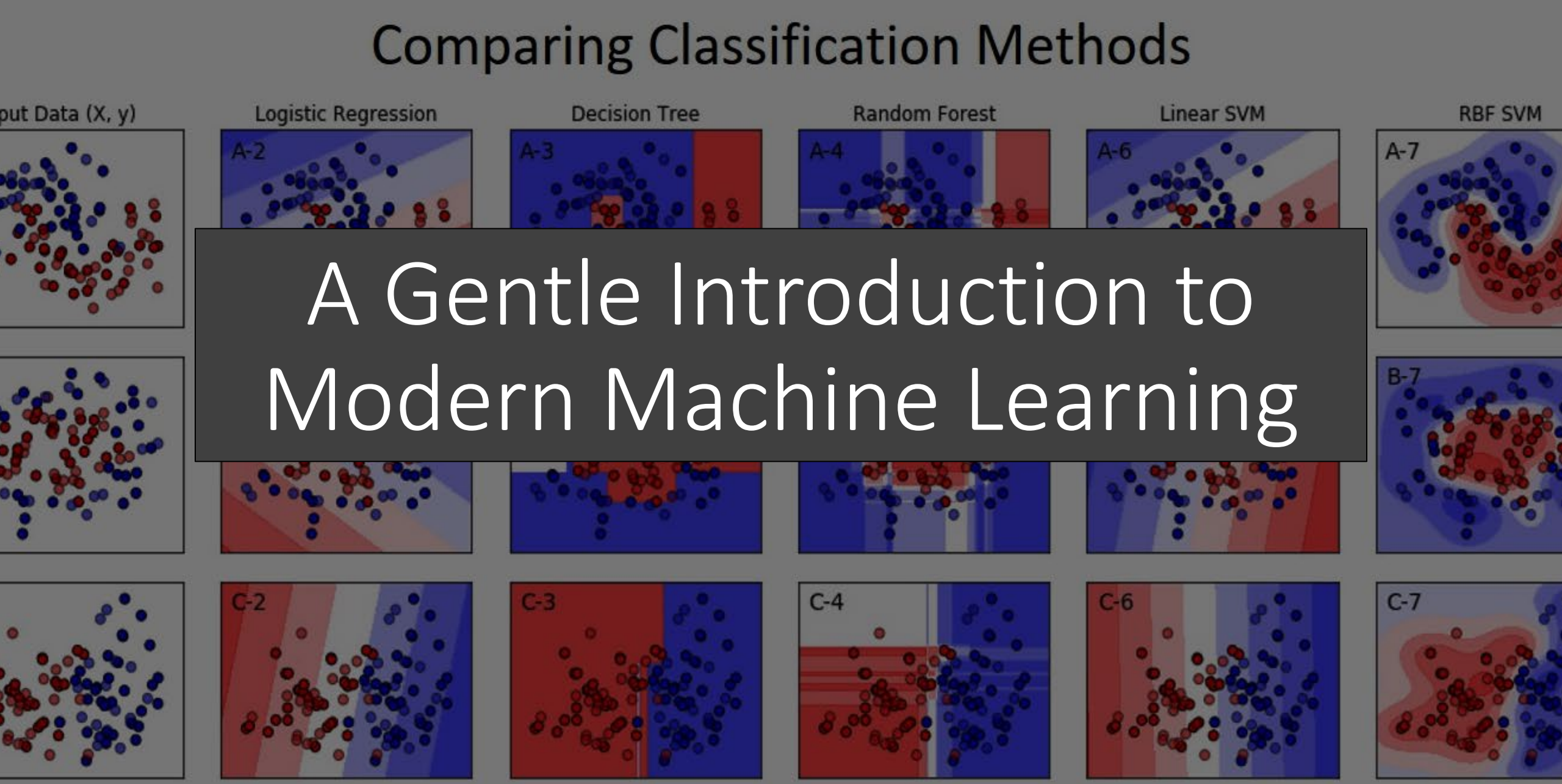
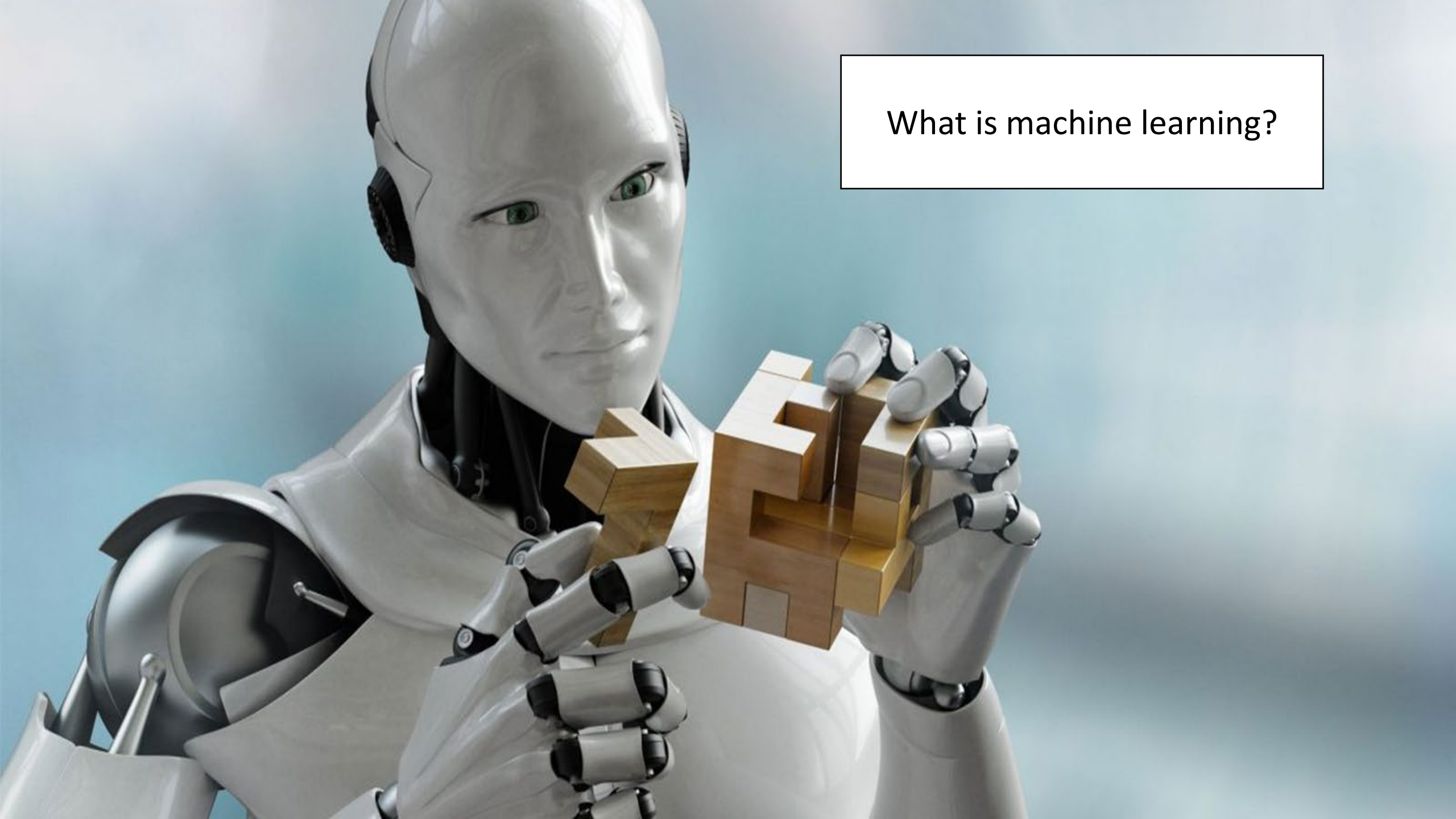
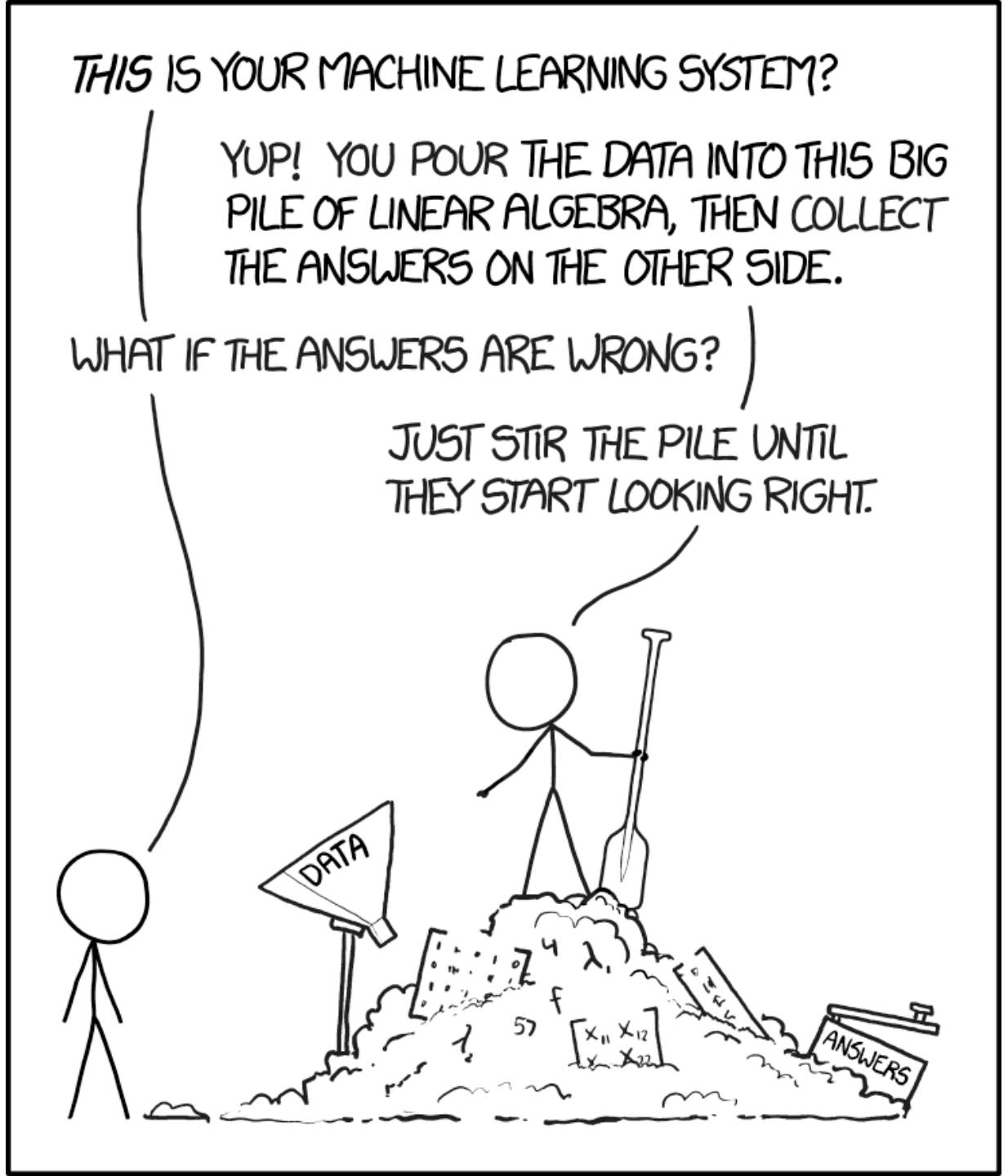


A Gentle Introduction to Modern Machine Learning



What is machine learning?


