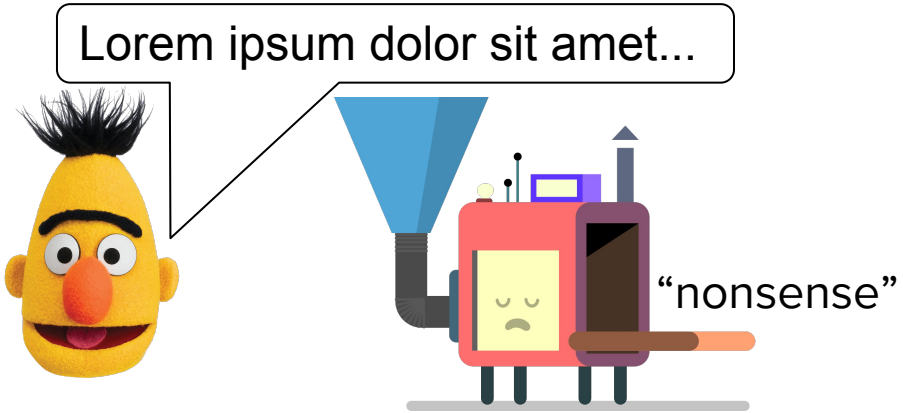
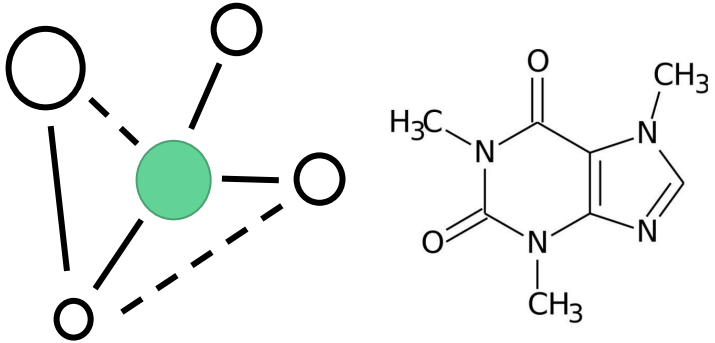
Seminar in Deep Neural Networks

Introduction

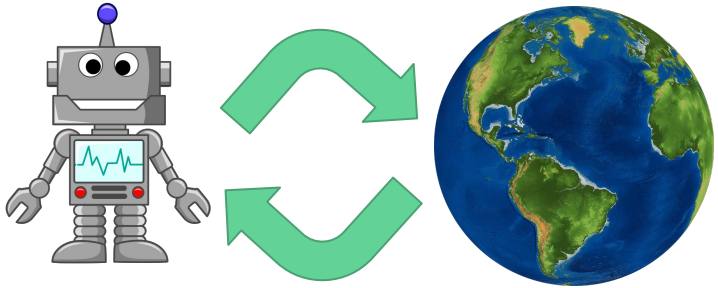
Natural Language Processing



Graph Neural Networks



Reinforcement Learning



Algorithmic Learning

Disclaimer: This is a seminar...

(almost) no basics

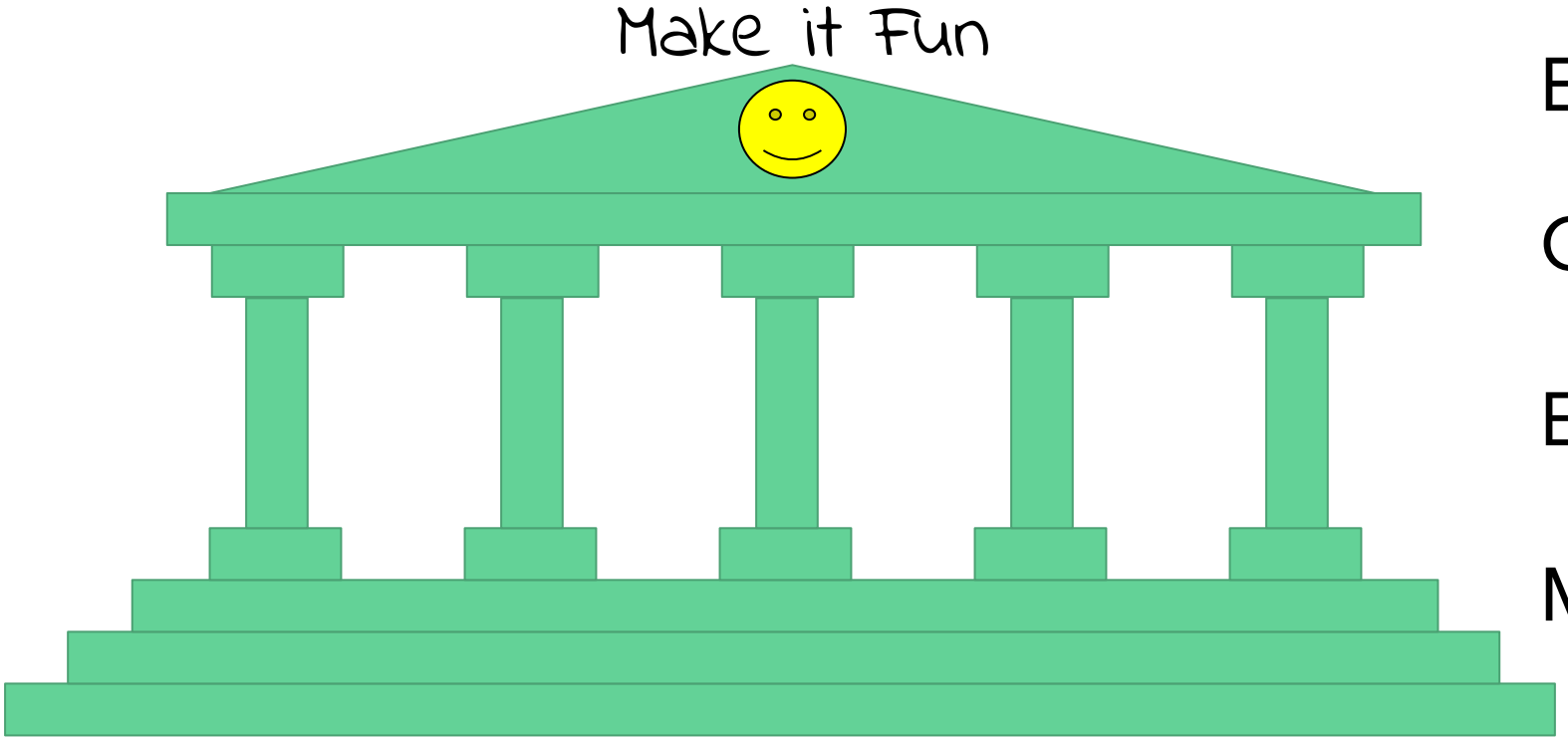
participation required

Format

- Assigned topics
- 2 x 35 min video presentation + Piazza
- 30 min facilitated discussion on zoom
- Feedback through Google Form

Grade = presentation + **active participation**

What makes a good talk?



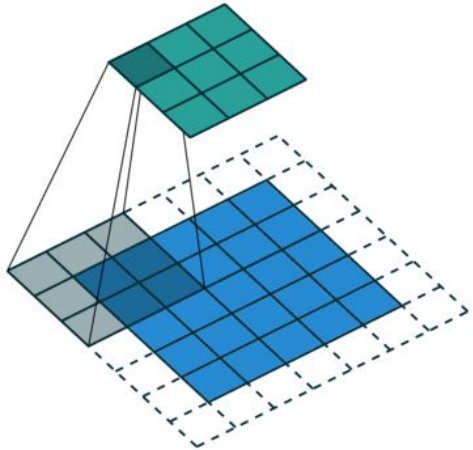
Explore

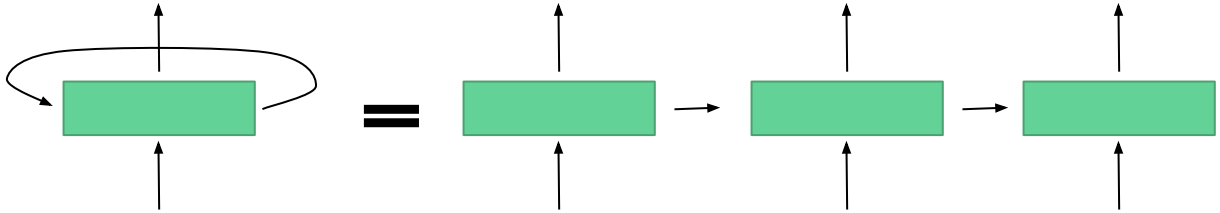
Connect

Build

Motivate

Neural architectures





$$\hat{f}(x) \approx f(x)$$

Natural Language Processing... Evaluation?

