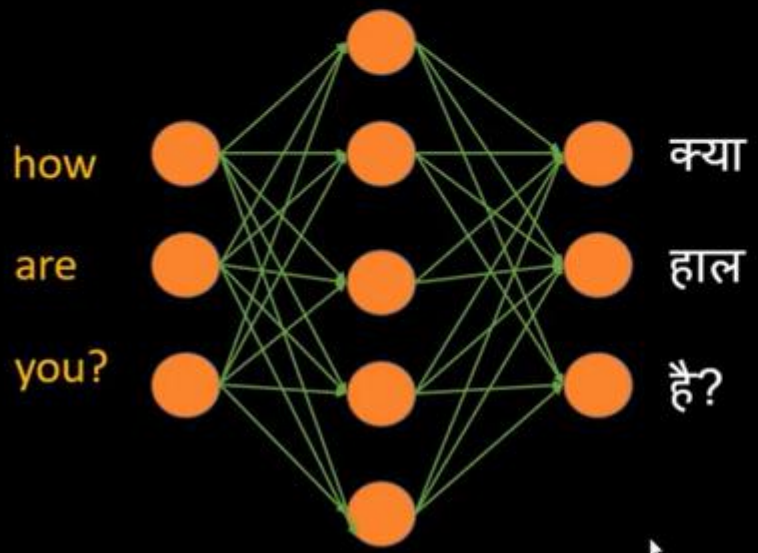
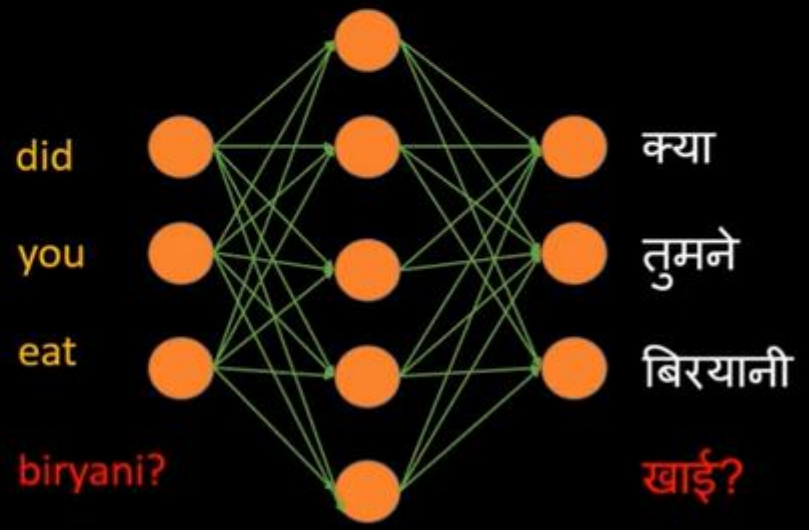
RNN- Recurrent Neural Networks (RNNs)

USE CASES

	x		y
auto complete	not interested at	→	this time
translation	how are you?	→	क्या हाल है?
NER	Rudolph Smith bought 1000 shares of tesla Inc. in March 2020	→	Person: Rudolph Smith bought 1000 shares of Company: tesla Inc. in time: March 2020
Sentiment Analysis	Not only the fan was expensive, but it was broken when it arrived.	→	★☆☆☆☆

Issue # 1: No fixed size of neurons in a layer





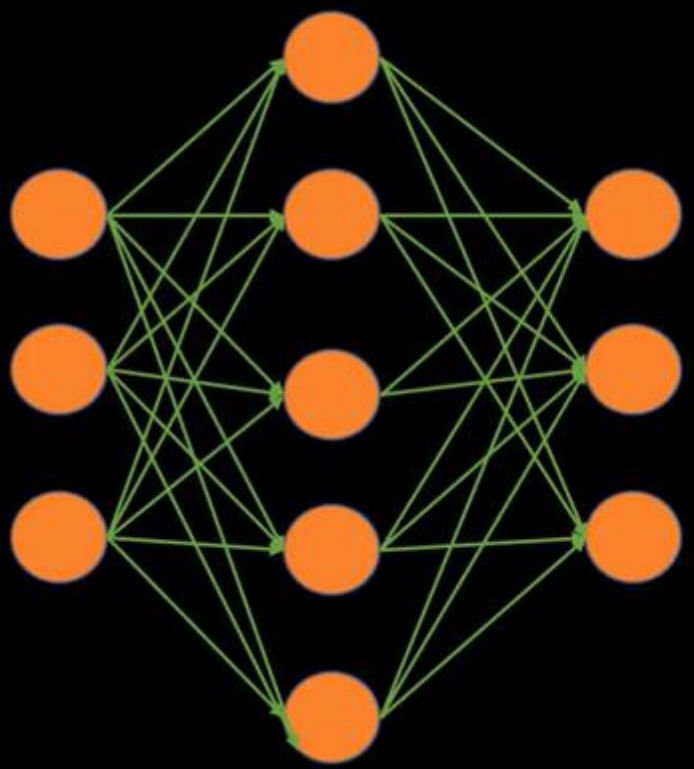
Issue # 2: Too much computation

25000 words in vocabulary

how → [0,0,0,..,1,0,0,...,0]

are → [0, 1,0,0,0,..0,0,...,0]

you? → [0, 0,0,0,..0,0,1,0,0]



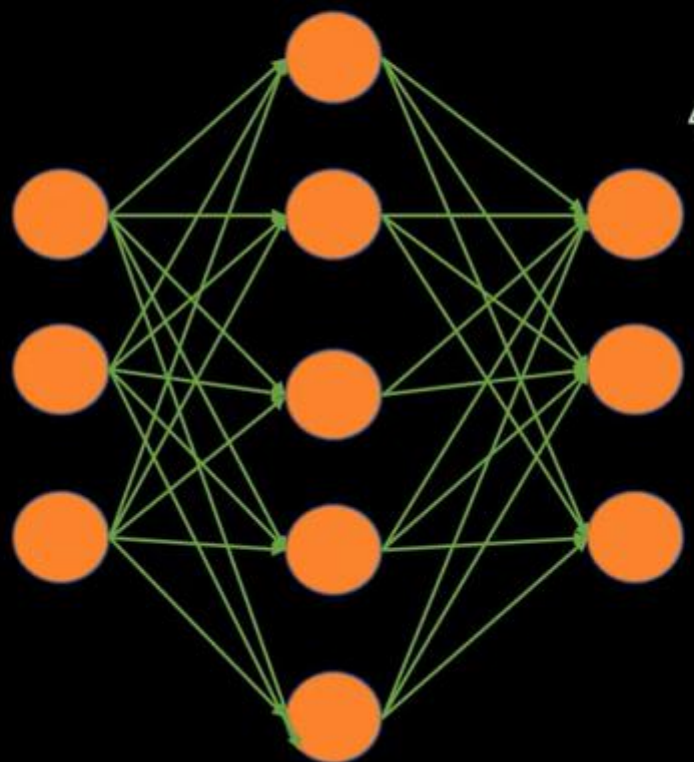
Issue # 2: Too much computation

25000 words in vocabulary

how → [0,0,0,..,1,0,0,...,0]

are → [0, 1,0,0,0,..0,0,...,0]

you? → [0, 0,0,0,..0,0,1,0,0]