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

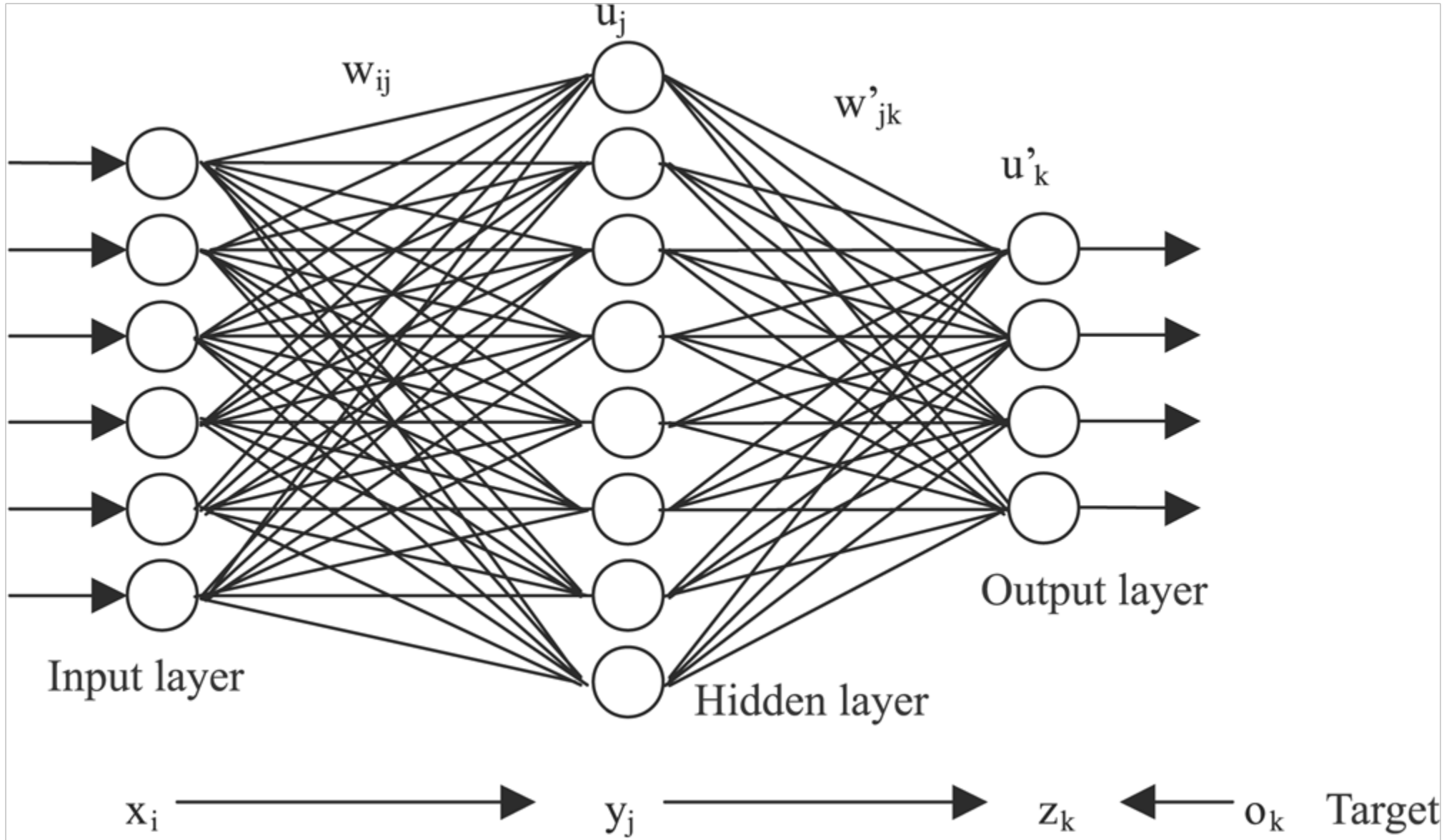
Machine Learning definition

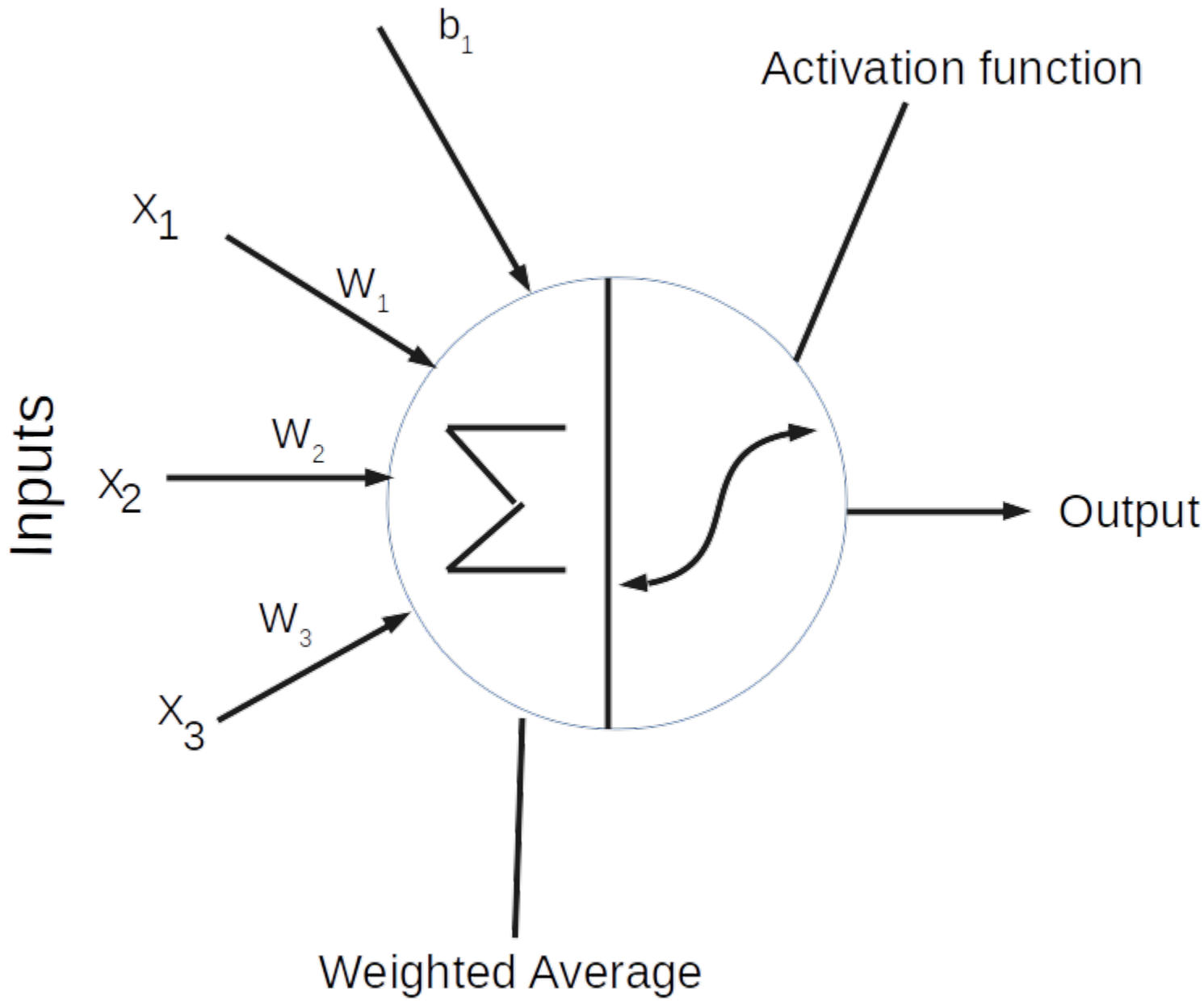
- Machine learning systems are those that can automatically improve their performance based on experience.

Supervised Learning

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.8	0.88	0	2.6	0.098	25	67	0.9968	3.2	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6
7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.4	0.66	0	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5
7.9	0.6	0.06	1.6	0.069	15	59	0.9964	3.3	0.46	9.4	5
7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	7
7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	7
7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5
6.7	0.58	0.08	1.8	0.097	15	65	0.9959	3.28	0.54	9.2	5
7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5
5.6	0.615	0	1.6	0.089	16	59	0.9943	3.58	0.52	9.9	5
7.8	0.61	0.29	1.6	0.114	9	29	0.9974	3.26	1.56	9.1	5
8.9	0.62	0.18	3.8	0.176	52	145	0.9986	3.16	0.88	9.2	5
8.9	0.62	0.19	3.9	0.17	51	148	0.9986	3.17	0.93	9.2	5
8.5	0.28	0.56	1.8	0.092	35	103	0.9969	3.3	0.75	10.5	7
8.1	0.56	0.28	1.7	0.368	16	56	0.9968	3.11	1.28	9.3	5
7.4	0.59	0.08	4.4	0.086	6	29	0.9974	3.38	0.5	9	4

Neural Networks

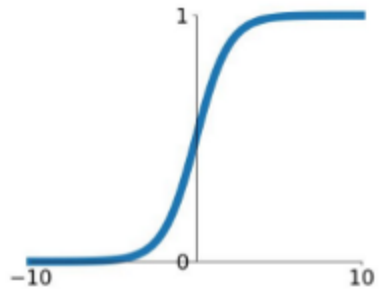

u_j
w_{ij}
w'_{jk}
u'_k
Input layer
Hidden layer
Output layer
x_i
y_j
z_k
o_k
Target


b1
Activation function
X1
W1
Inputs
W2
X2
Output
W3
X3
Weighted Average

Activation Functions

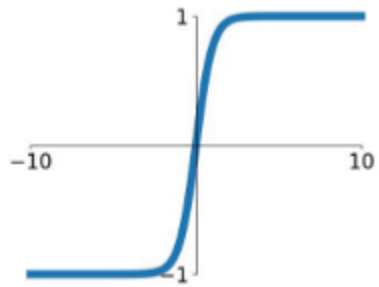
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



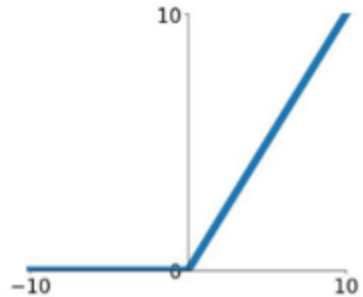
tanh

$$\tanh(x)$$



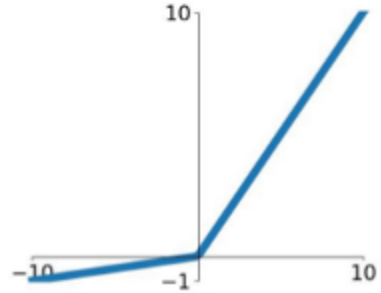
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

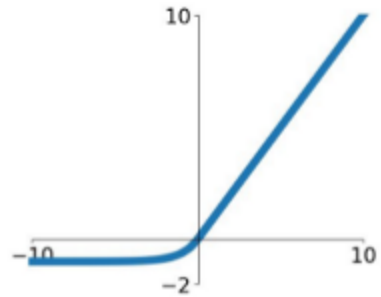


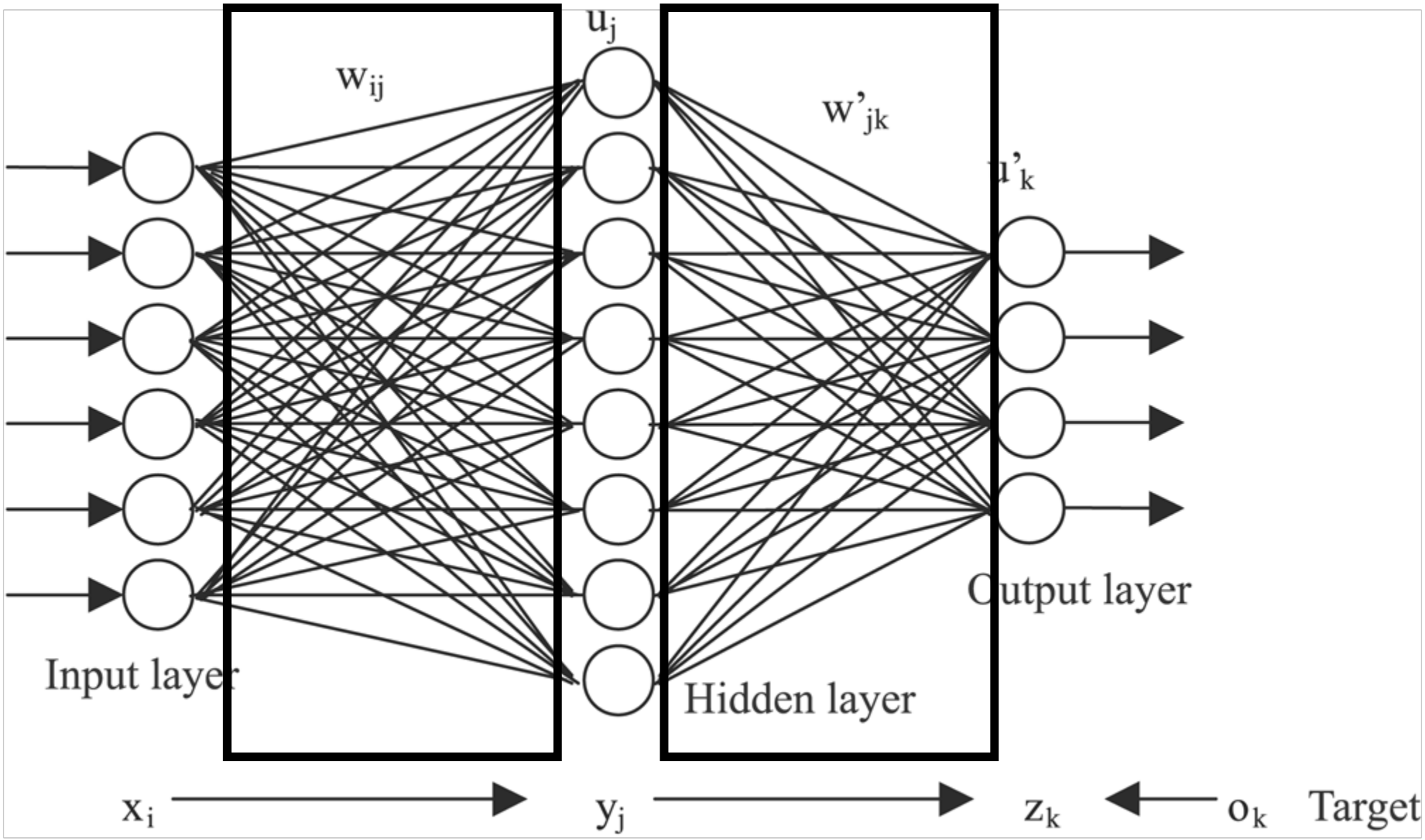
Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

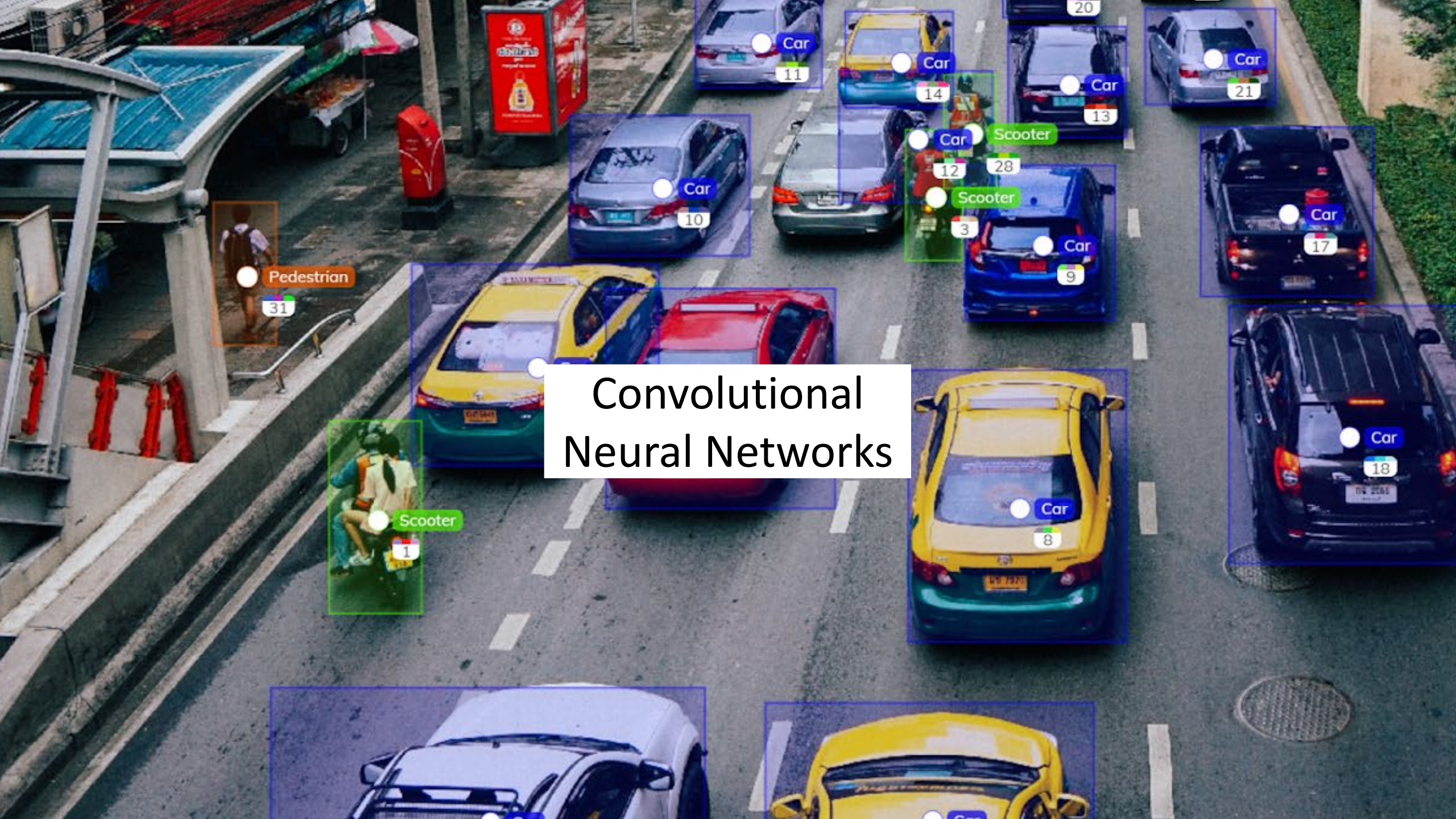
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