Chat Bots

Translation

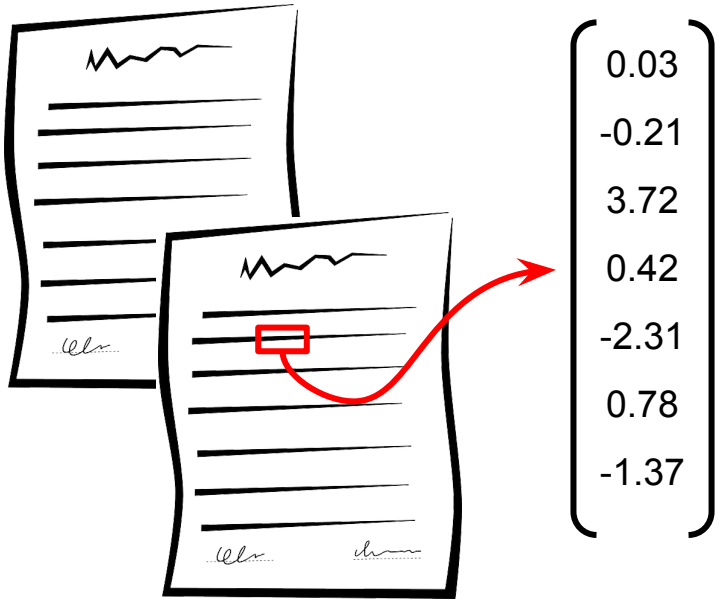
Trend Prediction

Answer Extraction

Spam Filter

Text Completion

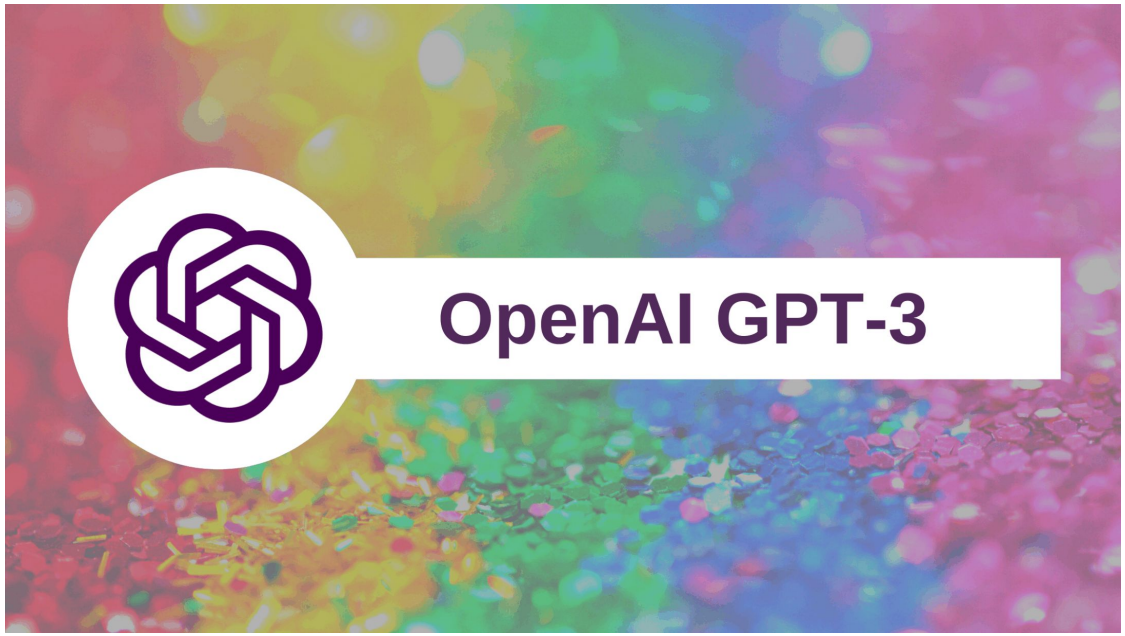
Natural Language Processing



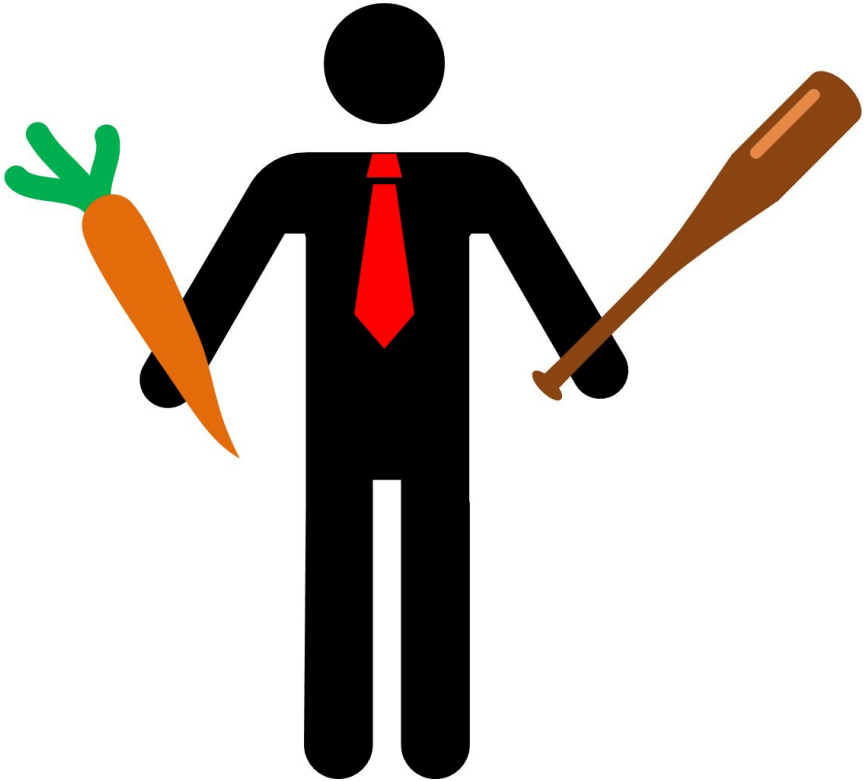
0.03
-0.21
3.72
0.42
-2.31
0.78
-1.37

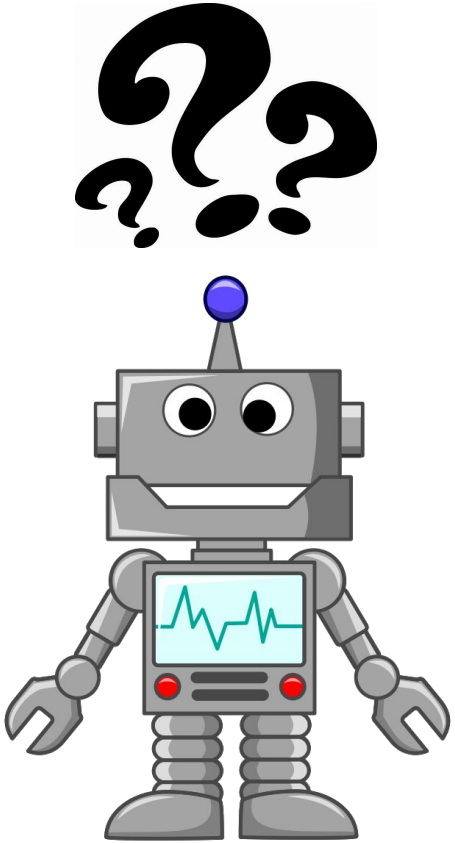
Self-Attention
BERT
GPT

How powerful are these models already?