42000 words in vocabulary

[0,0,0,..,1,0,0,...,0] क्या

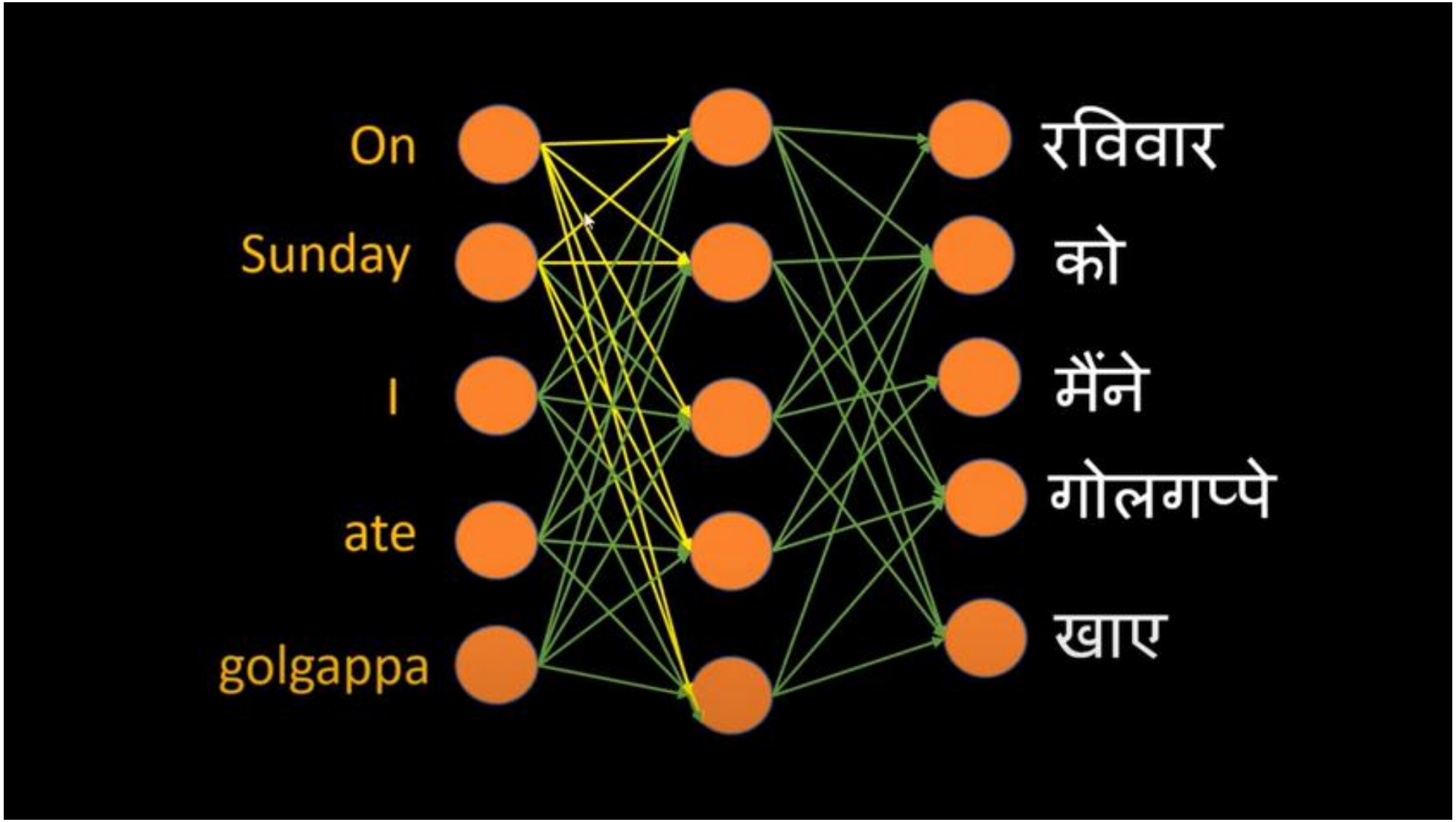
[0,0,0,..,1,0,0,...,0] हाल

[0,0,0,..,1,0,0,...,0] है?