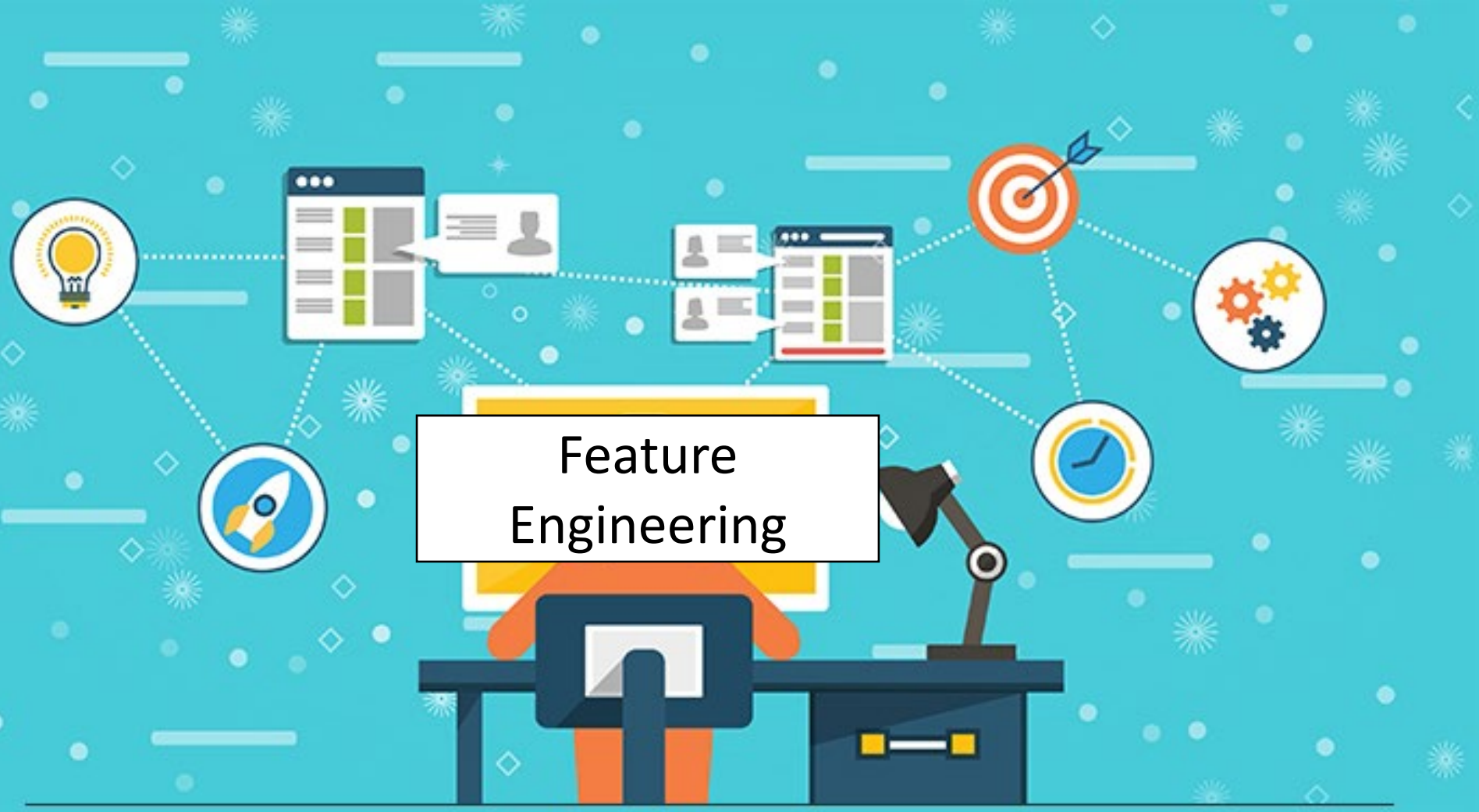



u_j
w_{ij}
w'_{jk}
u'_k
Input layer
Hidden layer
Output layer
x_i
y_j
z_k
o_k
Target