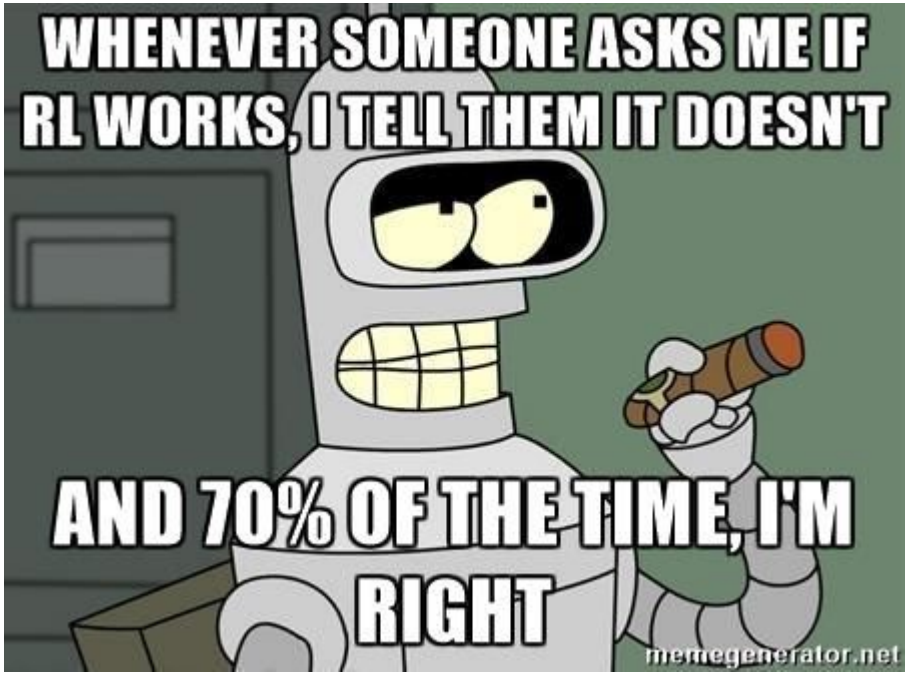


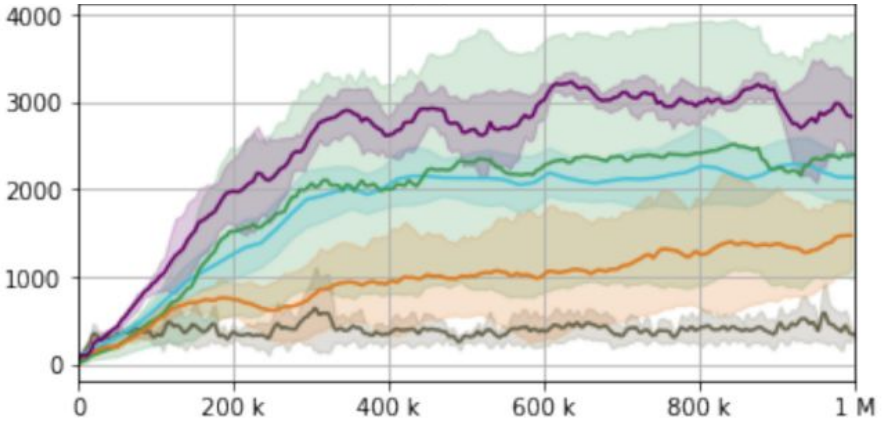
Reinforcement Learning...

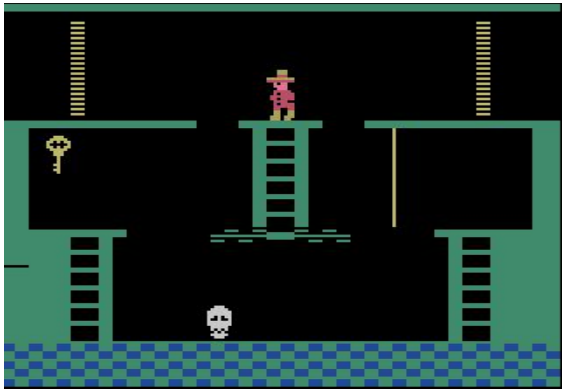




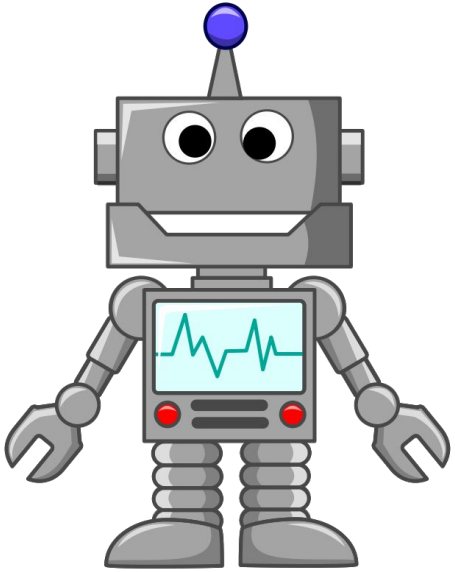
Why you should NOT use reinforcement learning...







Why you should NOT use reinforcement learning...



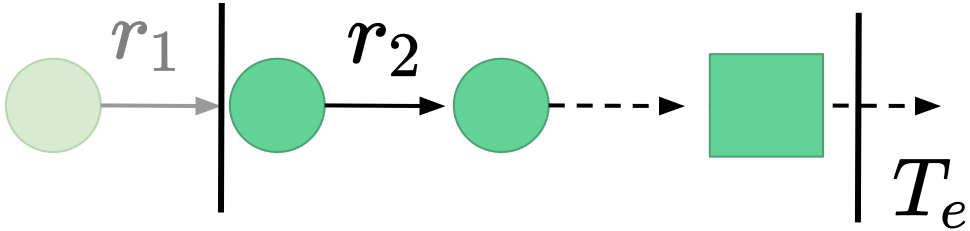
Agent

Action

State
Reward

Environment

How does RL work?



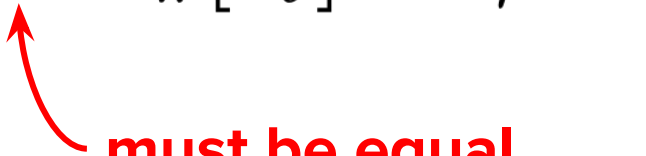
$$R = \sum_{t'=t}^{T_e} \gamma^{t'-t} r_{t'}$$

$$\gamma \in (0, 1]$$

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}[\sum_{t'=t}^{T_e} \gamma^{t'-t} r_{t'}]$$

$$= \mathbb{E}_{\pi}[r_t] + \gamma \mathbb{E}_{\pi}[\sum_{t'=t+1}^{T_e} \gamma^{t'-t-1} r_{t'}]$$

$$= \mathbb{E}_{\pi}[r_t] + \gamma V^{\pi}(s_{t+1})$$



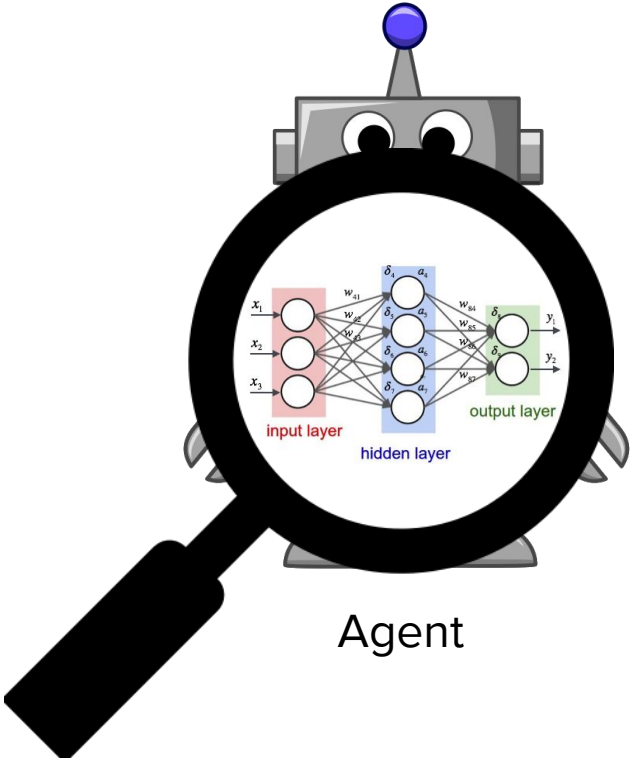
Q-Learning - Watkins (1989)