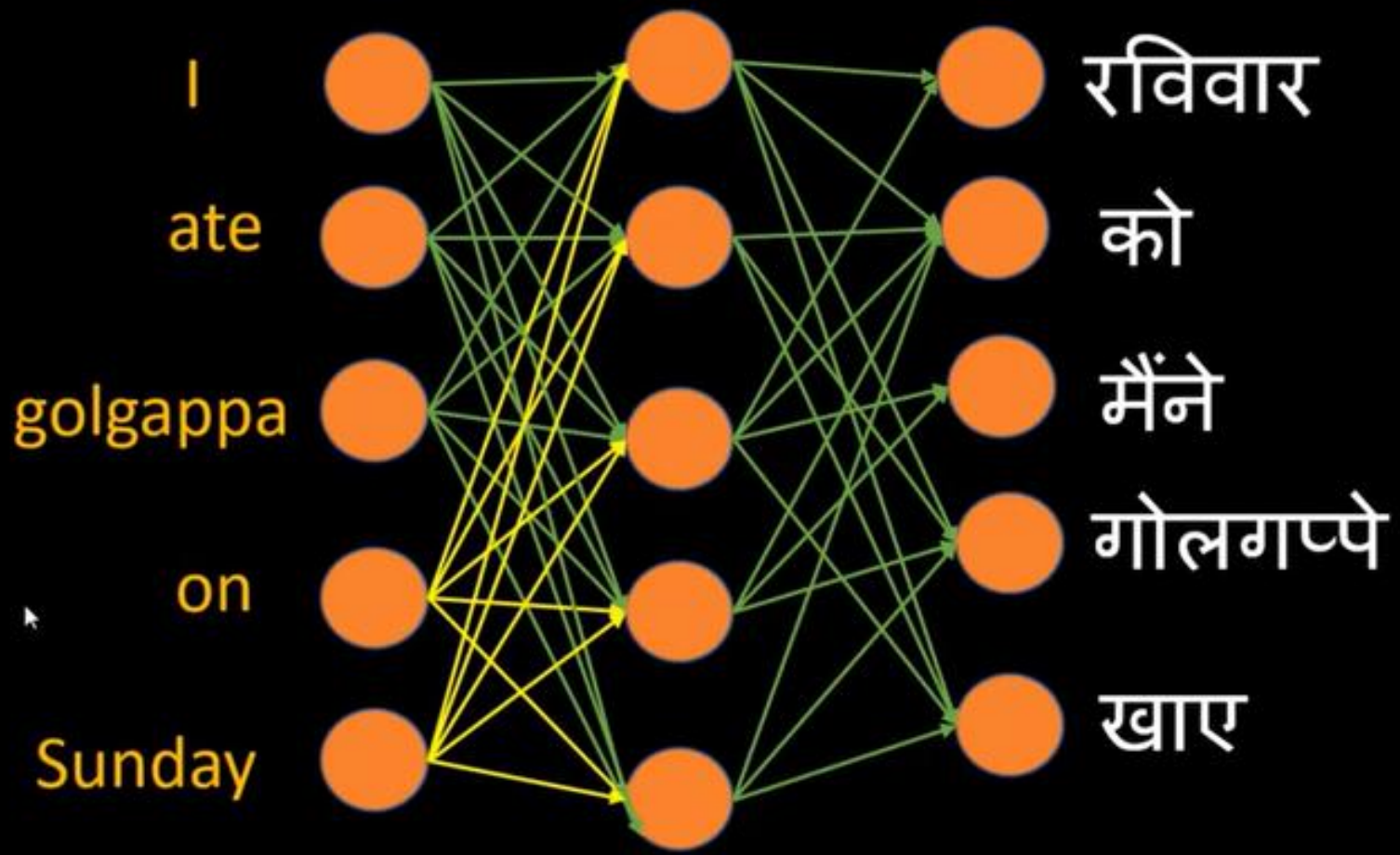
On sunday I ate golgappa

I ate golgappa on Sunday

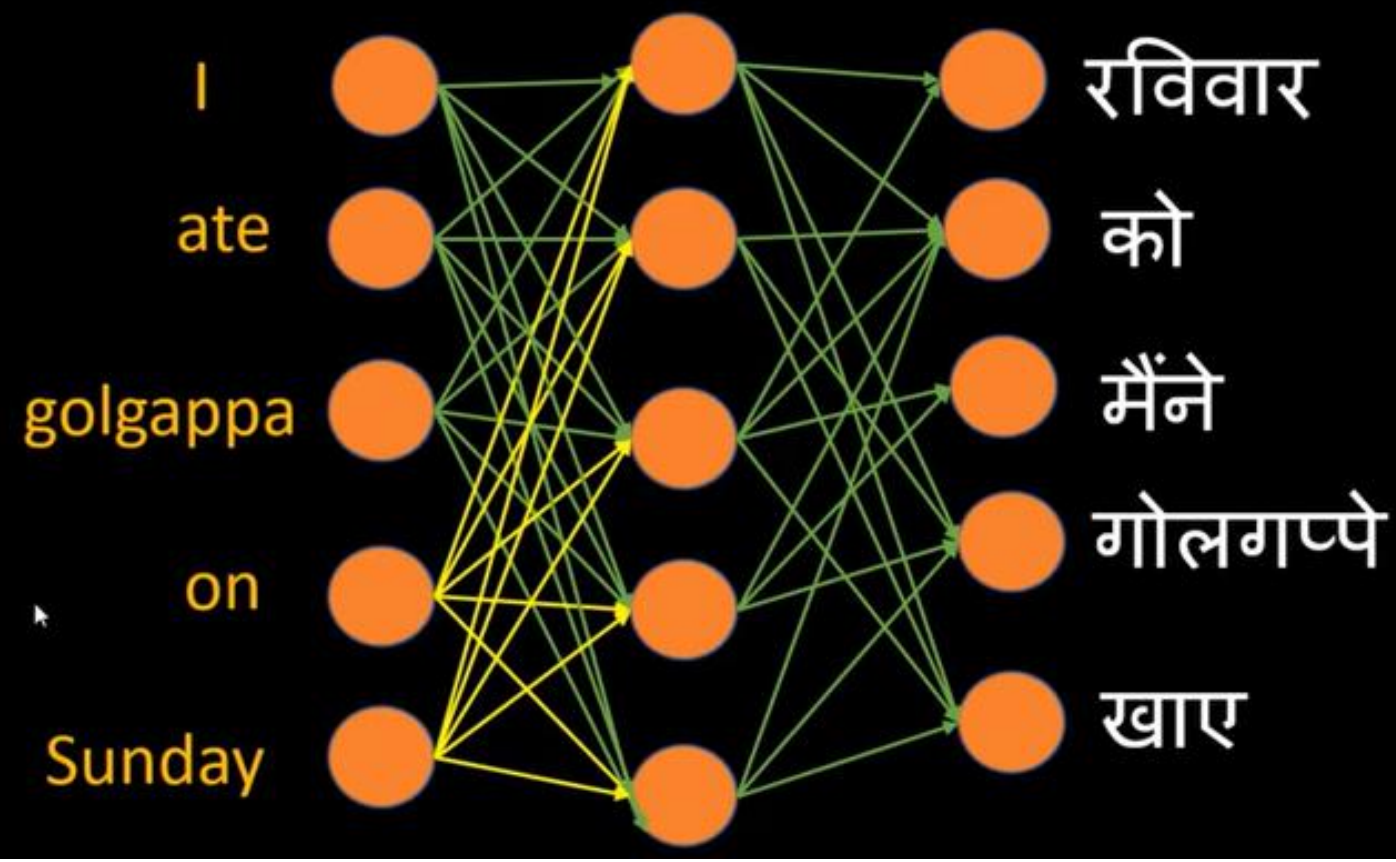
रविवार को मैंने गोलगप्पे खाए

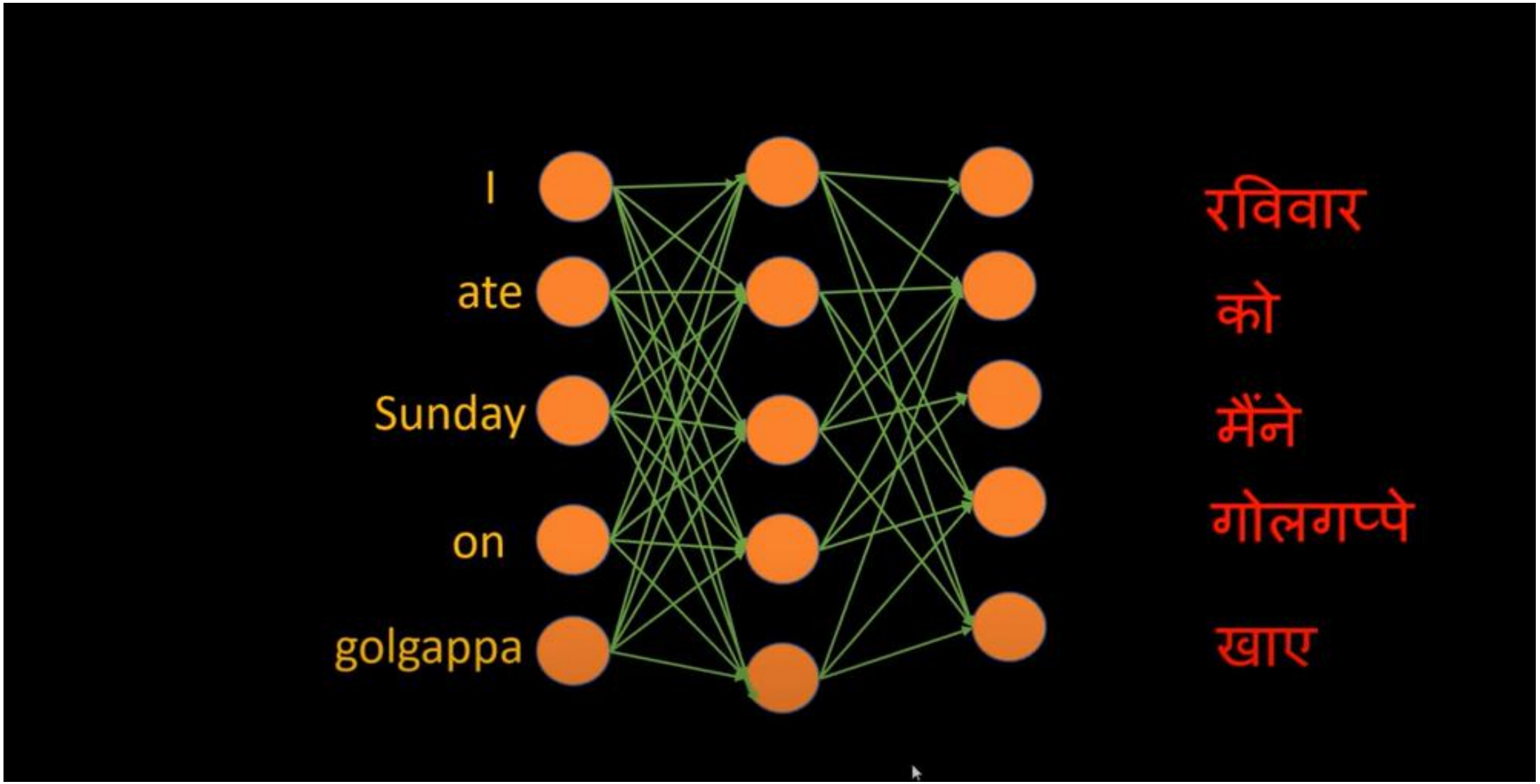

On
Sunday
I
ate
golgappa
रविवार
को
मैंने
गोलगप्पे
खाए

Issue # 3: Parameters are not shared