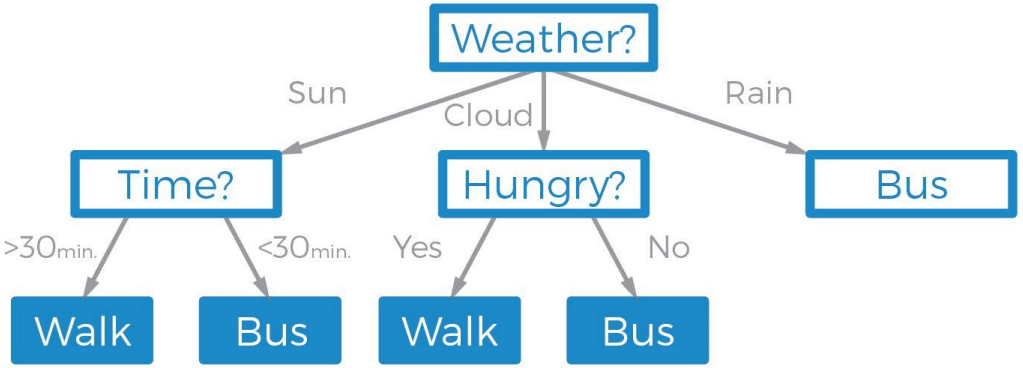

Pros and Cons



Convolutional Neural Networks


Feature
Engineering

Input

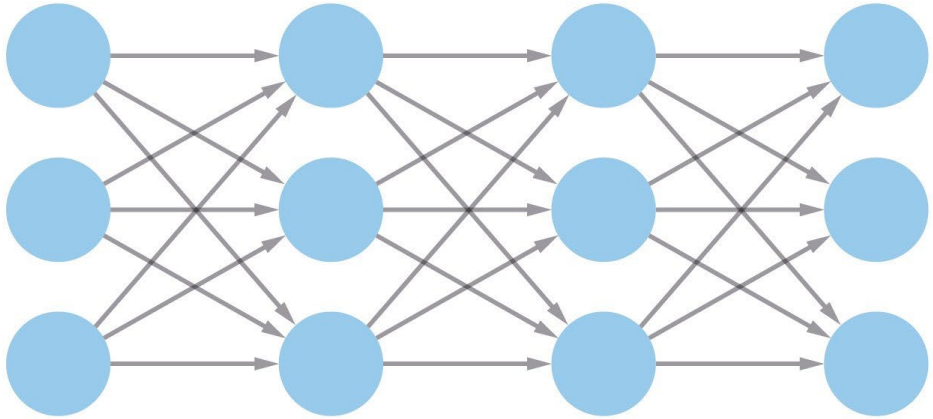


Decision tree

Output

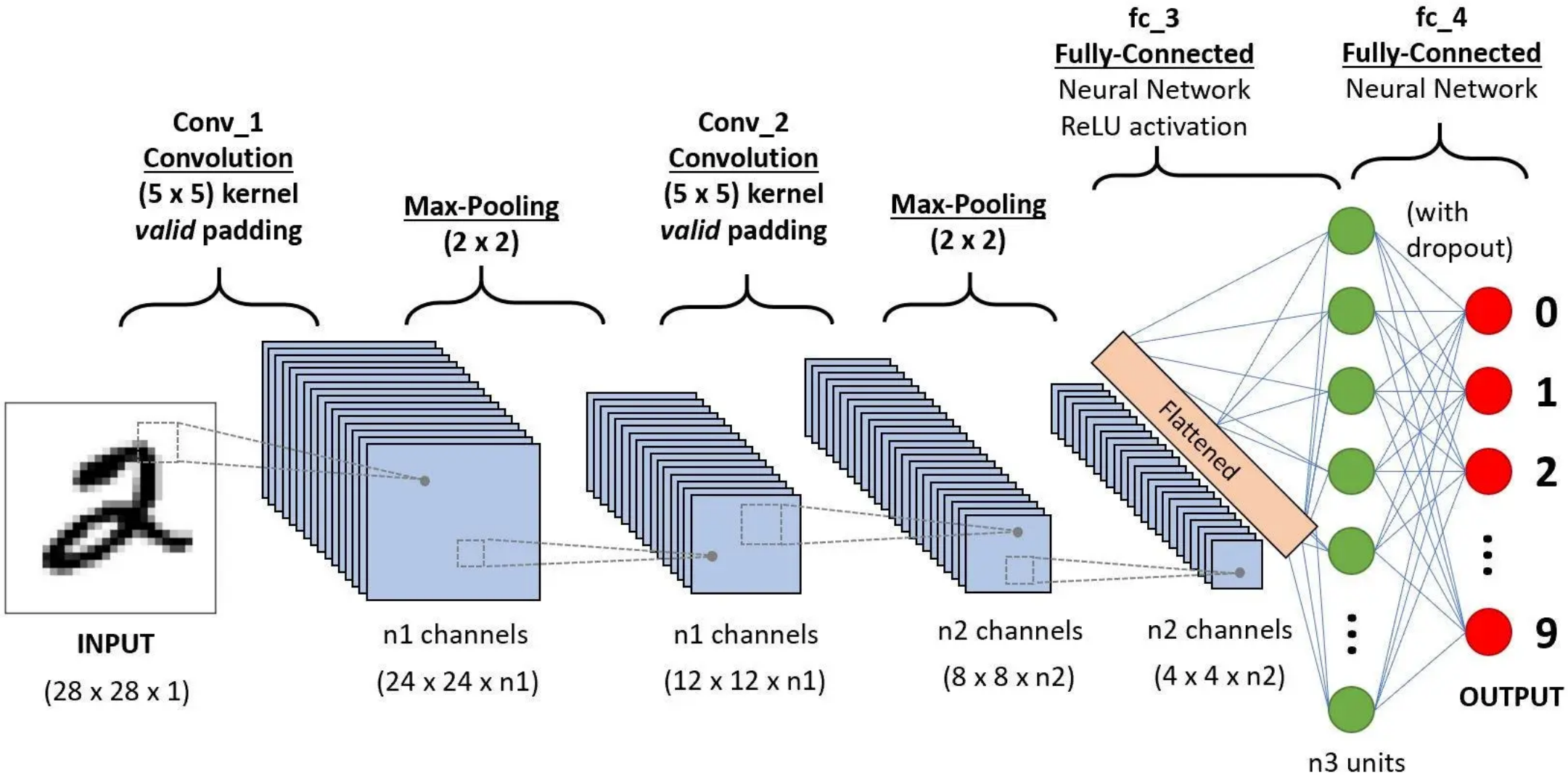
Deep Learning

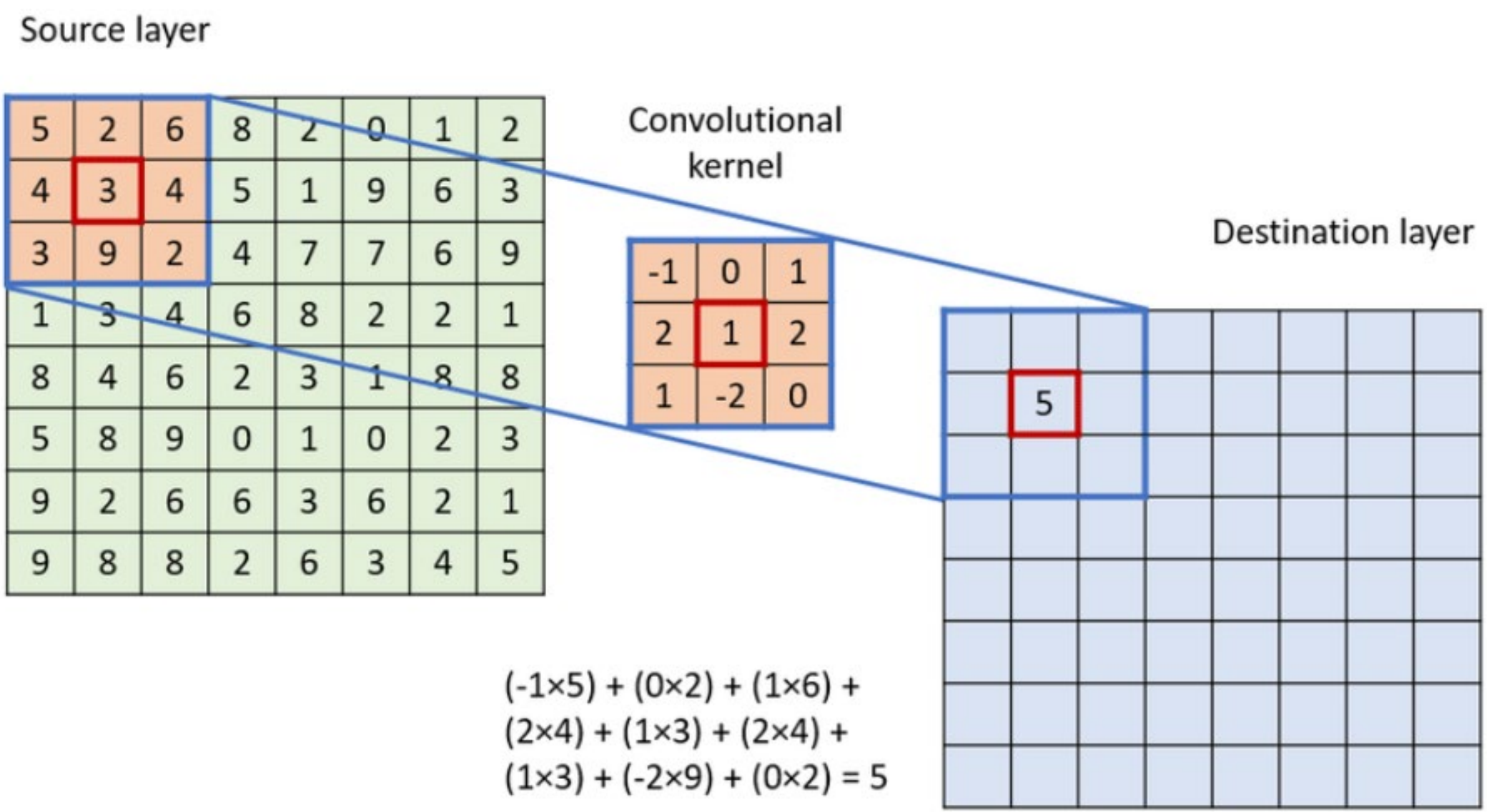
Input

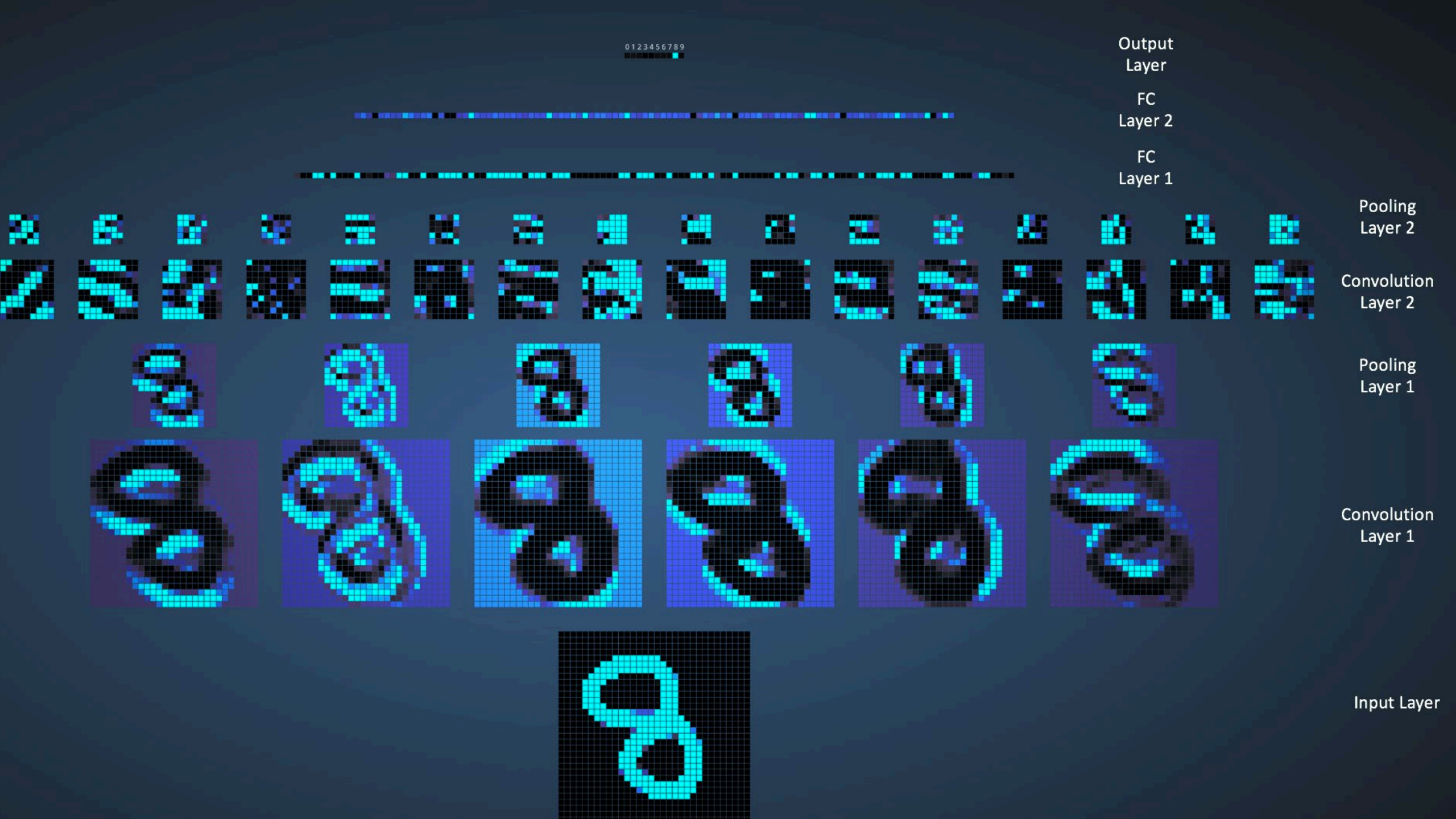


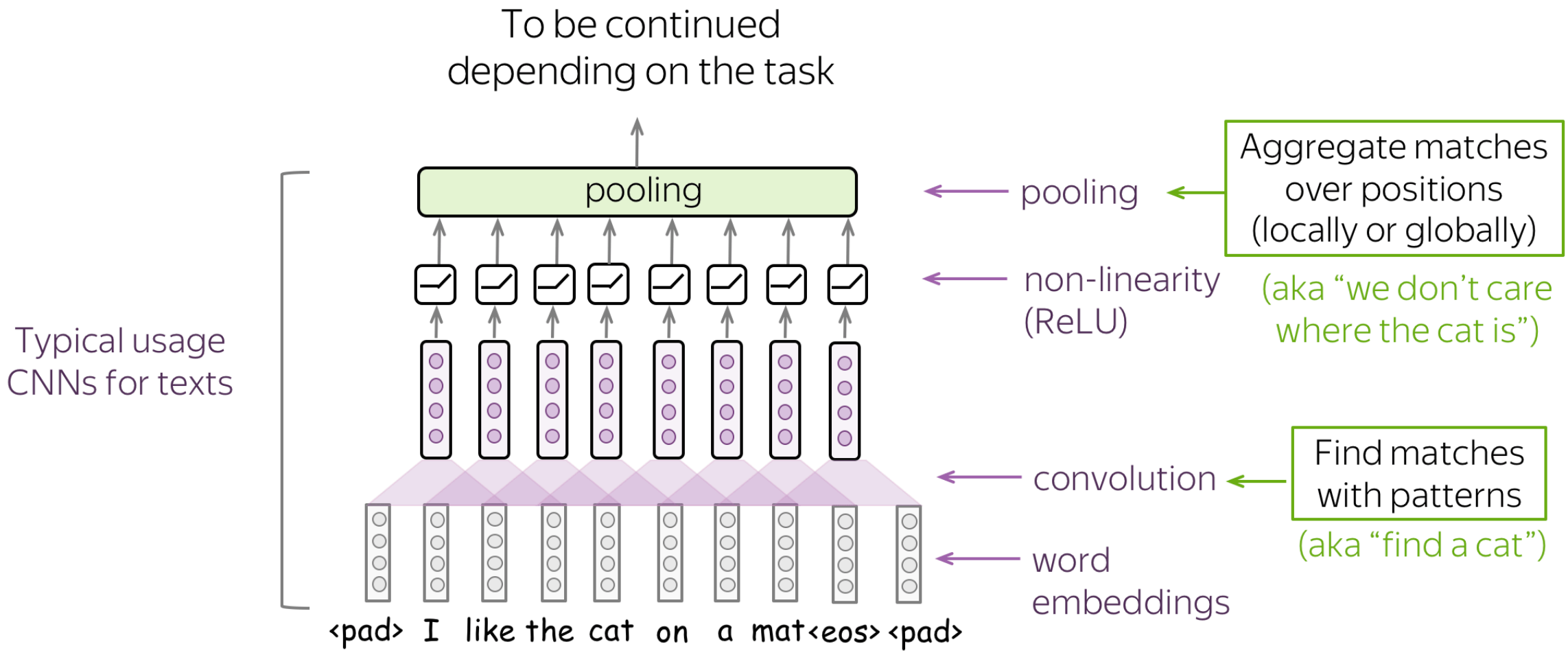
Feature extraction + Classification

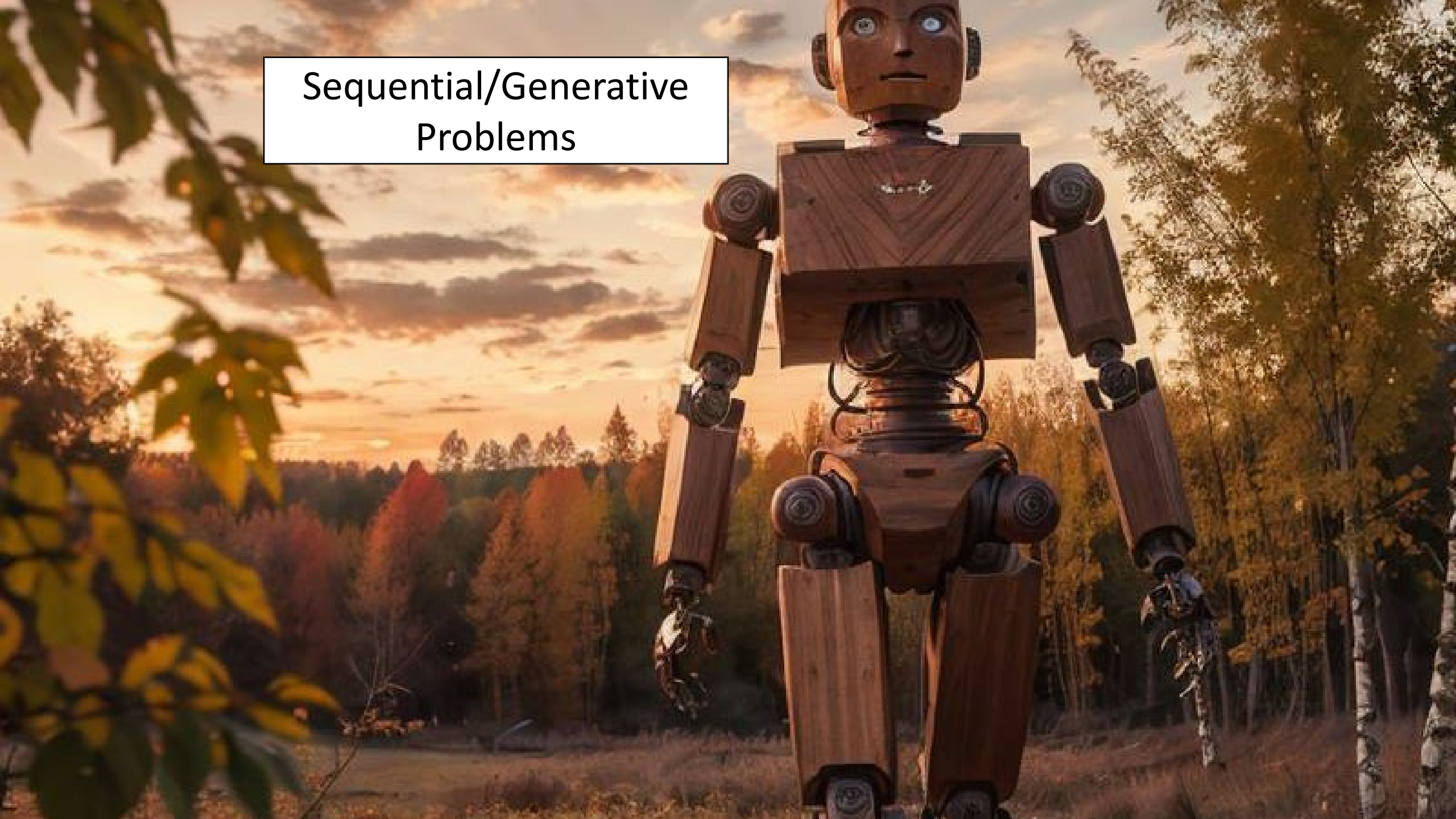
Output

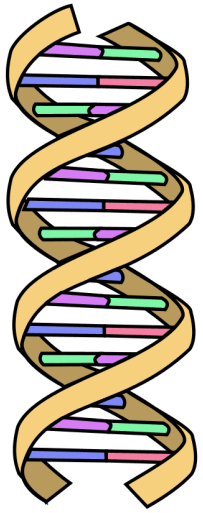

Conv_1
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
Conv_2
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
fc_3
Fully-Connected
Neural Network
ReLU activation
fc_4
Fully-Connected
Neural Network
(with dropout)
Flattened
INPUT
(28 x 28 x 1)
n1 channels
(24 x 24 x n1)
n1 channels
(12 x 12 x n1)
n2 channels
(8 x 8 x n2)
n2 channels
(4 x 4 x n2)
n3 units
0
1
2
9
OUTPUT


Source layer
5 2 6 8 2 0 1 2
4 3 4 5 1 9 6 3
3 9 2 4 7 7 6 9
1 3 4 6 8 2 2 1
8 4 6 2 3 1 8 8
5 8 9 0 1 0 2 3
9 2 6 6 3 6 2 1
9 8 8 2 6 3 4 5
Convolutional kernel
-1 0 1
2 1 2
1 -2 0
Destination layer
5
(-1×5) + (0×2) + (1×6) +
(2×4) + (1×3) + (2×4) +
(1×3) + (-2×9) + (0×2) = 5


0 1 2 3 4 5 6 7 8 9
Output Layer
FC Layer 2
FC Layer 1
Pooling Layer 2
Convolution Layer 2
Pooling Layer 1
Convolution Layer 1
Input Layer


To be continued
depending on the task
pooling
Typical usage
CNNs for texts
pooling
Aggregate matches
over positions
(locally or globally)
non-linearity
(ReLU)
(aka “we don’t care
where the cat is”)
convolution
Find matches
with patterns
(aka “find a cat”)
word
embeddings
<pad> I like the cat on a mat<eos> <pad>