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}[r_t] + \gamma V^{\pi}(s_{t+1})$$

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{a_t}[r_t] + \gamma V^{\pi}(s_{t+1})$$

$$V^{*}(s) = \max_a Q^{*}(s, a)$$

$$Q^{*}(s_t, a_t) = \mathbb{E}_{a_t}[r_t] + \gamma \max_a Q^{*}(s_{t+1}, a)$$

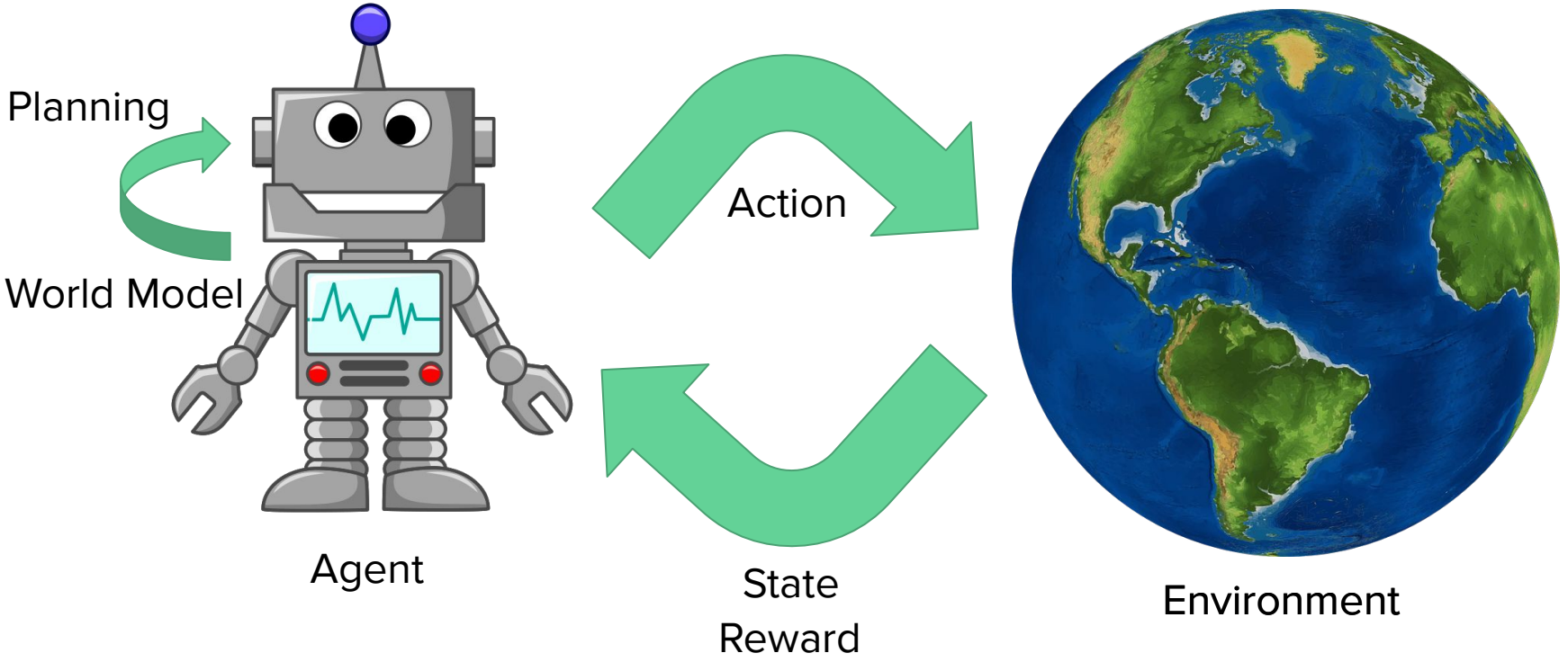


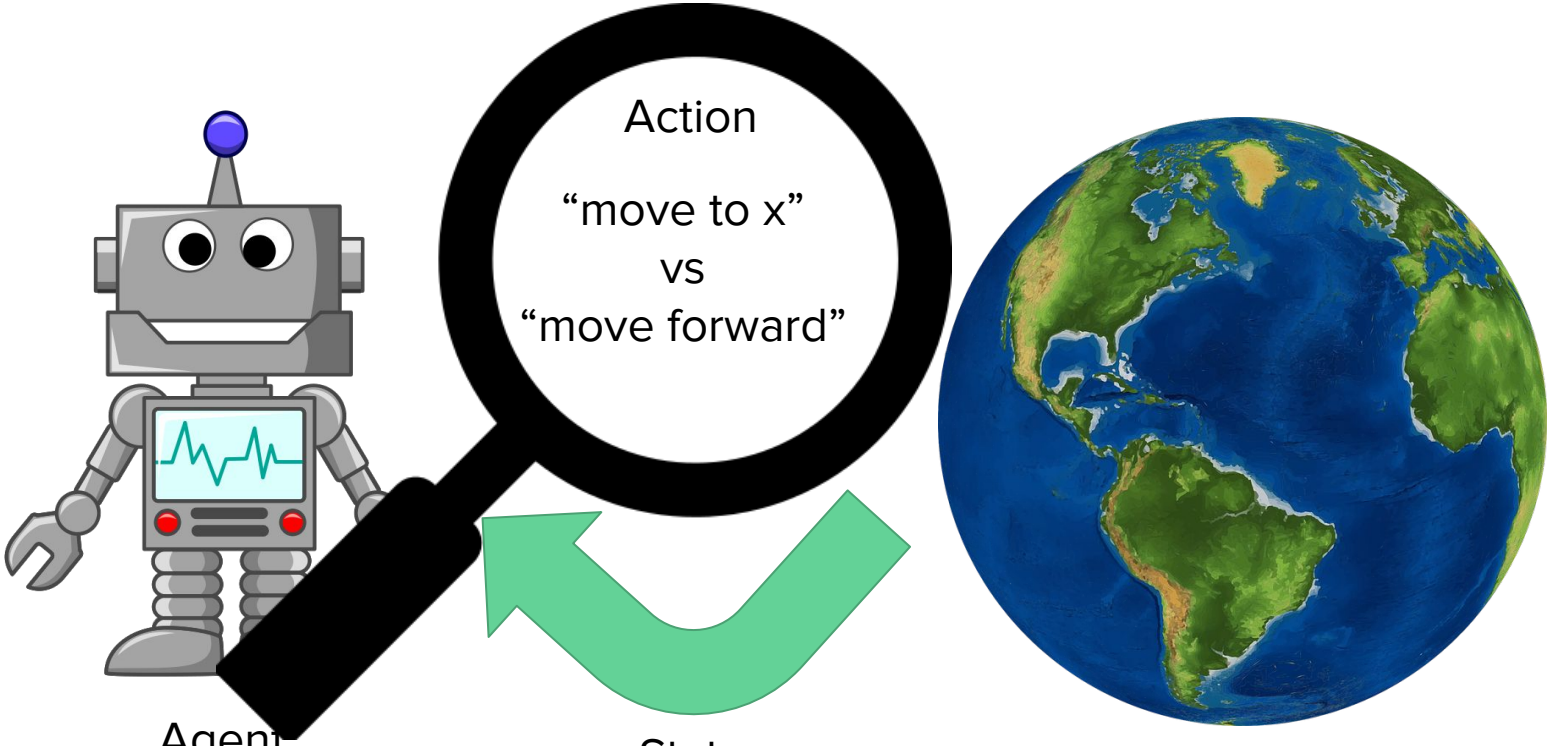
Agent

Action

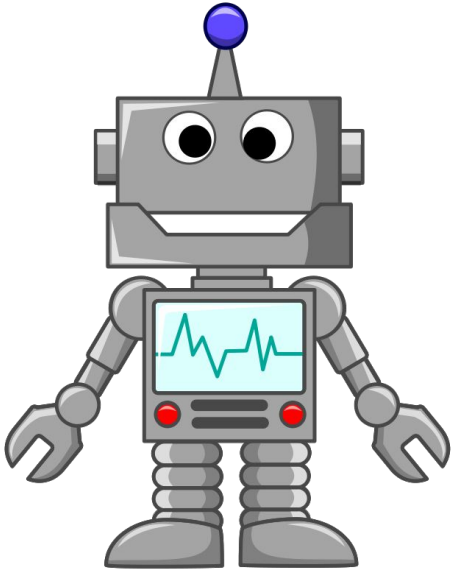
State
Reward

Environment


Planning
World Model
Agent
Action
State
Reward
Environment


Action
“move to x”
vs
“move forward”
Agent
State
Reward
Environment

non-differentiable



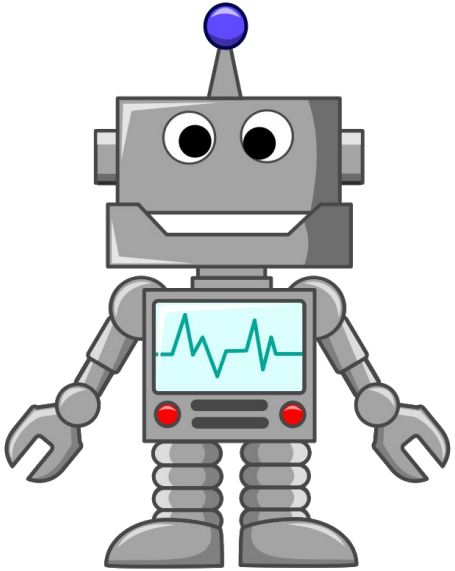
Agent

Action

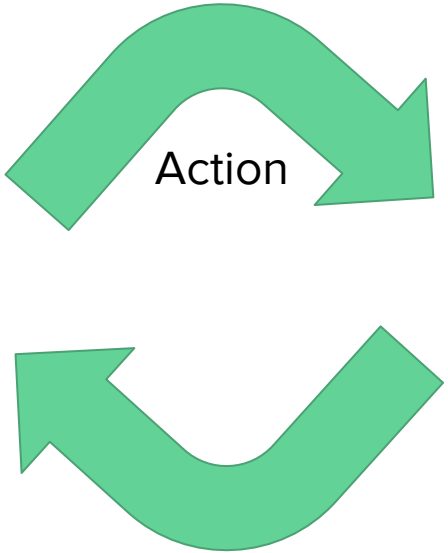
State
Reward

Environment

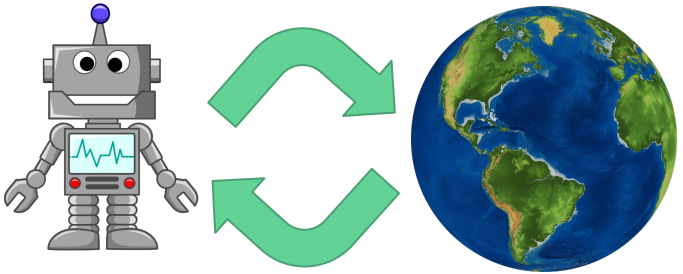