Issue # 3: Parameters are not shared




I
ate
Sunday
on
golgappa
रविवार
को
मैंने
गोलगप्पे
खाए

3 Issues using ANN for sequence problems

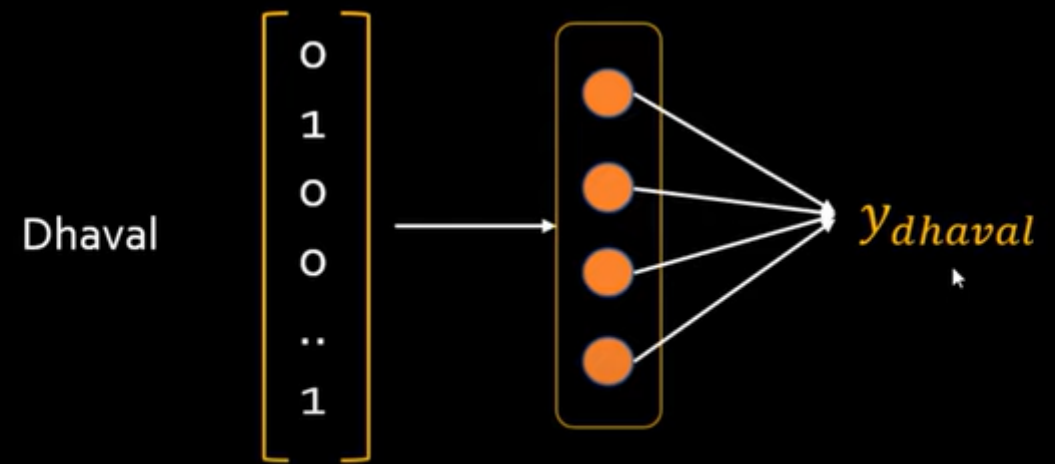
Variable size of input/output neurons

Too much computation

No parameter sharing

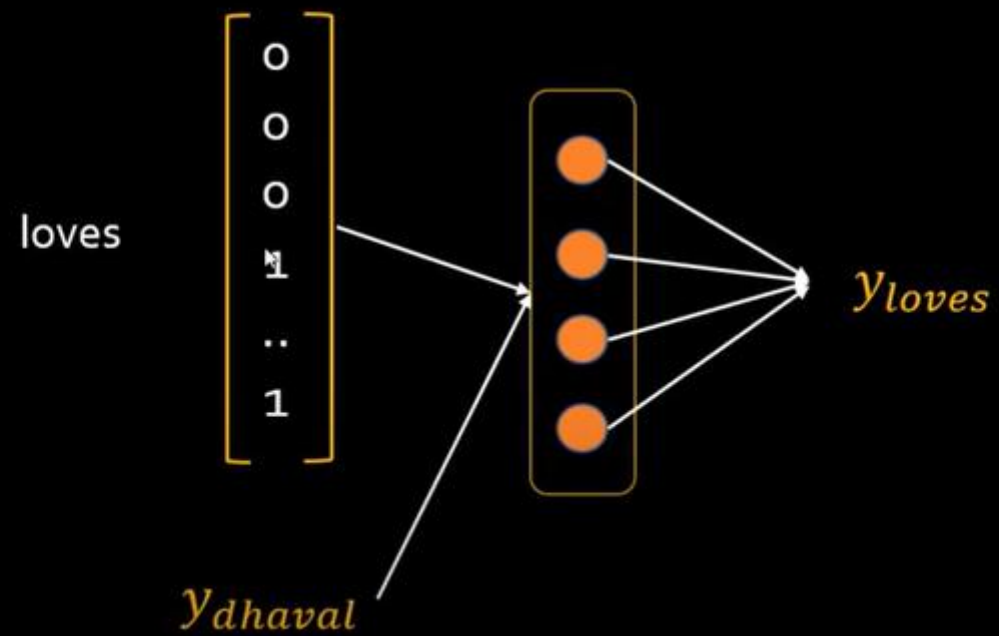
