Baker u. a. 2019)

Categorization of Reinforcement Learning Agents

- Value Based
  - No Policy (Implicit)
  - Value Function
- Policy Based
  - Policy
  - No Value Function
- Actor Critic
  - Policy
  - Value Function



(Silver 2015)

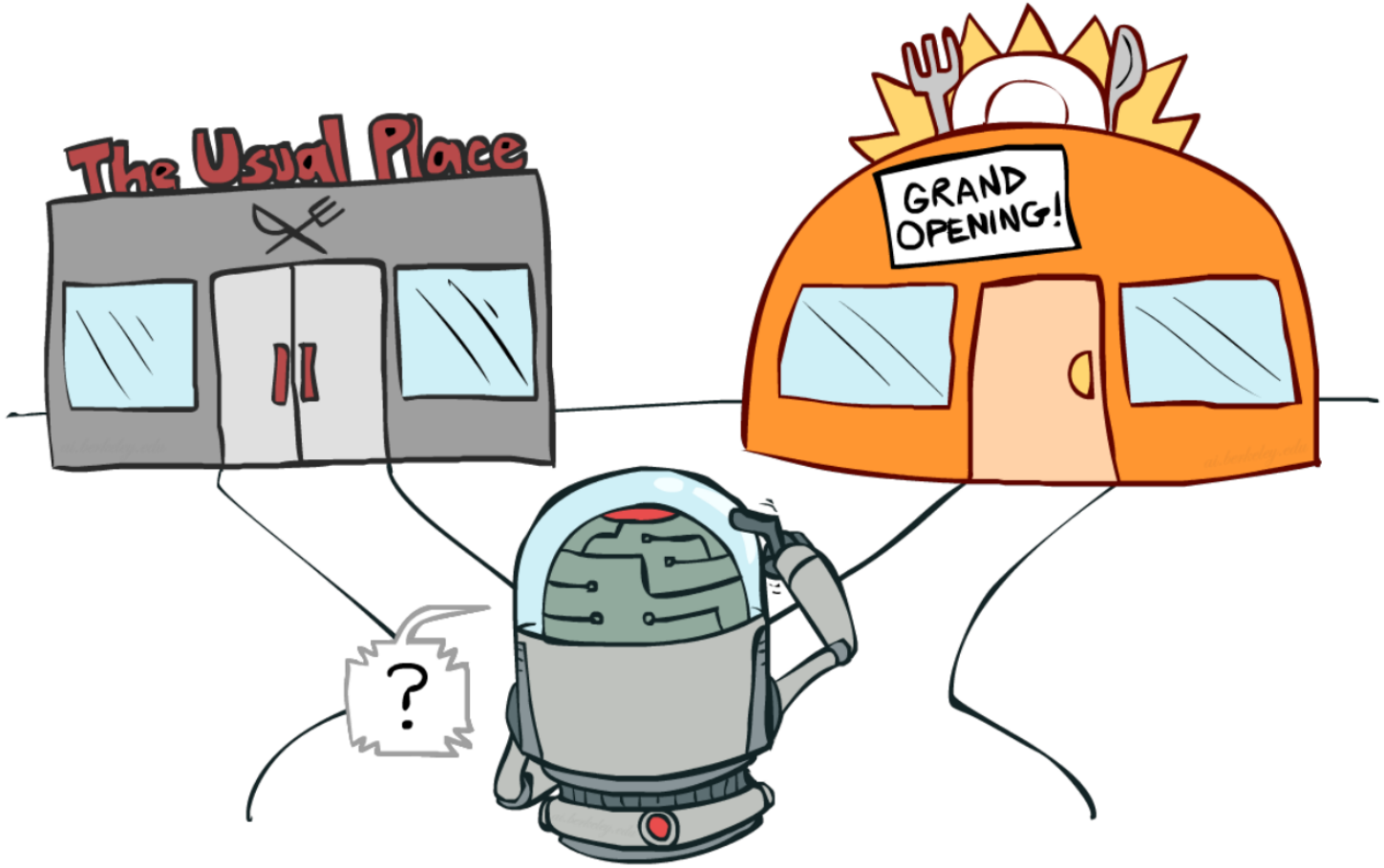
Overview Lecture

DATE	LECTURE
11.10.2022	Introduction
18.10.2022	Formalization of agent and MDP
25.10.2022	Dynamic Programming
8.11.2022	Model-free Prediction
15.11.2022	Model-free Control
22.11.2022	Function Approximation (and RL)

Overview course 2 – tentative!

DATE	LECTURE
29.11.2022	Recap Deep Neural Networks
6.12.2022	Value Function Approximation
13.12.2022	Deep RL, DQN
20.12.2022	Policy Gradient Methods
10.1.2023	Actor-Critic Methods
17.1.2023	Model-based RL
24.1.2023	Recap
31.1.2023	Exam?

Explore or Exploit Information for Decision Making



Decision Making: sticking to a good past experience might make you miss out on even better options, but at least you can be confident to get something good.

(Klein und Abbeel 2014)

References

Alammar, Jay. 2022. „The Illustrated Stable Diffusion“. http://jalammar.github.io/illustrated-stable-diffusion/.

Baker, Bowen, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, und Igor Mordatch. 2019. „Choosing the right estimator“. https://openai.com/blog/emergent-tool-use/.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

„Choosing the right estimator“. 2018. https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html.

„Dall-E 2“. 2022. https://openai.com/dall-e-2/.

Goodfellow, Ian, Yoshua Bengio, und Aaron Courville. 2016. *Deep Learning*. MIT Press.

Gu, Xinxing. 2019. „Google Translate’s instant camera translation gets an upgrade“. https://www.blog.google/products/translate/google-translates-instant-camera-translation-gets-upgrade/.

Ho, Jonathan, Ajay Jain, und Pieter Abbeel. 2020. „Denoising Diffusion Probabilistic Models“. arXiv. doi:10.48550/ARXIV.2006.11239.

Klein, Dan, und Pieter Abbeel. 2014. „UC Berkeley CS188 Intro to AI“. http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html.

Liang, Percy. 2018. „Artificial Intelligence: Principles and Techniques“. Course CS221, Stanford University, Lecture Notes.

„Machine Learning - an online comic from google AI“. 2019. https://cloud.google.com/products/ai/ml-comic-1/.

Miki, Takahiro, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, und Marco Hutter. 2022. „Learning robust perceptive locomotion for quadrupedal robots in the wild“. *Science Robotics* 7 (62): eabk2822. doi:10.1126/scirobotics.abk2822.

OpenAI, Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, u. a. 2019. „Solving Rubik’s Cube with a Robot Hand“. https://arxiv.org/abs/1910.07113.

Schilling, Malte, Andrew Melnik, Frank W. Ohl, Helge J. Ritter, und Barbara Hammer. 2022. „Decentralized control and local information for robust and adaptive decentralized Deep Reinforcement Learning“. *Neural Networks* 144: 699–725. doi:https://doi.org/10.1016/j.neunet.2021.09.017.

Silver, David. 2015. „UCL Course on RL UCL Course on RL UCL Course on Reinforcement Learning“. http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html.