non-differentiable



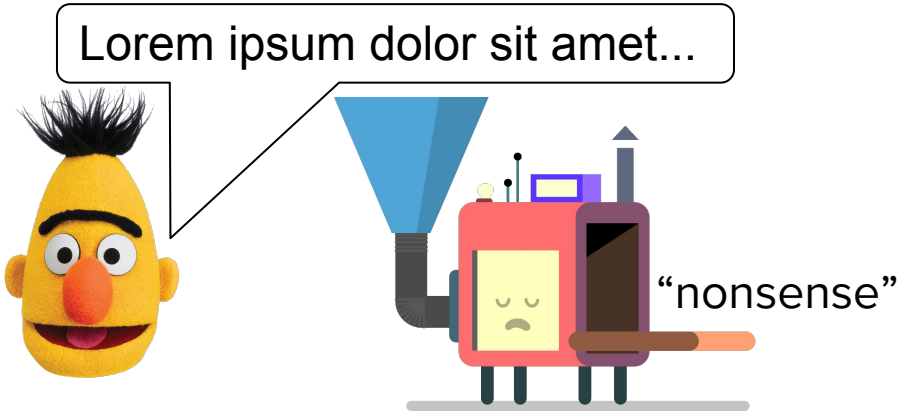
Agent



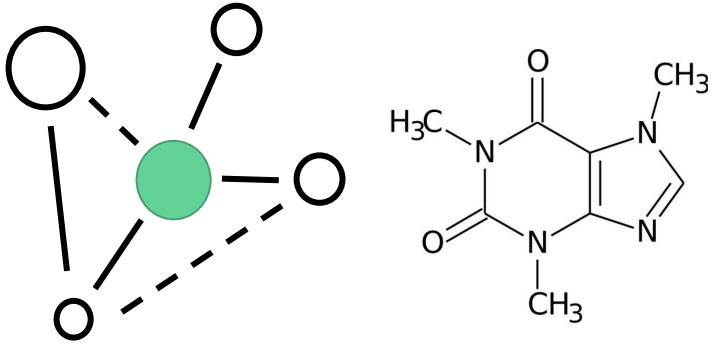
State
Reward



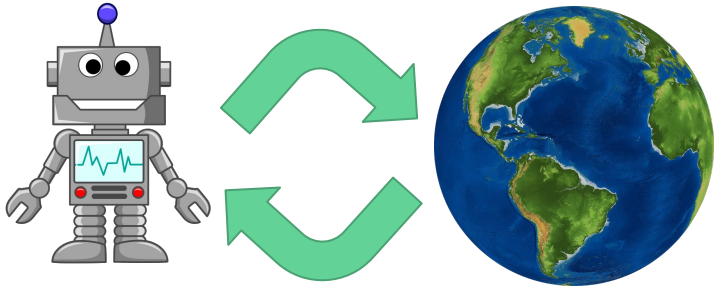
Natural Language Processing



Graph Neural Networks

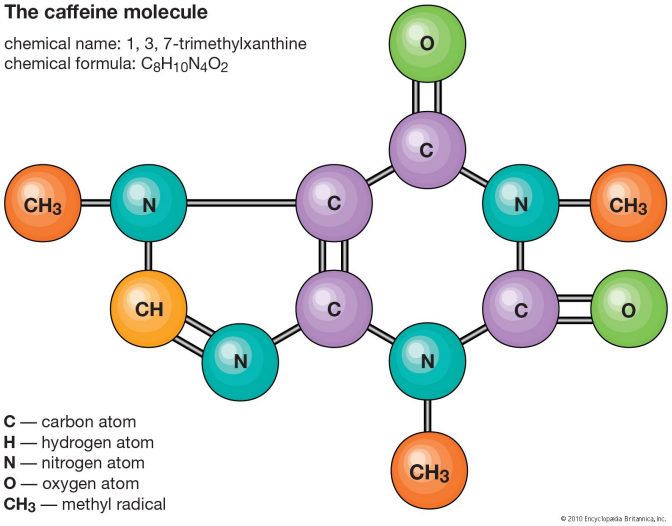


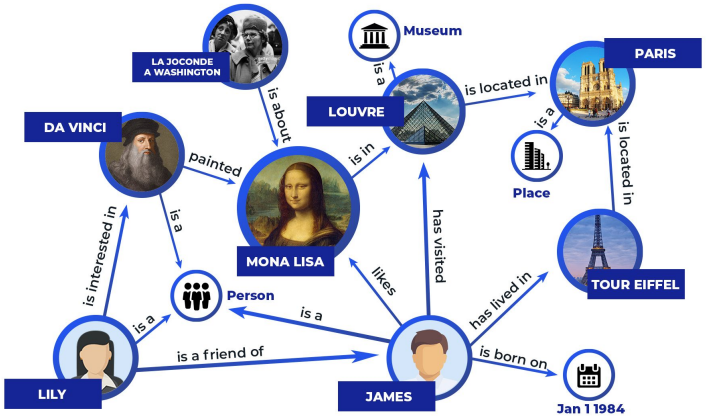
Reinforcement Learning

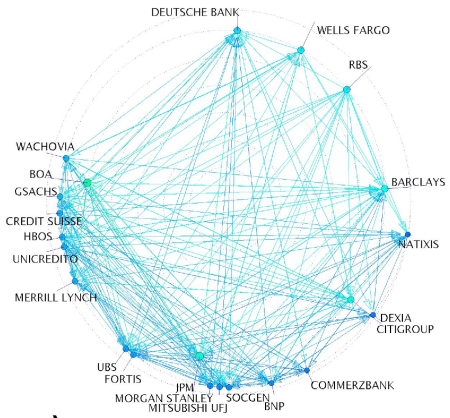


Algorithmic Learning

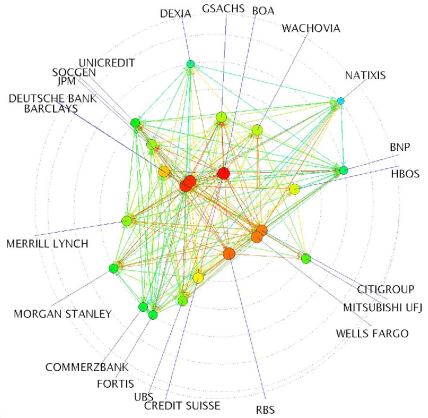
Graph Neural Networks


The caffeine molecule
chemical name: 1, 3, 7-trimethylxanthine
chemical formula: $C_8H_{10}N_4O_2$
O
C
CH_3
N
C
N
CH_3
CH
C
C
O
N
N
CH_3
C — carbon atom
H — hydrogen atom
N — nitrogen atom
O — oxygen atom
CH_3 — methyl radical
© 2010 Encyclopædia Britannica, Inc.