Named Entity Recognition

Dhaval loves baby yoda

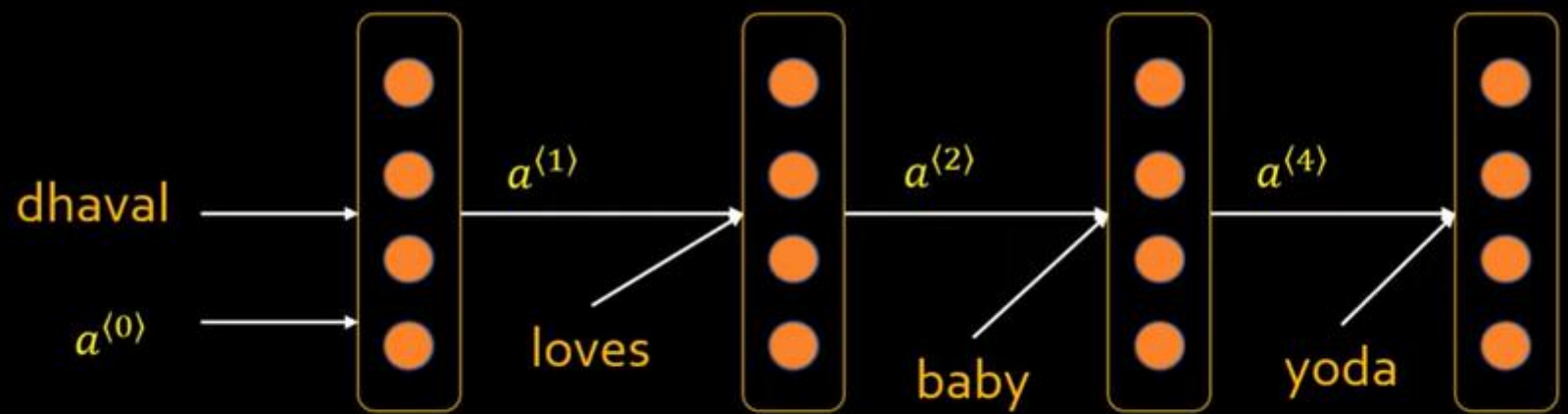


Named Entity Recognition

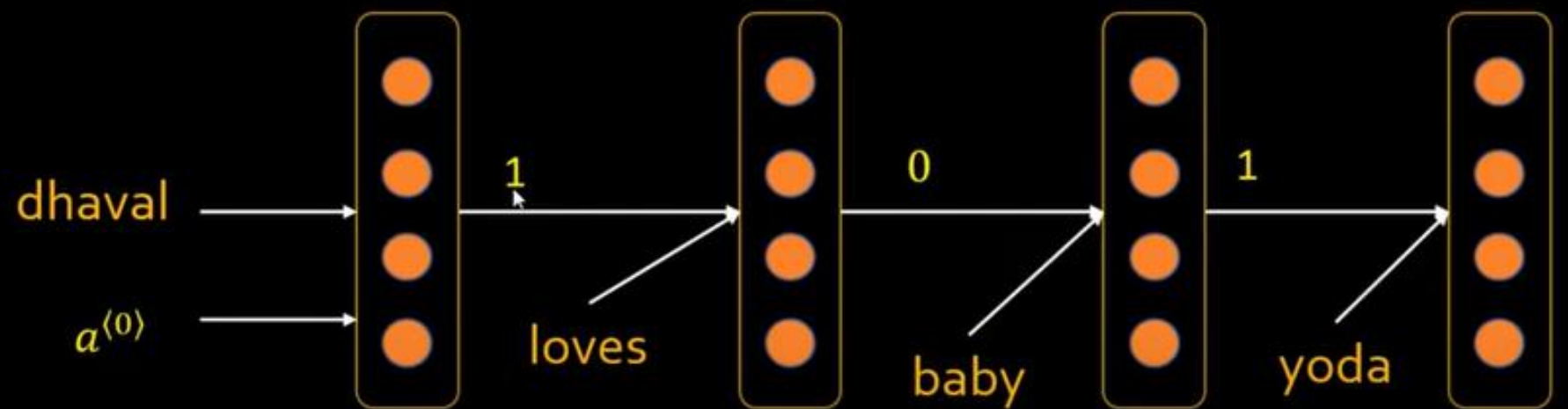
Dhaval loves baby yoda



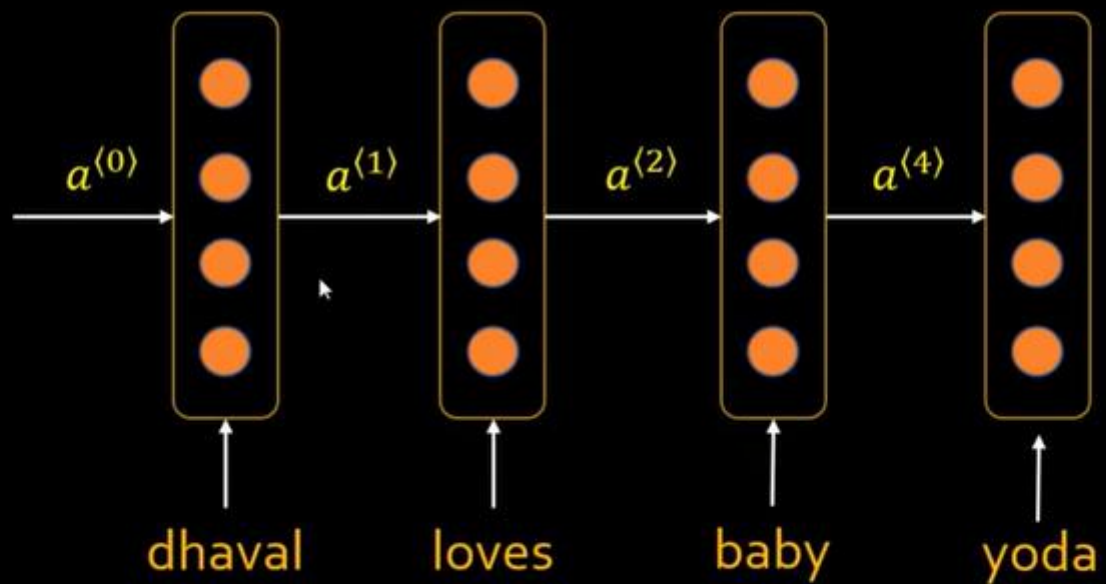
Named Entity Recognition



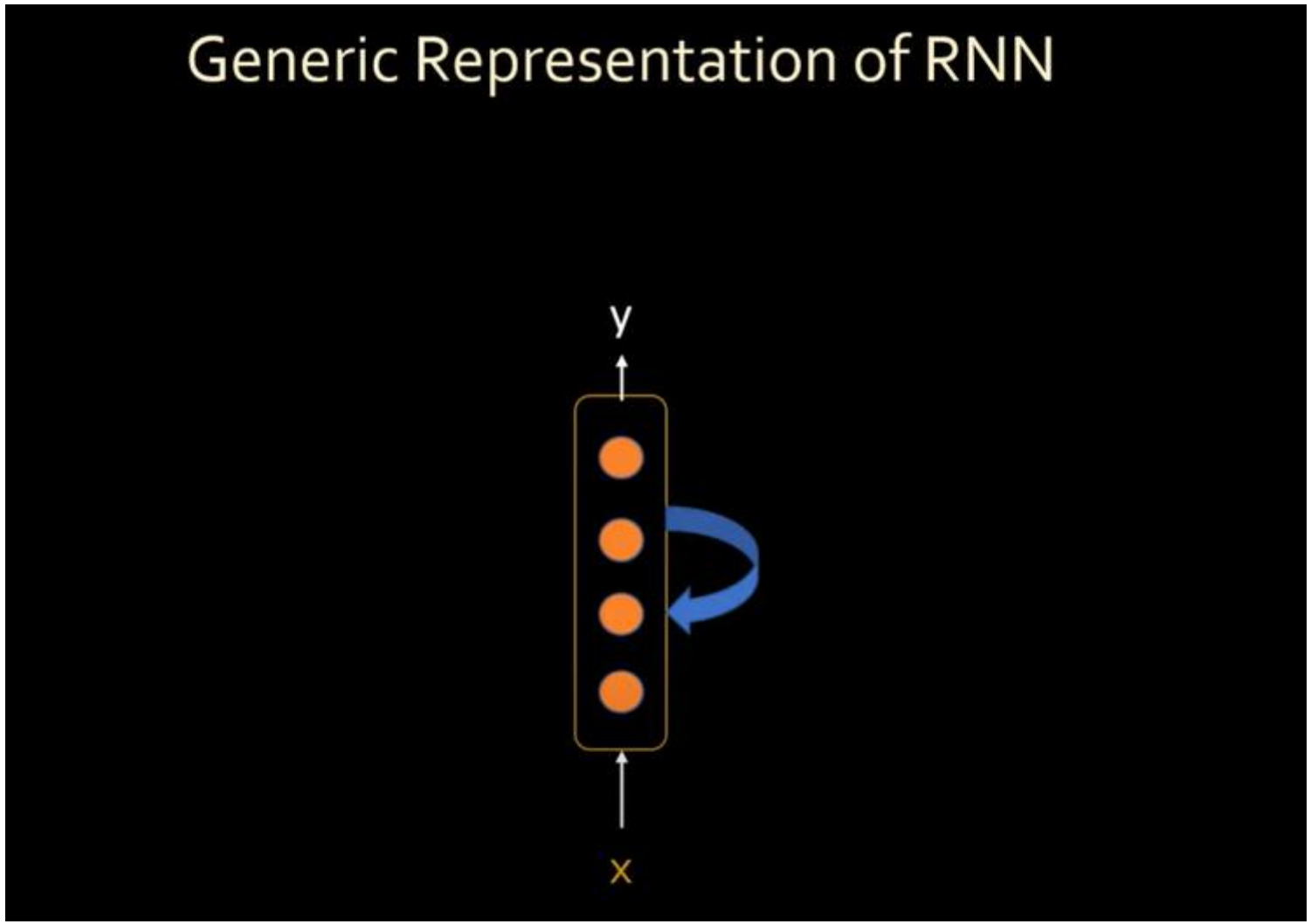