Sequential/Generative
Problems



DNA

=Adenine

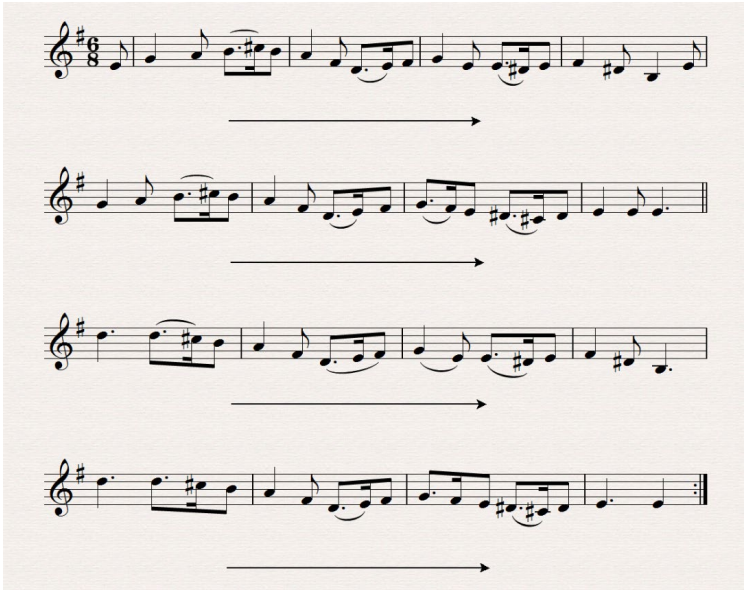
= Thymine

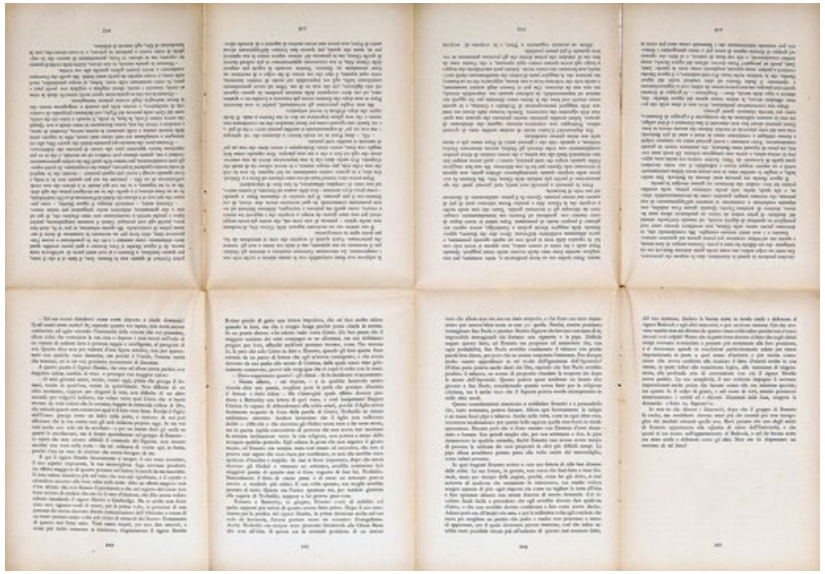
= Cytosine

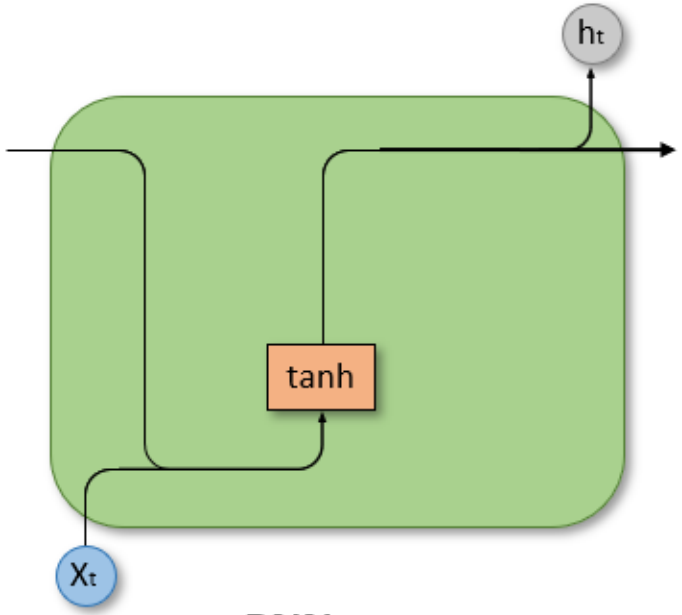
= Guanine

= Phosphate backbone

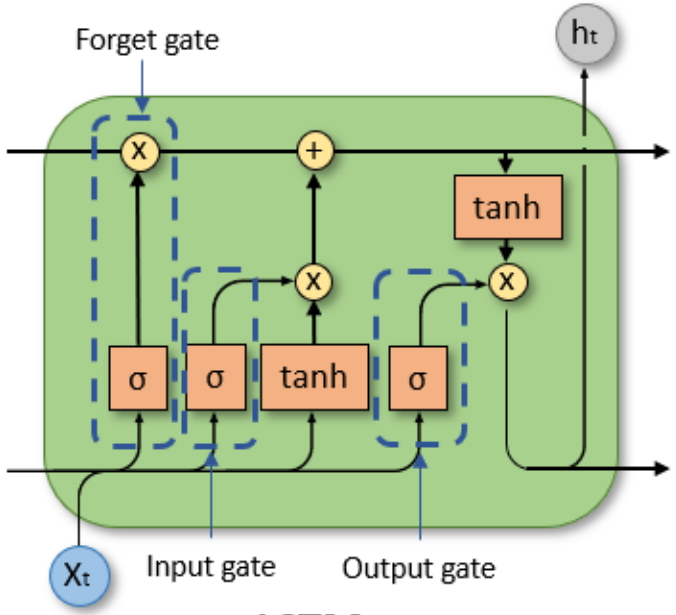
Sequential Data



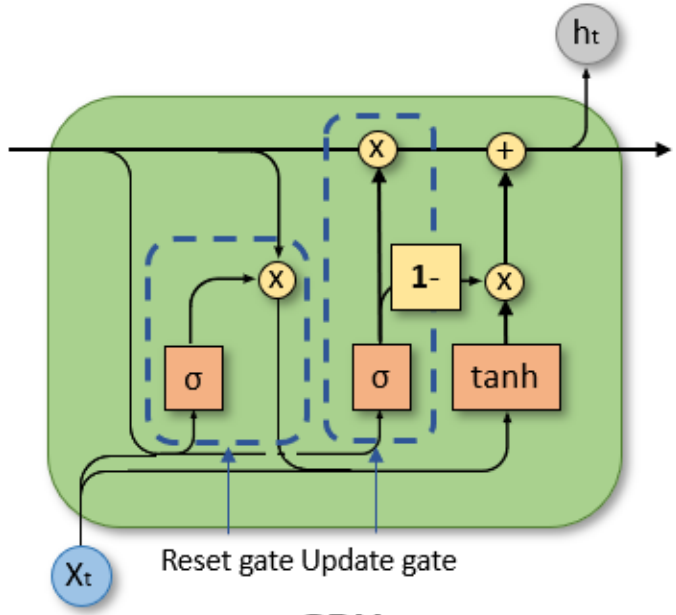




RNN

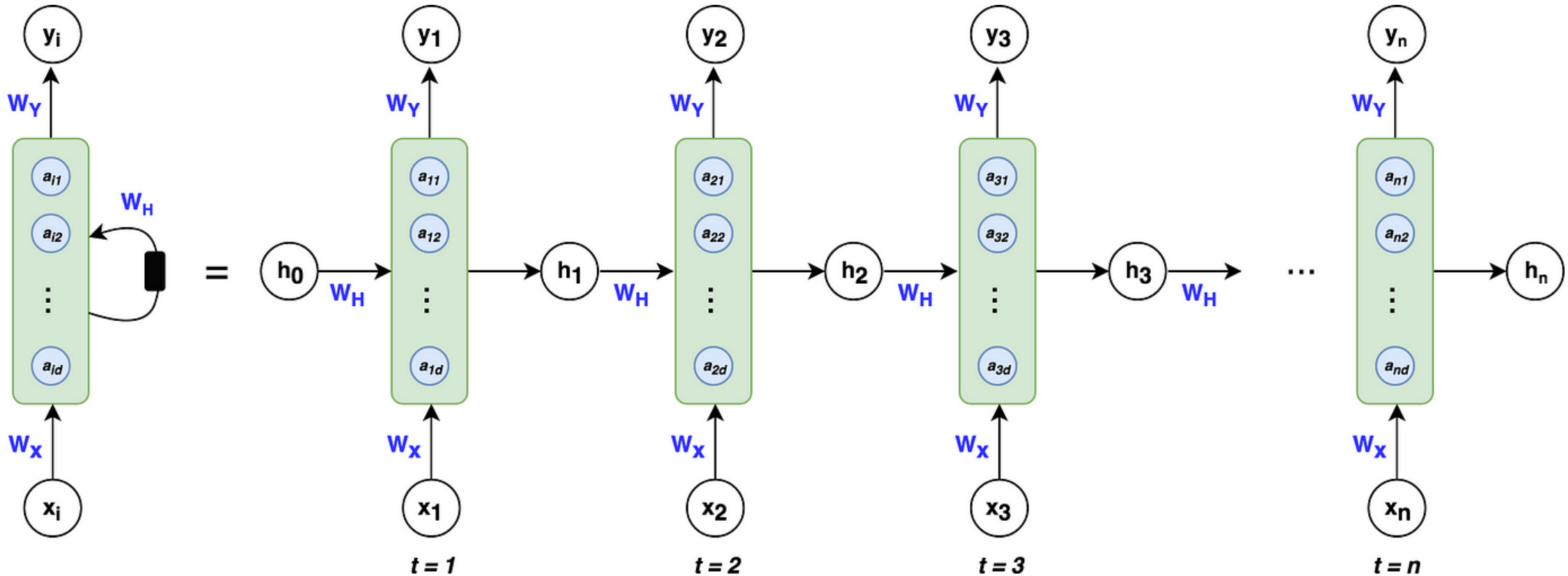


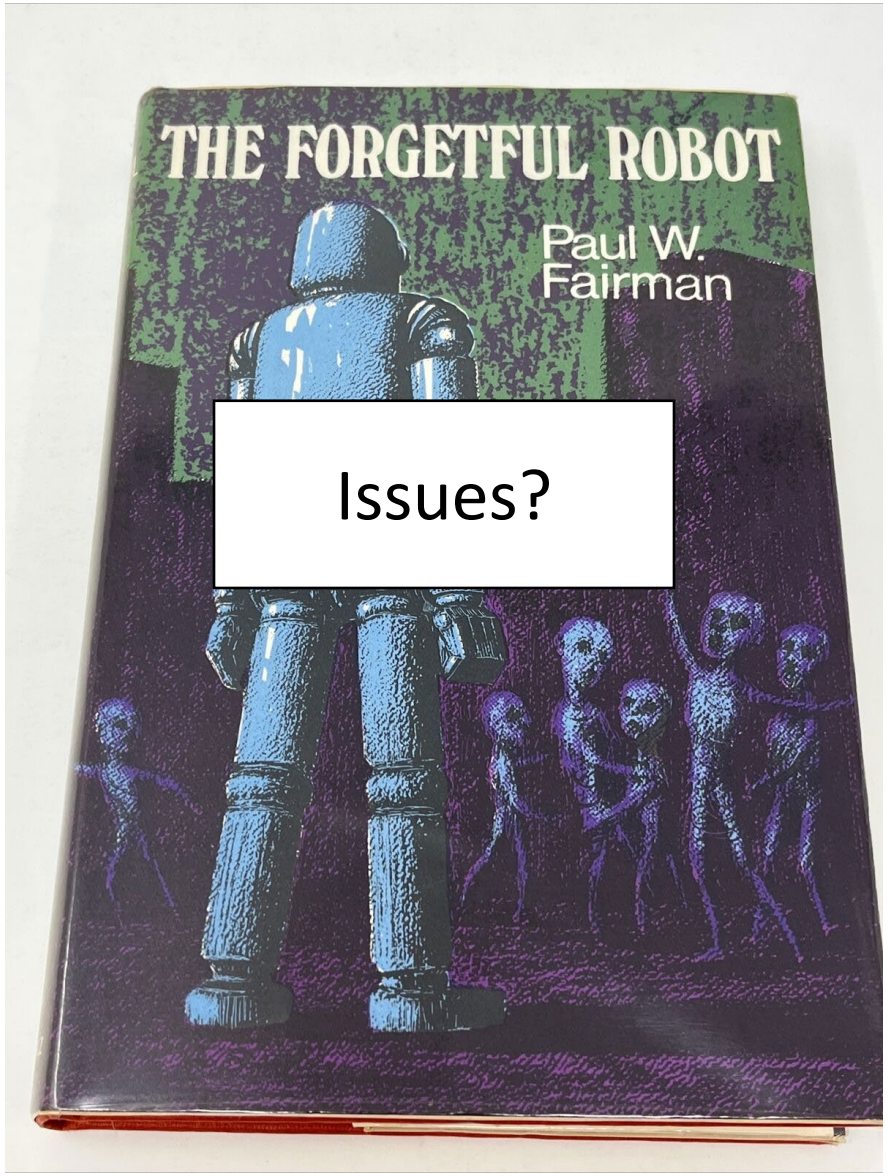
LSTM

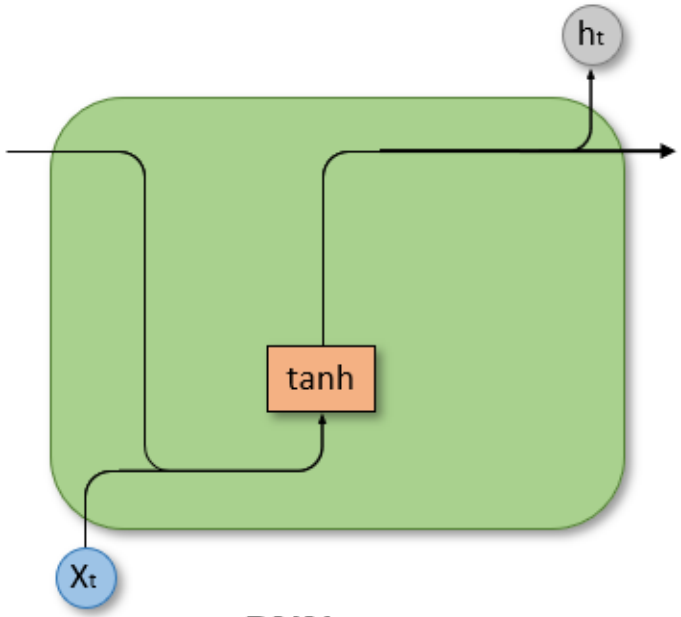


GRU

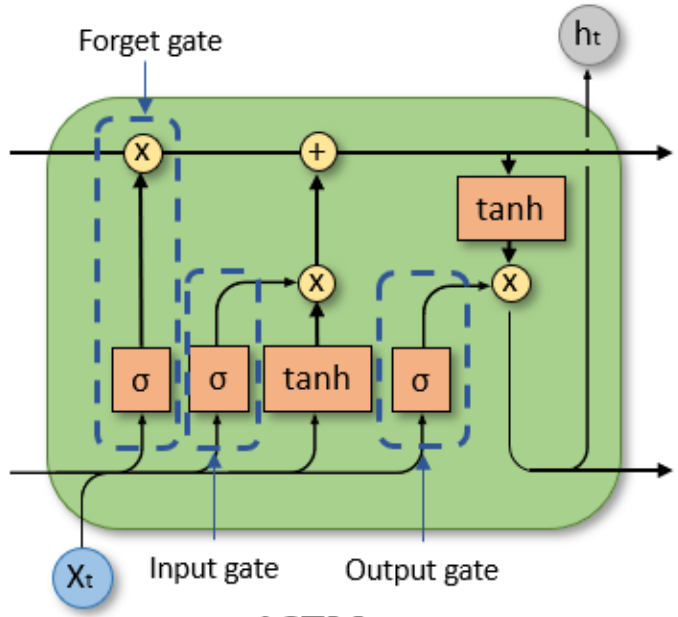
Recurrent Neural Networks


y_i
W_Y
a_{i1}
a_{i2}
a_{id}
W_H
W_X
x_i
=
h_0
W_H
y_1
W_Y
a_{11}
a_{12}
a_{1d}
W_X
x_1
$t = 1$
h_1
W_H
y_2
W_Y
a_{21}
a_{22}
a_{2d}
W_X
x_2
$t = 2$
h_2
W_H
y_3
W_Y
a_{31}
a_{32}
a_{3d}
W_X
x_3
$t = 3$
h_3
W_H
...
y_n
W_Y
a_{n1}
a_{n2}
a_{nd}
W_X
x_n
$t = n$
h_n

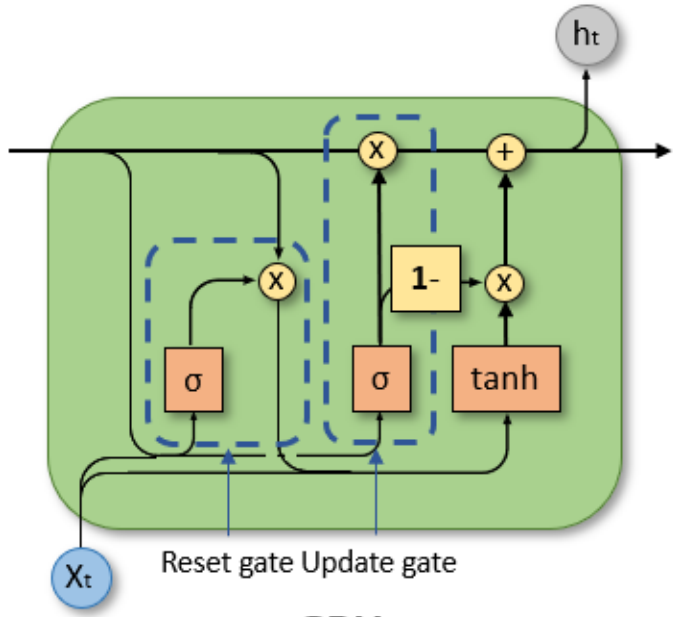

THE FORGETFUL ROBOT
Paul W. Fairman
Issues?