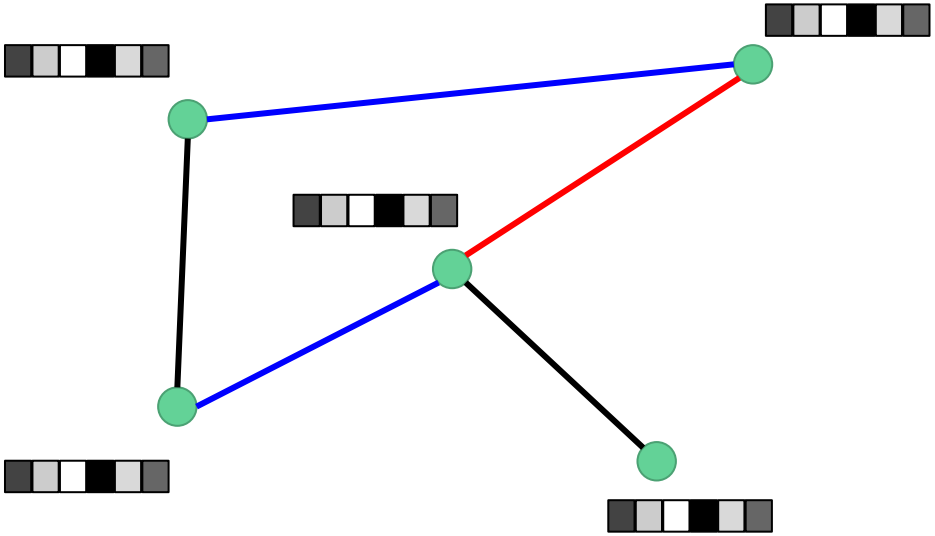

LA JOCONDE A WASHINGTON
is about
Museum
is a
LOUVRE
is located in
PARIS
is a
Place
is located in
DA VINCI
painted
is in
MONA LISA
has visited
likes
TOUR EIFFEL
is interested in
is a
Person
is a
is a friend of
has lived in
is born on
LILY
JAMES
Jan 1 1984


DEUTSCHE BANK
WELLS FARGO
RBS
BARCLAYS
NATIXIS
DEXIA
CITIGROUP
COMMERZBANK
BNP
SOCGEN
JPM
MORGAN STANLEY
MITSUBISHI UFJ
UBS
FORTIS
MERRILL LYNCH
UNICREDITO
HBOS
CREDIT SUISSE
GSACHS
BOA
WACHOVIA

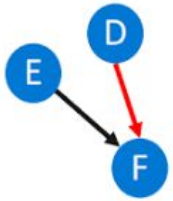
a)


DEXIA
GSACHS
BOA
WACHOVIA
NATIXIS
BNP
HBOS
CITIGROUP
MITSUBISHI UFJ
WELLS FARGO
RBS
CREDIT SUISSE
UBS
FORTIS
COMMERZBANK
MORGAN STANLEY
MERRILL LYNCH
DEUTSCHE BANK
BARCLAYS
UNICREDIT
SOCGEN
JPM

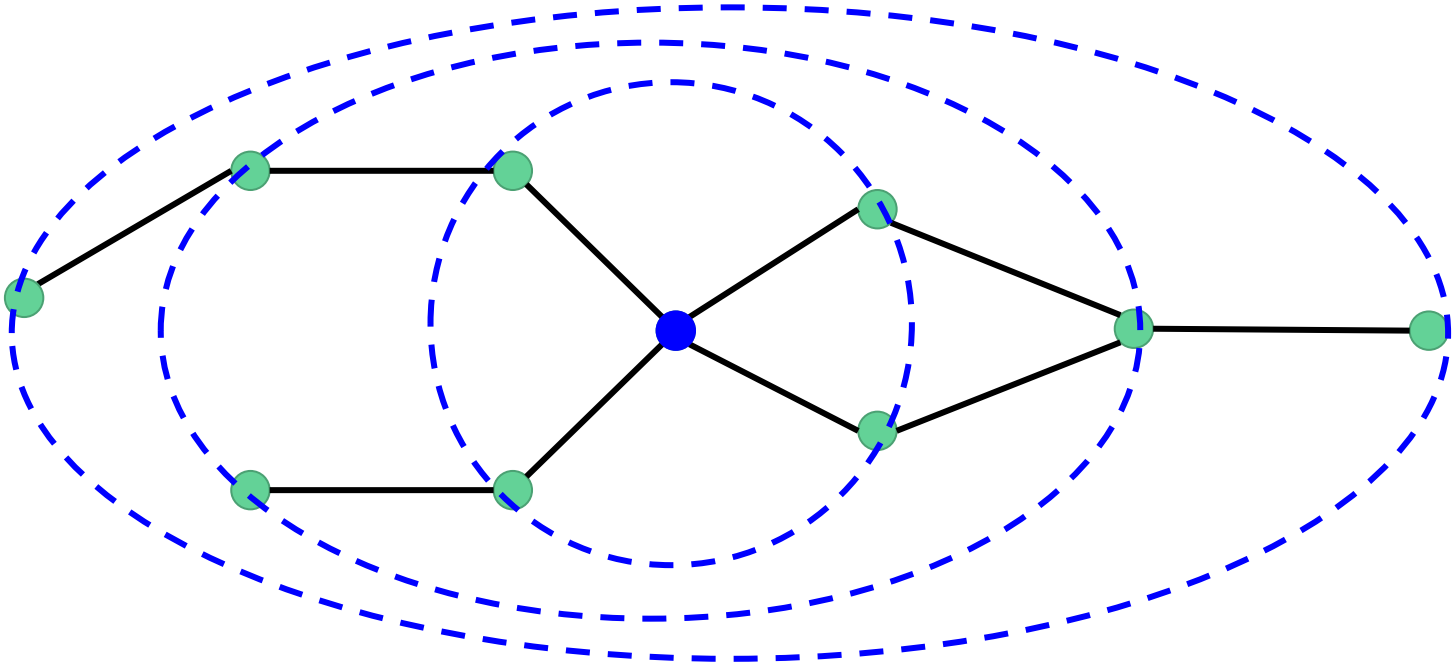
b)