Named Entity Recognition: once network is trained



Named Entity Recognition



Generic Representation of RNN

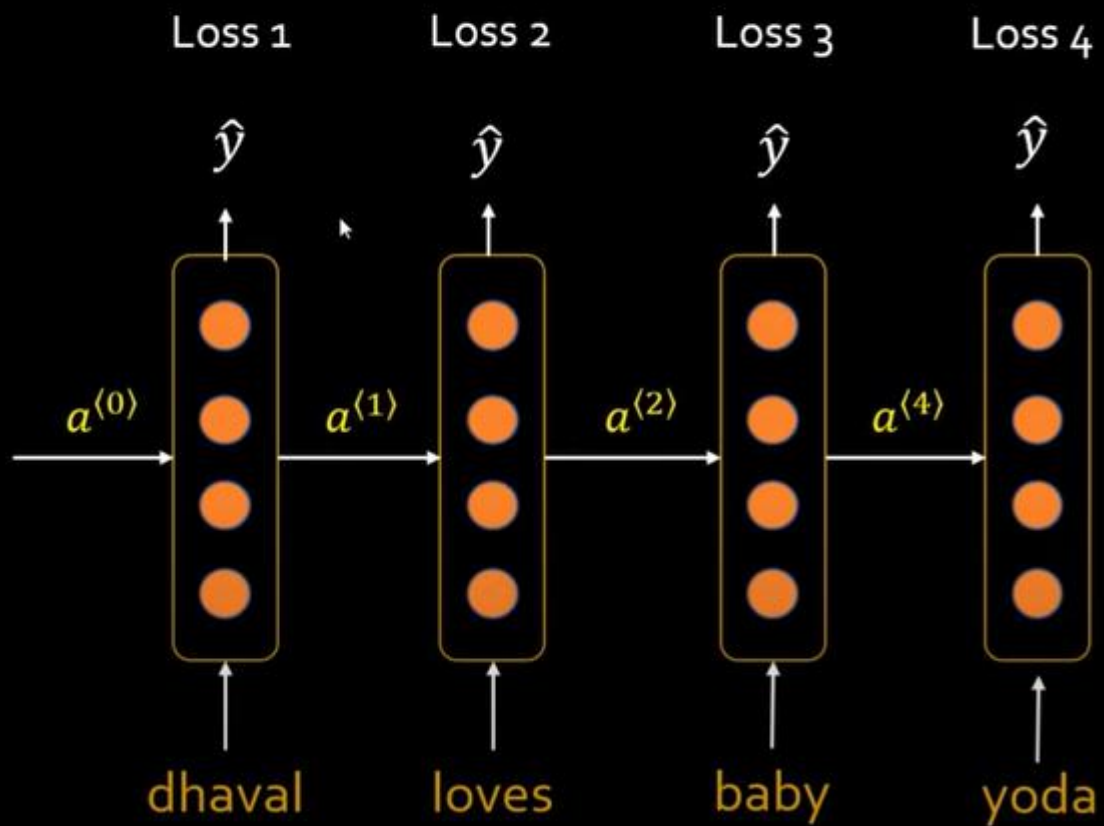


Training : Named Entity Recognition (NER)

X	y
Dhaval loves baby yoda	1 0 1 1
Bob told Ahmed that pizza is delivered	1 0 1 0 0 0 0
Ironman punched on hulk's face	1 0 0 1 1