ht
tanh
Xt

RNN


Forget gate
ht
x
+
tanh
x
x
σ
σ
tanh
σ
Xt
Input gate
Output gate

LSTM

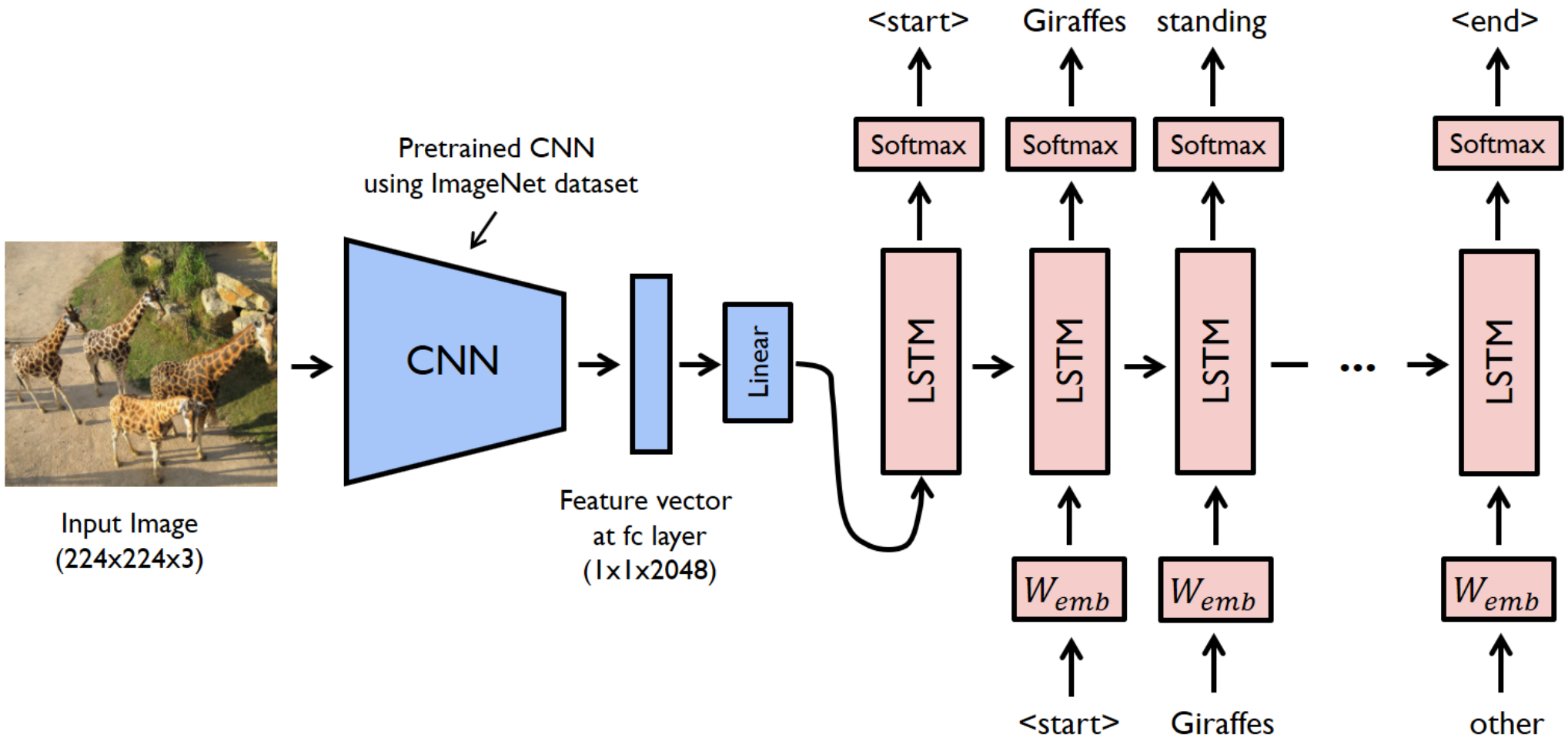

ht
x
+
x
1-
x
σ
σ
tanh
Xt
Reset gate
Update gate

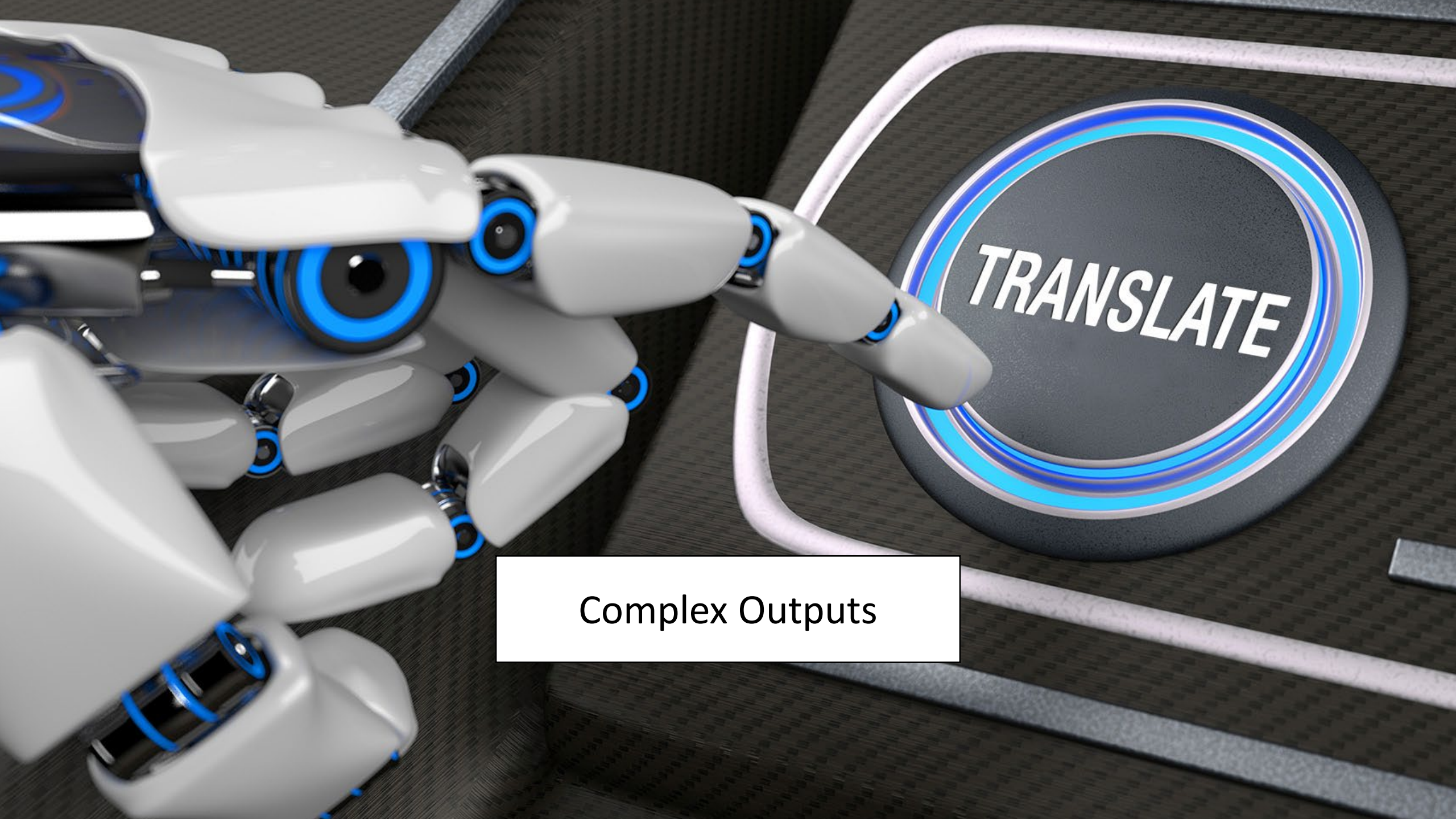
GRU

Captioning Model

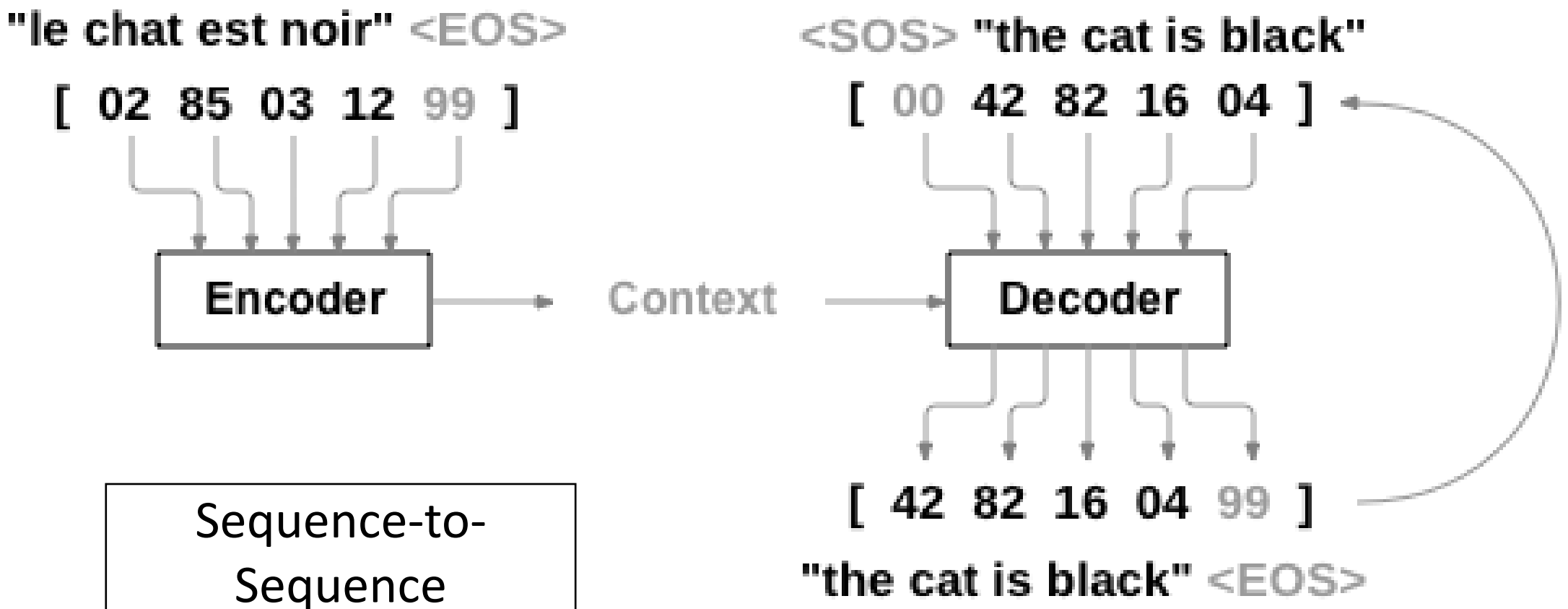
A happy dog is standing in the ocean

Enables Captioning

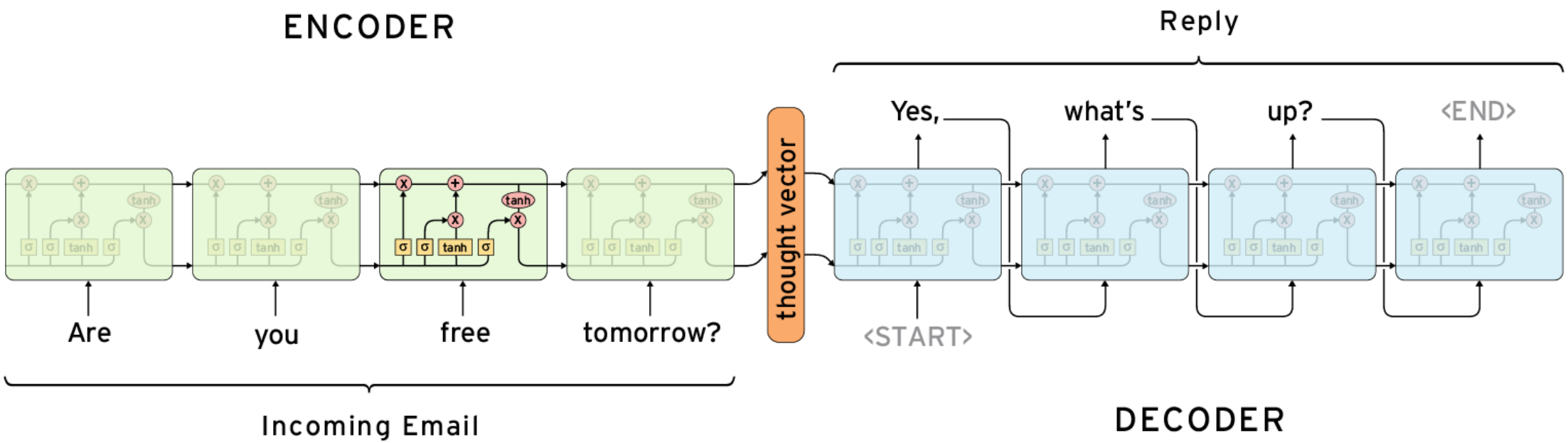

Pretrained CNN
using ImageNet dataset
Input Image
(224x224x3)
CNN
Feature vector
at fc layer
(1x1x2048)
Linear
LSTM
LSTM
LSTM
LSTM
Softmax
Softmax
Softmax
Softmax
<start>
Giraffes
standing
<end>
W_{emb}
W_{emb}
W_{emb}
<start>
Giraffes
other