Neural Message Passing



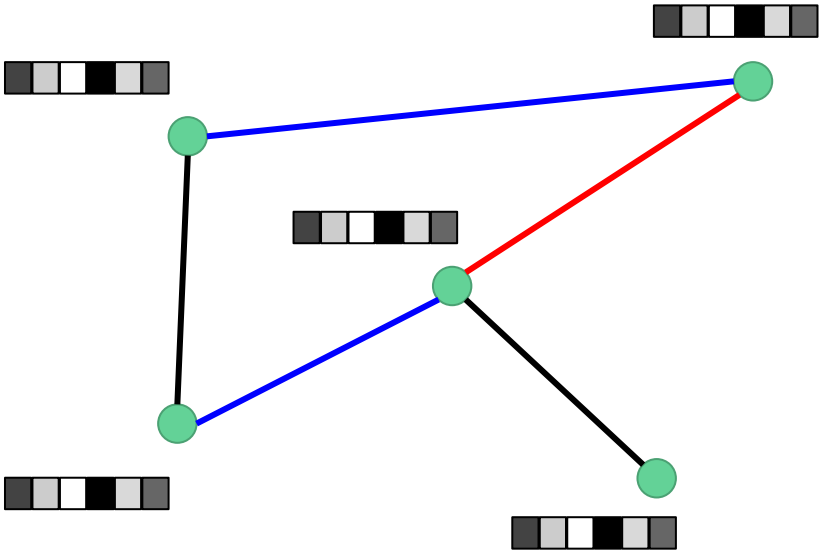
K-hop neighbourhood

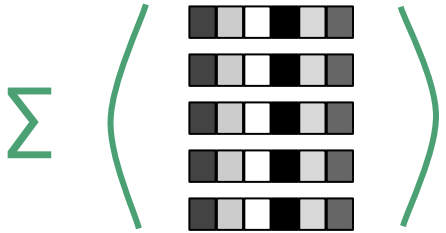


Readout

Node selection

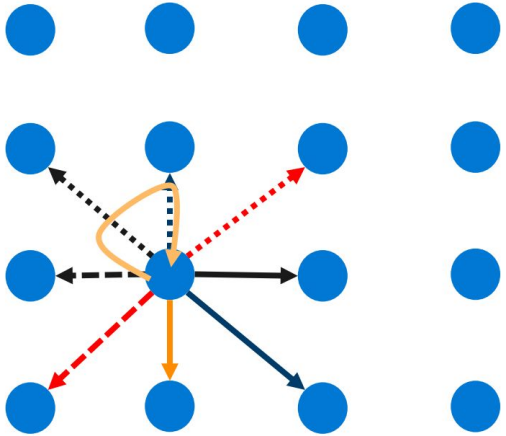
Node Classification

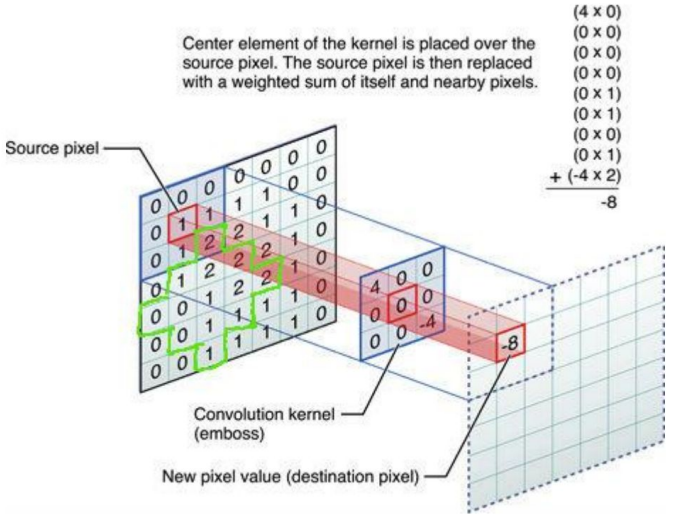




Graph Classification

GNNs vs CNNs




Center element of the kernel is placed over the source pixel. The source pixel is then replaced with a weighted sum of itself and nearby pixels.
(4 x 0)
(0 x 0)
(0 x 0)
(0 x 0)
(0 x 1)
(0 x 1)
(0 x 0)
(0 x 1)
+ (-4 x 2)
-8
Source pixel
Convolution kernel (emboss)
New pixel value (destination pixel)

What are they good at?

What are they **not** good at?