Training

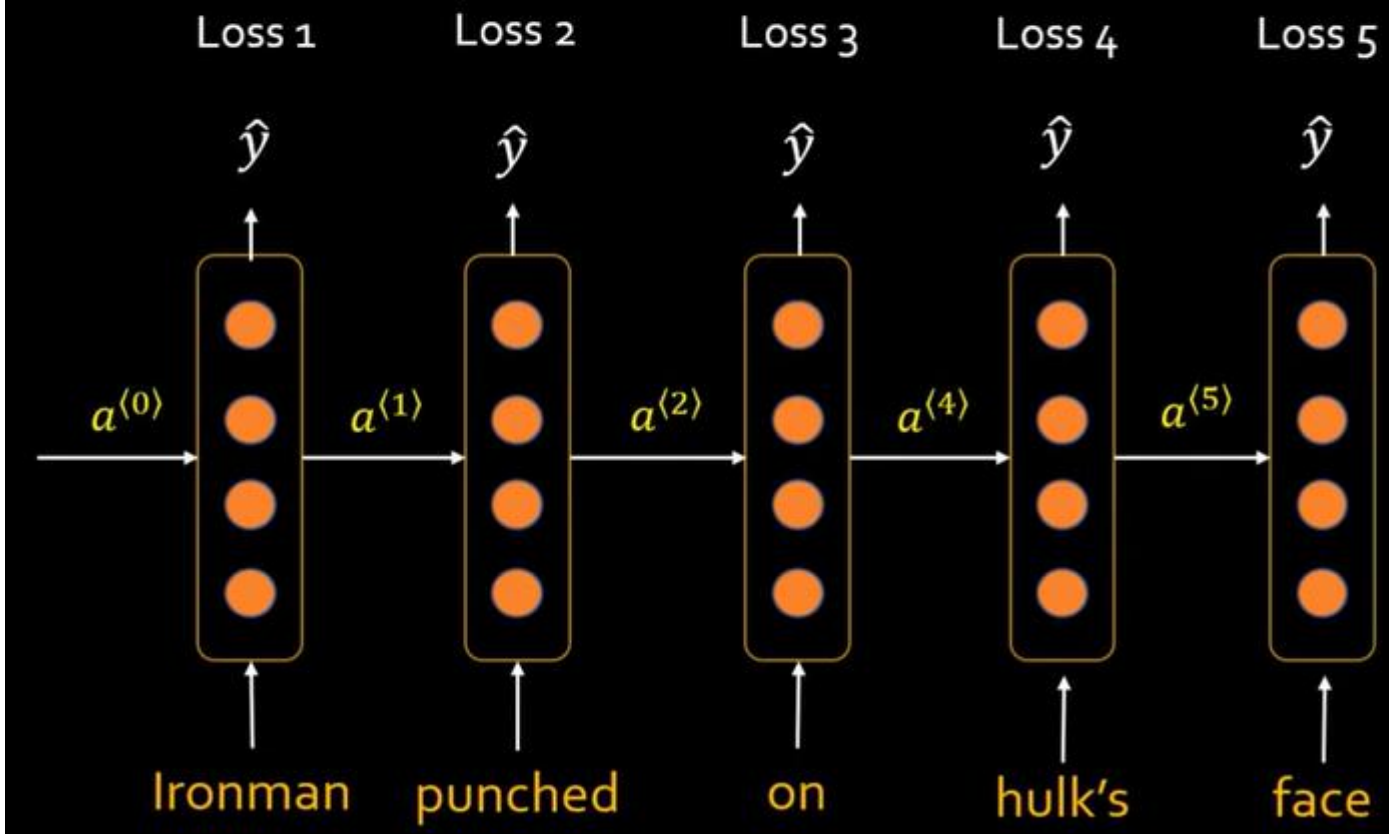
Dhaval loves baby yoda → 1 0 1 1



Total Loss = Loss 1 + Loss 2 + Loss 3 + Loss4

Training

Ironman punched on hulk's face → 1 0 0 1 0



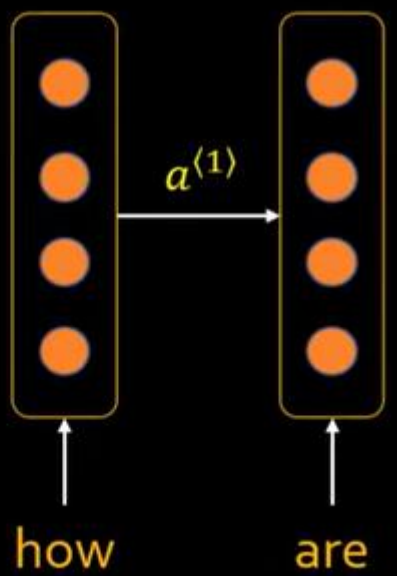
Total Loss = Loss 1 + Loss 2 + Loss 3 + Loss4 + Loss 5

Language translation

how are you? क्या हाल है?

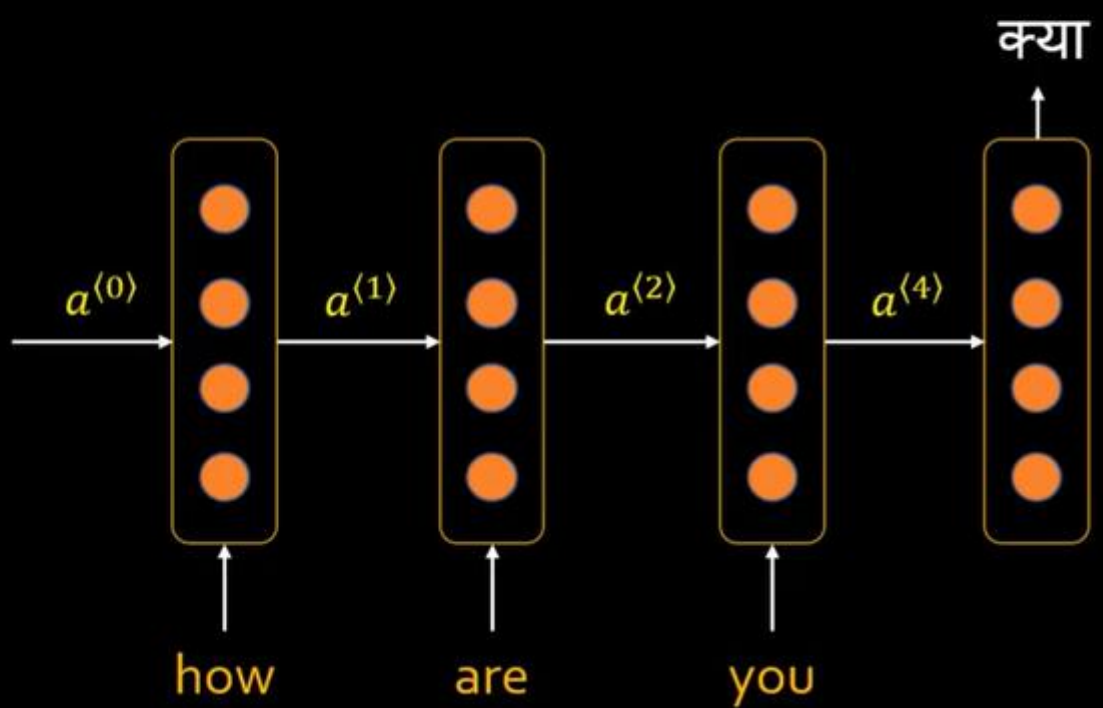
Language translation

how are you? क्या हाल है?