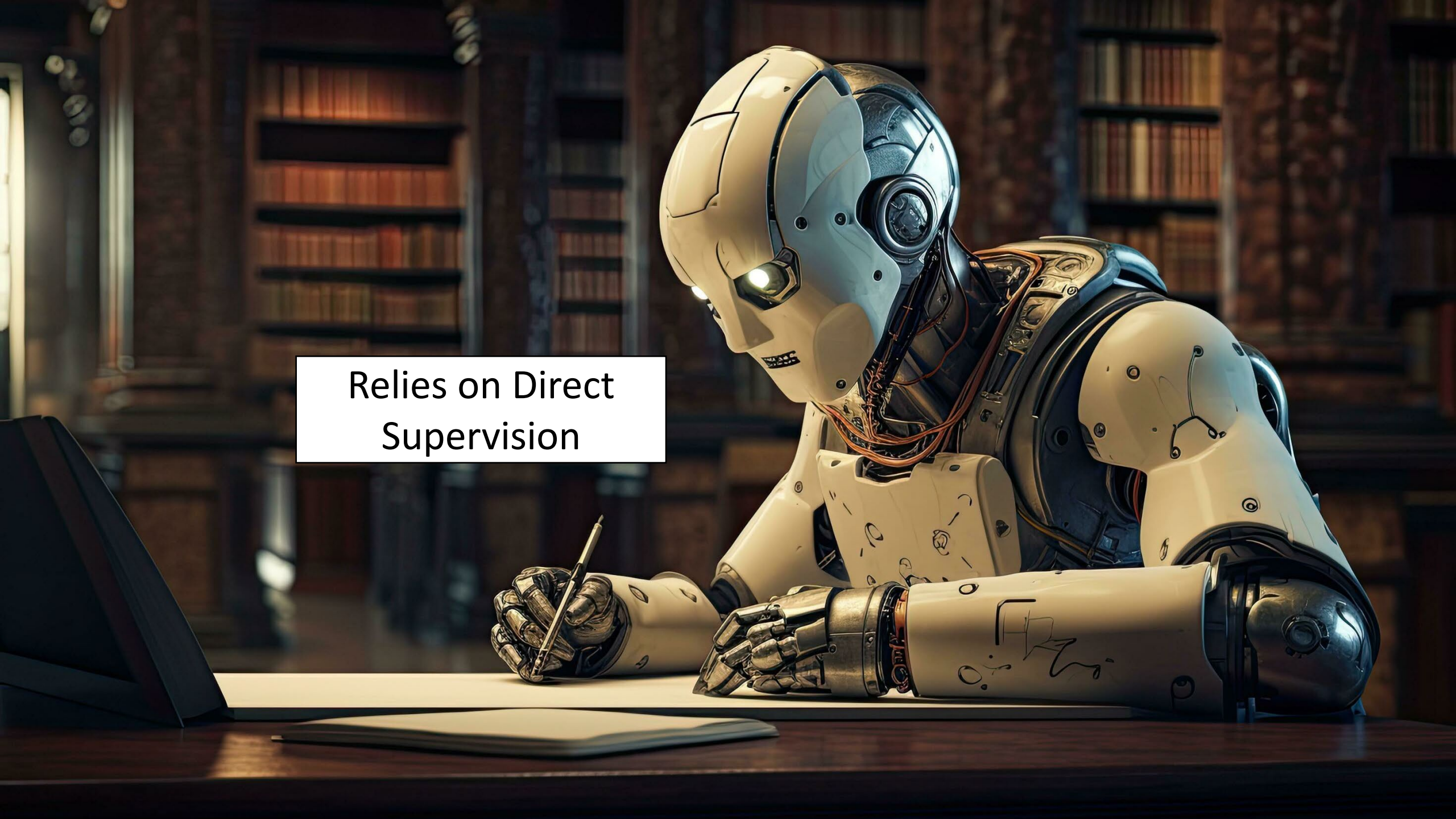


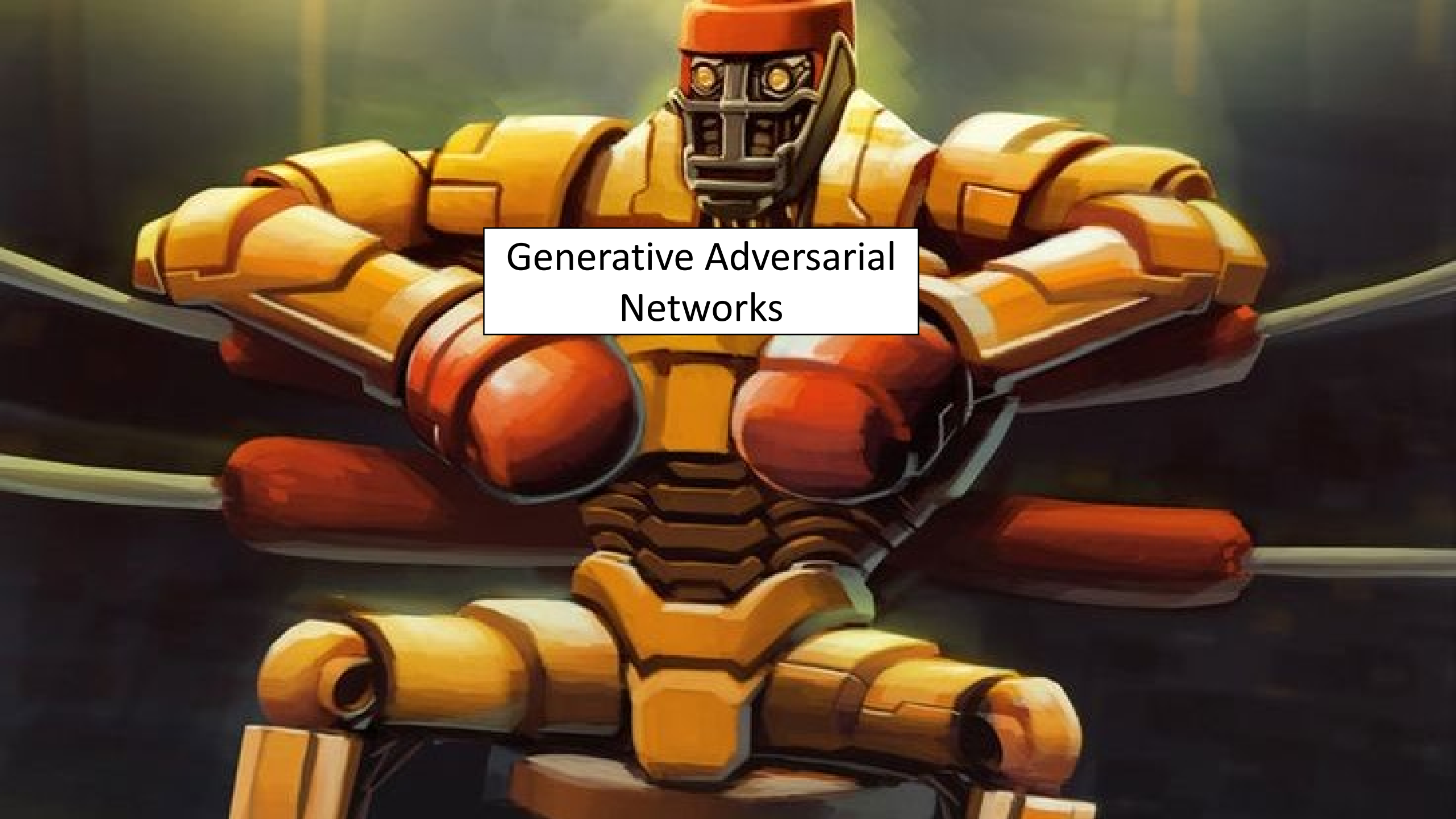
Complex Outputs



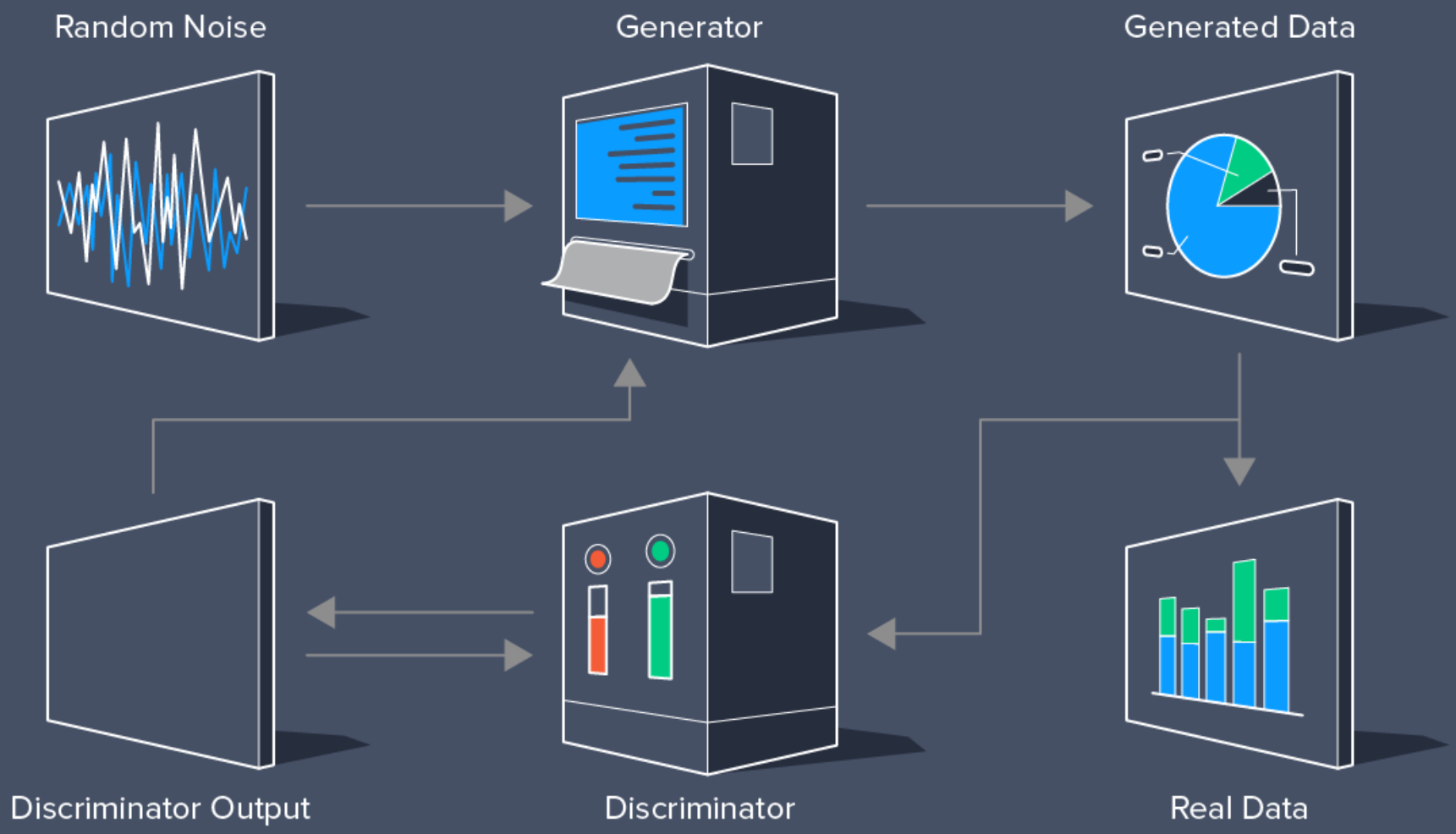
Sequence-to-Sequence


ENCODER
Reply
Yes,
what's
up?
<END>
thought vector
Are
you
free
tomorrow?
<START>
Incoming Email
DECODER

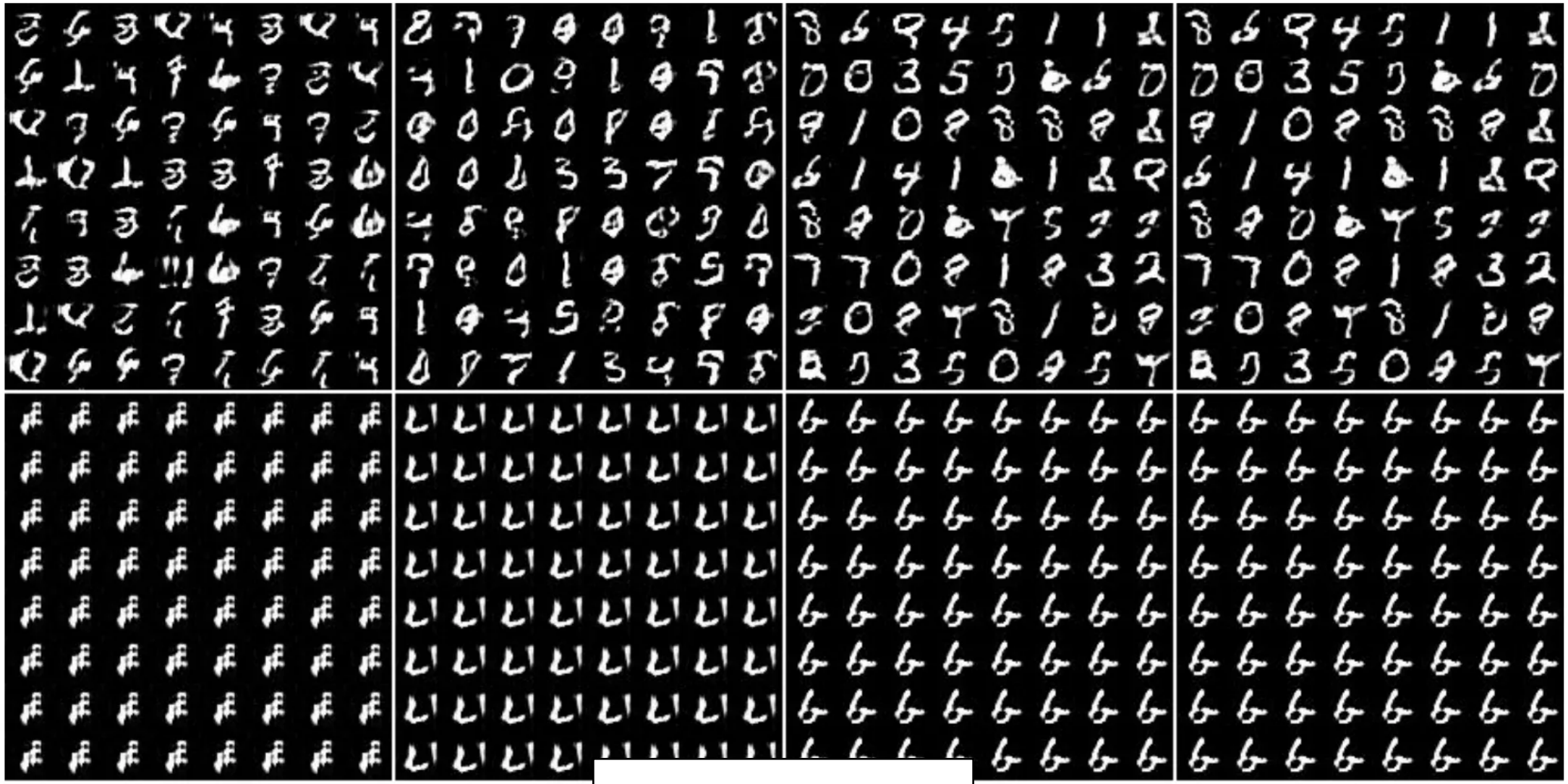

Relies on Direct Supervision


Generative Adversarial Networks

Prompt: robot playing against magnus carlsen in chess


Random Noise
Generator
Generated Data
Discriminator Output
Discriminator
Real Data