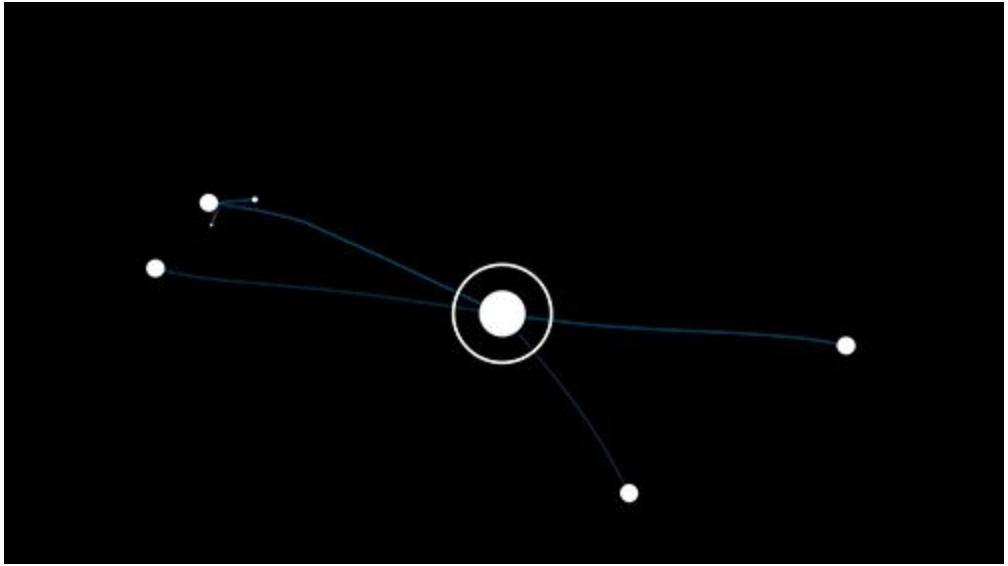
Theoretical Limitations

- K-hop neighbourhood
- WL test for isomorphism
- LOCAL/CONGEST

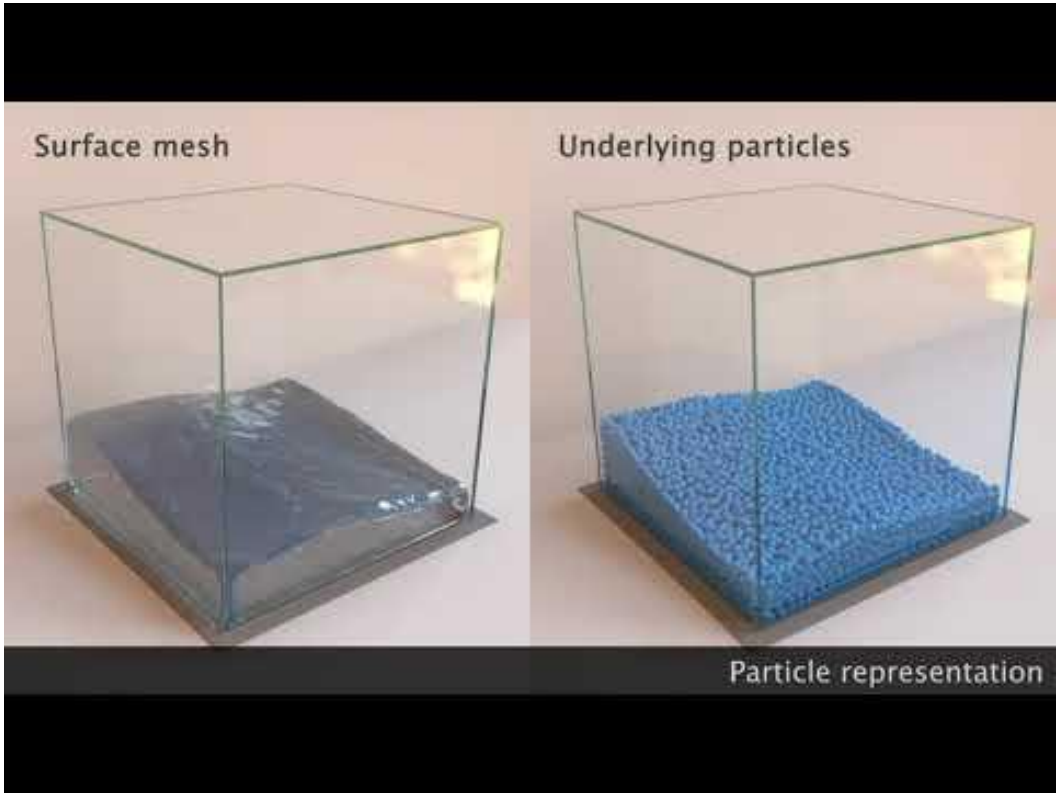
Why not do all-to-all message passing?

Oversmoothing

Graph Generation



Simulation


Surface mesh
Underlying particles
Particle representation

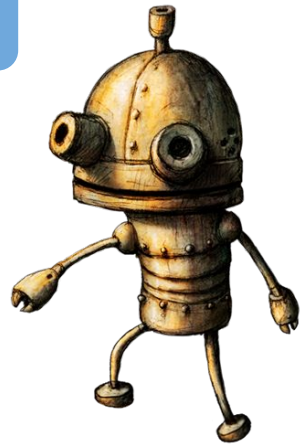
Algorithmic Learning

3 + 4 = ?

18467238957 + 67836423785 = ?

6.9809932742

😲?!!



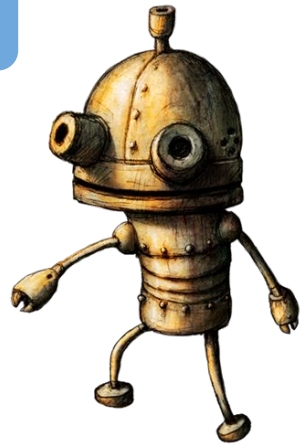
Algorithmic Learning

1 4 7 10 ?

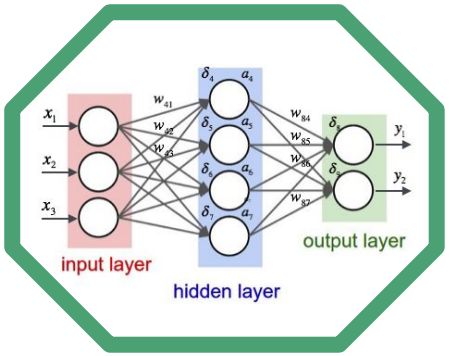
1 1 2 3 5 8 ?

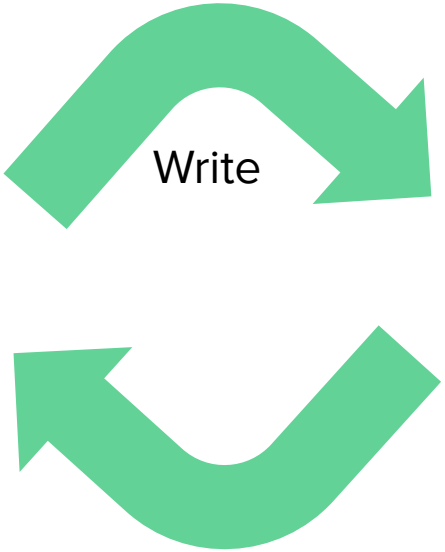
13

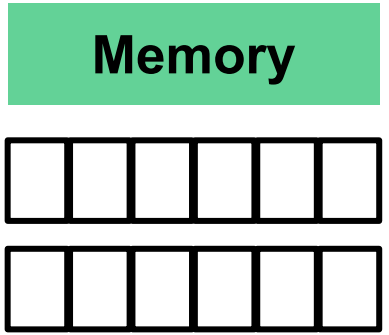
8?



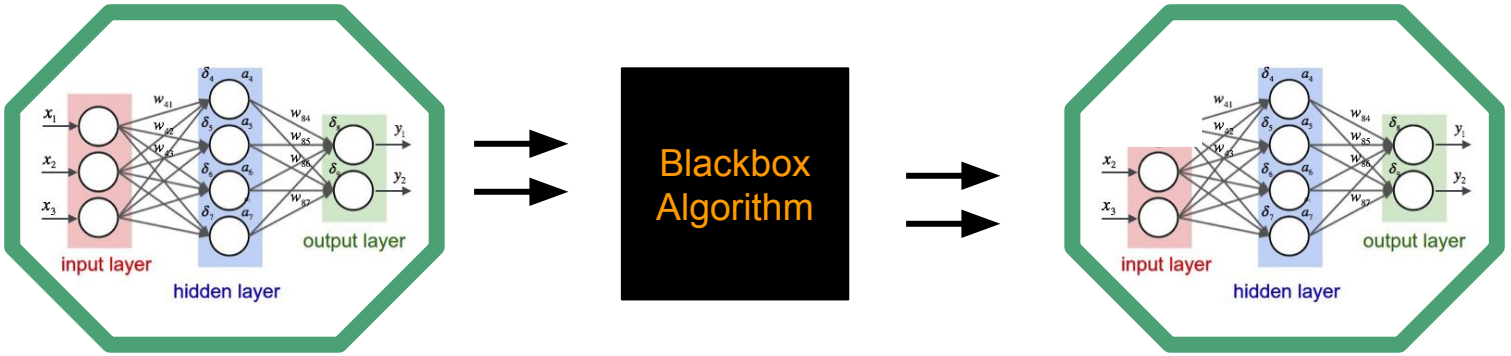
Augmenting NNs with Memory







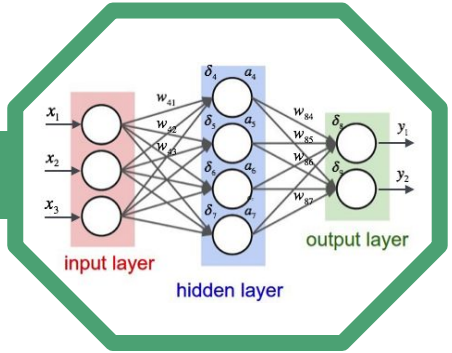
Augmenting NNs with Algorithms



Augmenting Algorithms with NNs

```
def knapsack(items, capacity):
  if len(items) == 0:
    return 0
  first, *rest = items
  take = 0


  skip = knapsack(rest, capacity)
  return max(take, skip)
```




"YOUR COUNTRY NEEDS
YOU"