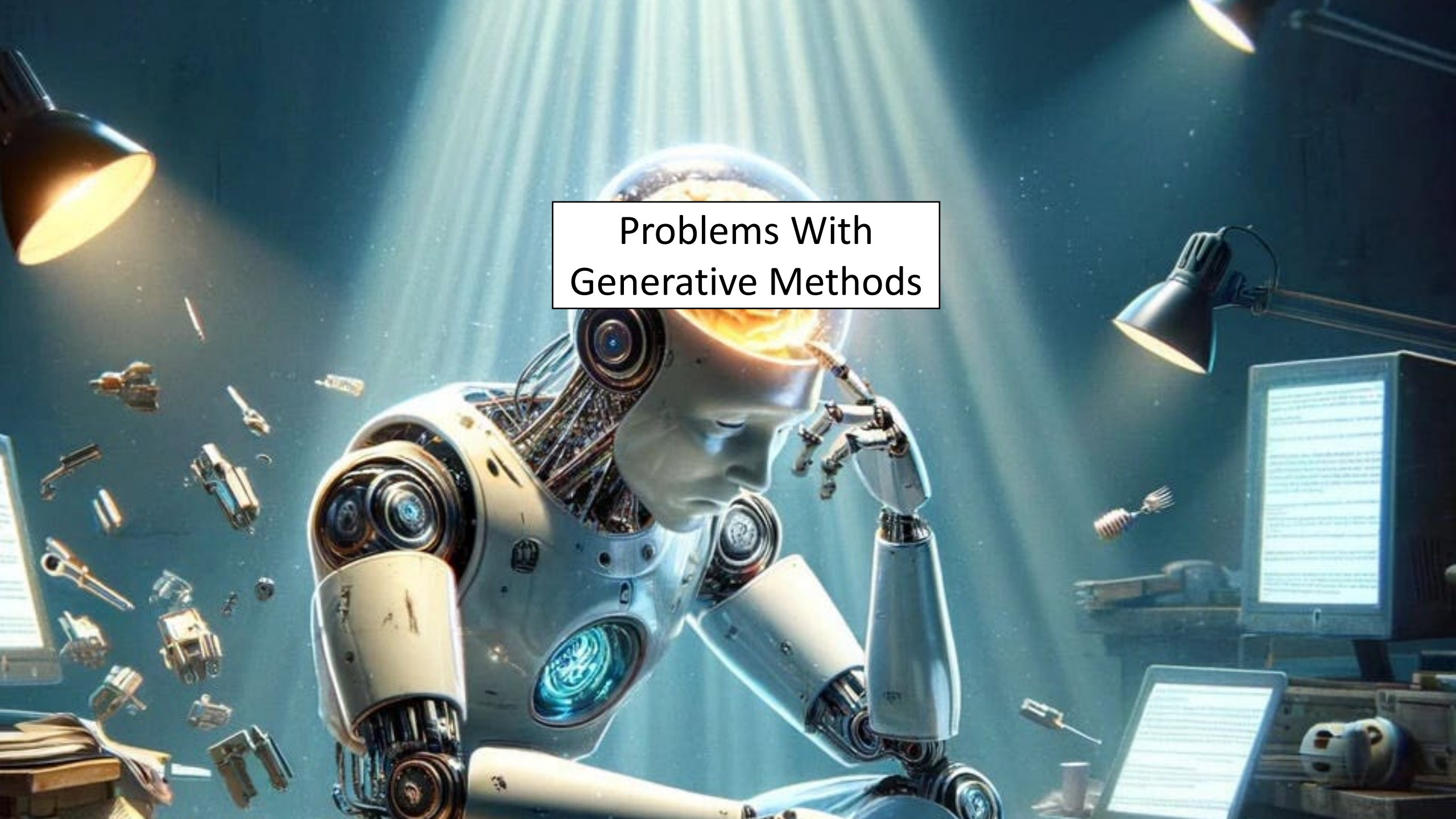
Minimax Interpretation



Problems

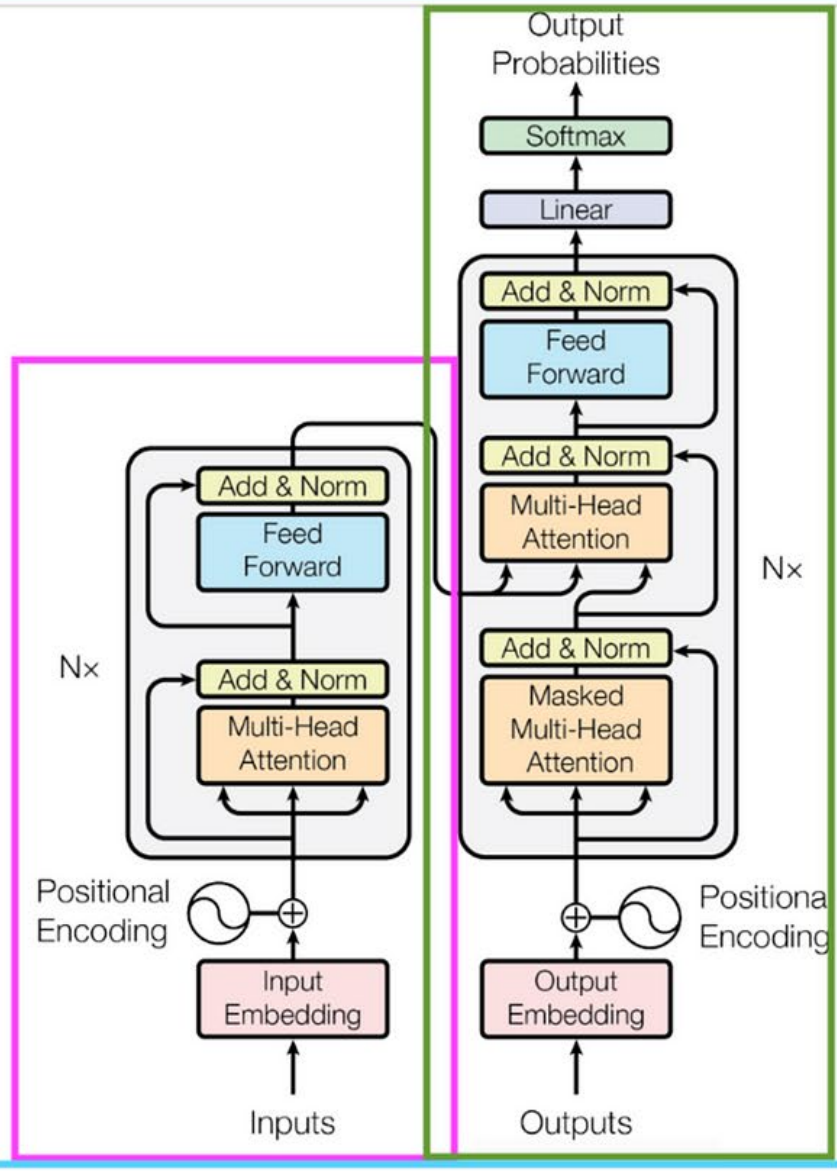

Resolving
Problems

Transformers

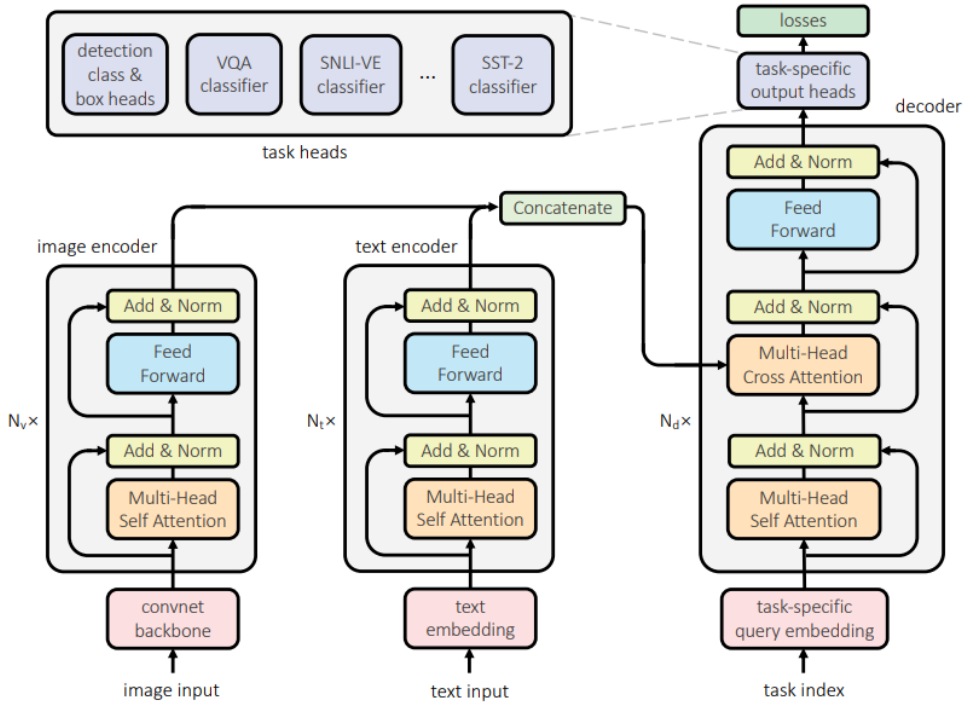

Problems With
Generative Methods

Encoder-Decoder
T5
BART

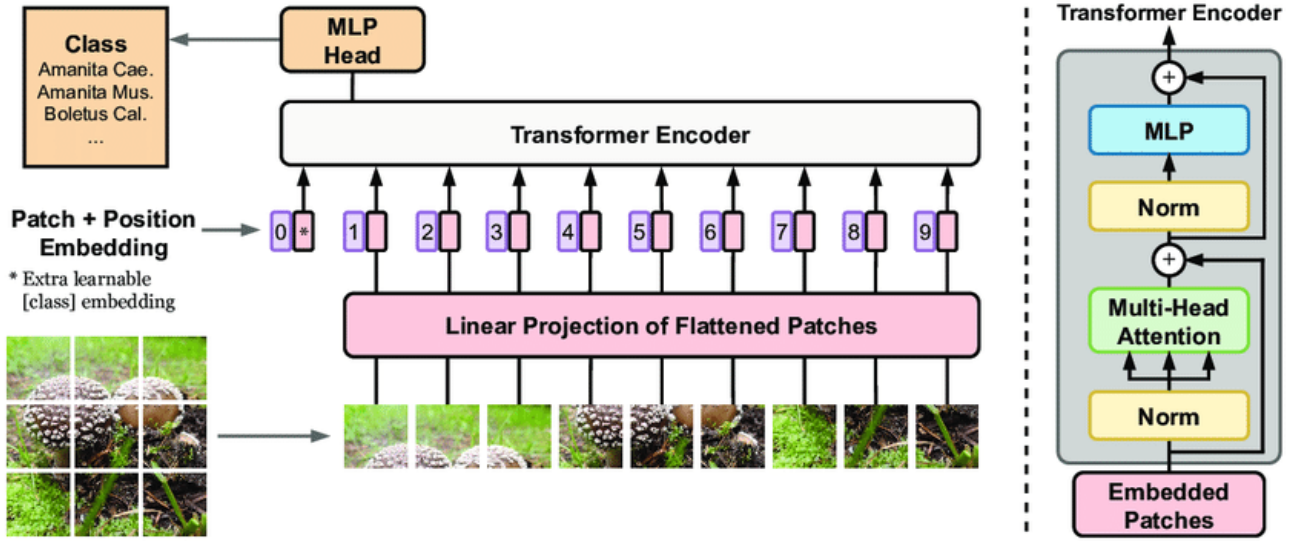
Encoder-only
BERT
ROBERTA



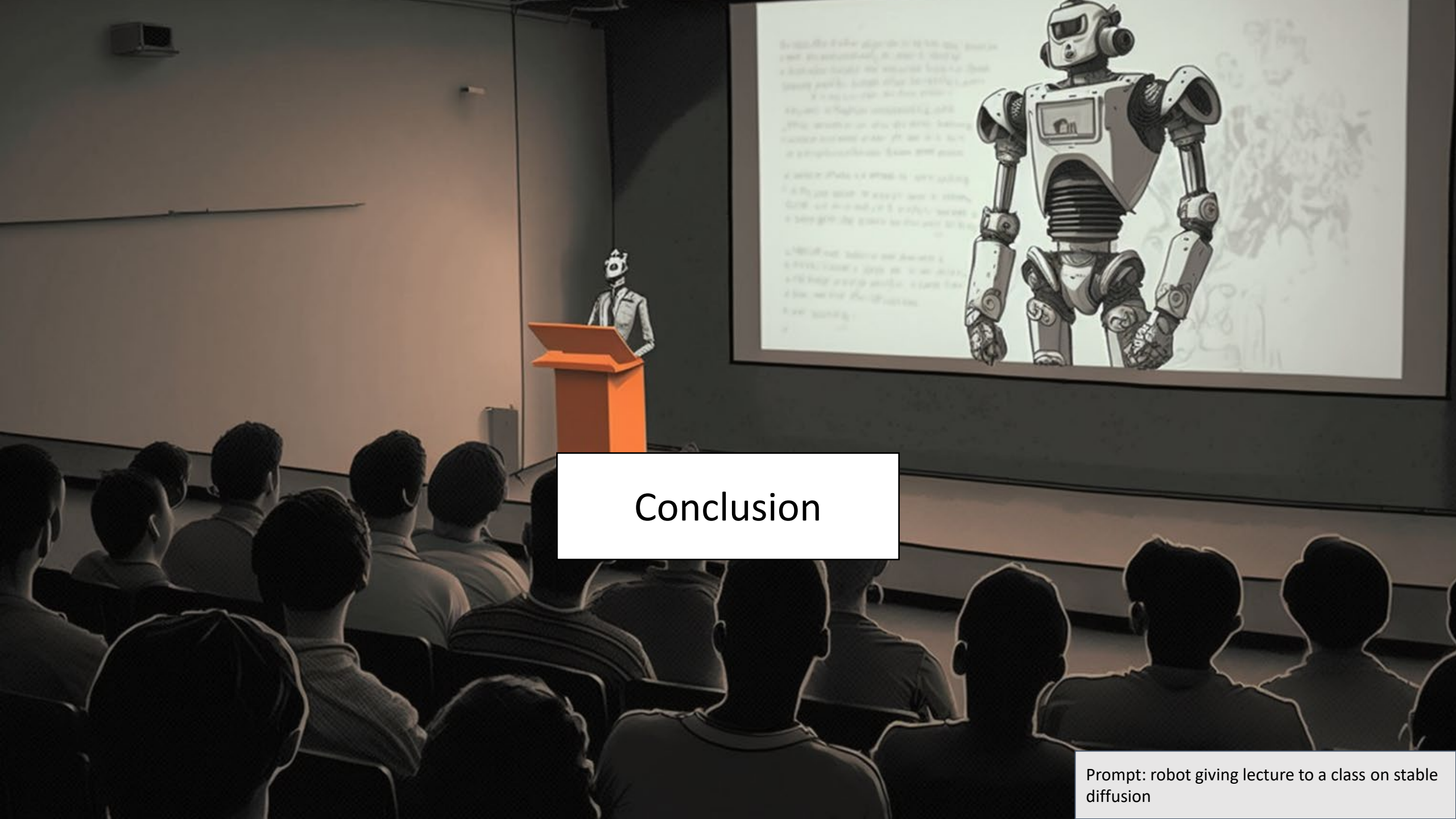
Decoder-only
GPT
BLOOM



Multimodal Transformer



Visual Transformer (ViT)



Conclusion

Prompt: robot giving lecture to a class on stable diffusion