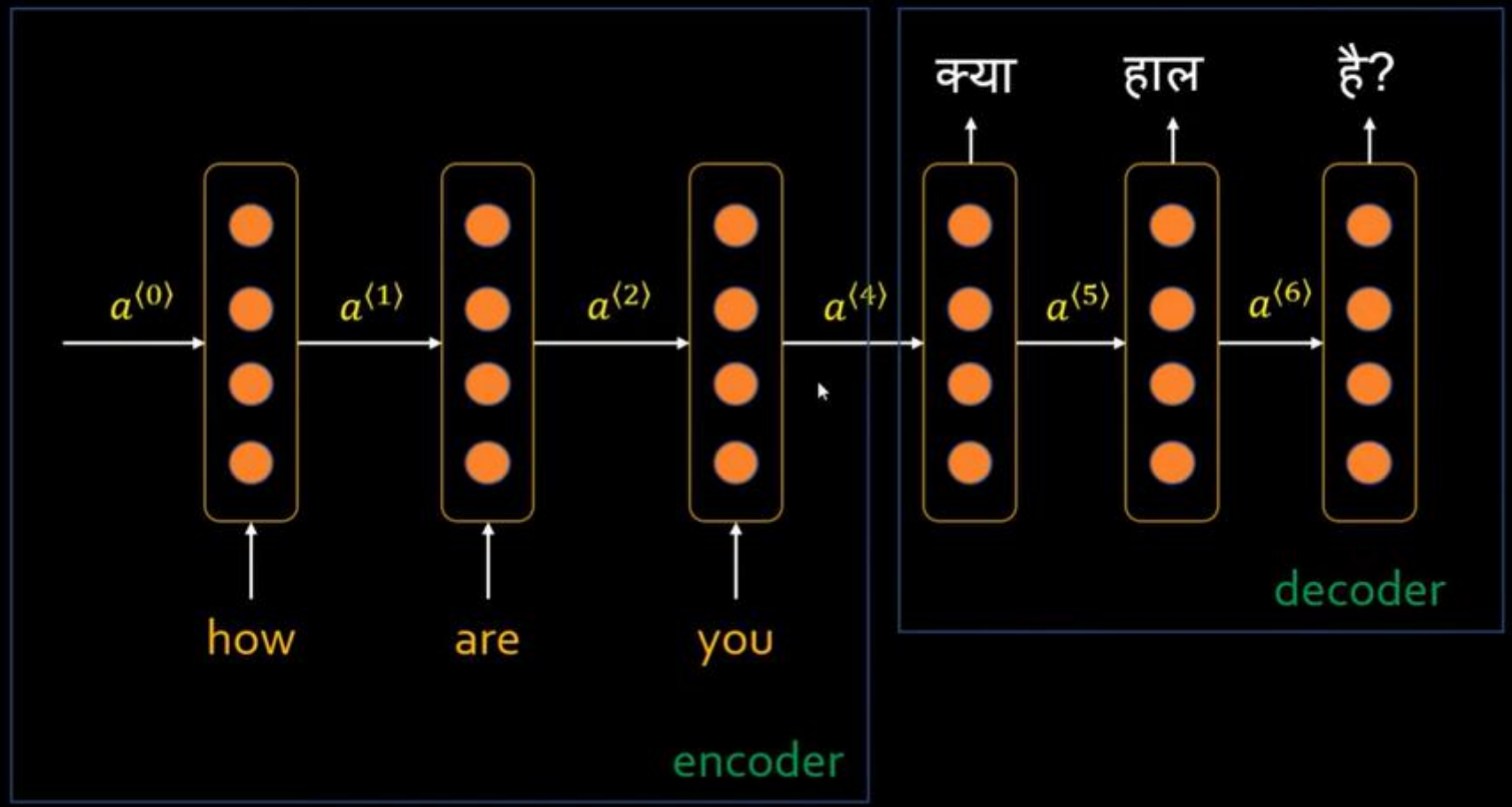
Language translation

how are you? क्या हाल है?

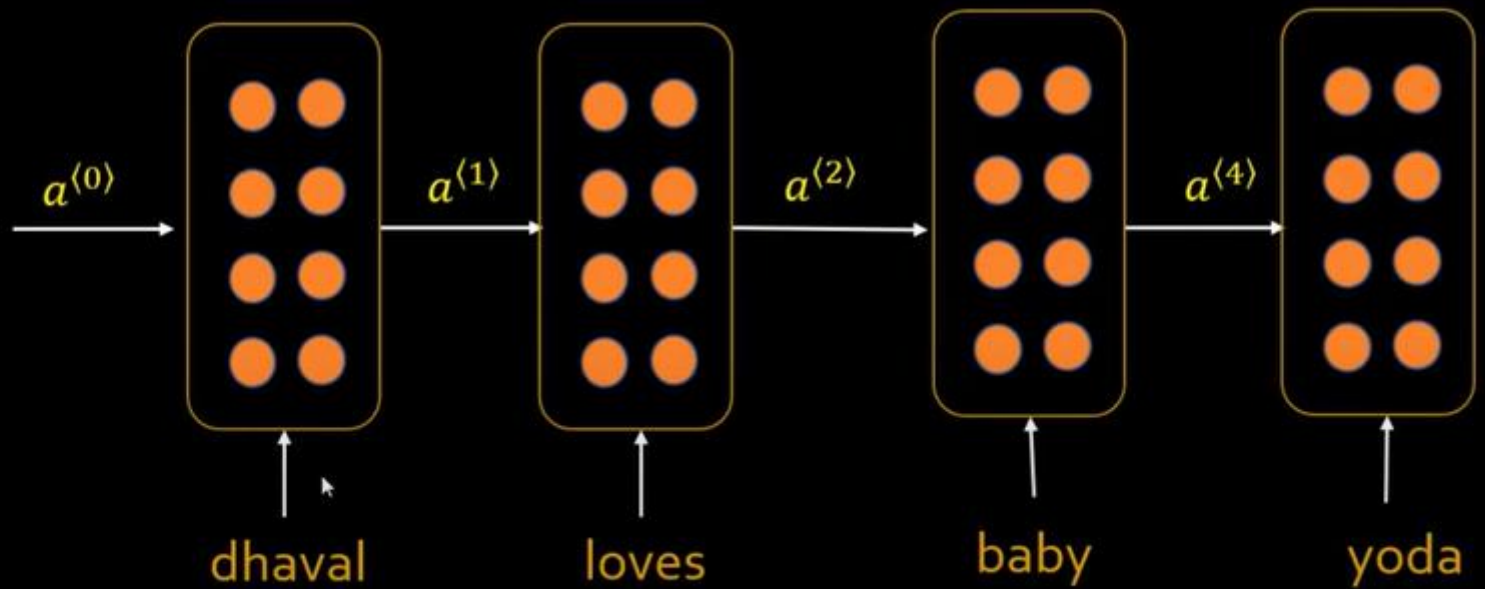


Language translation

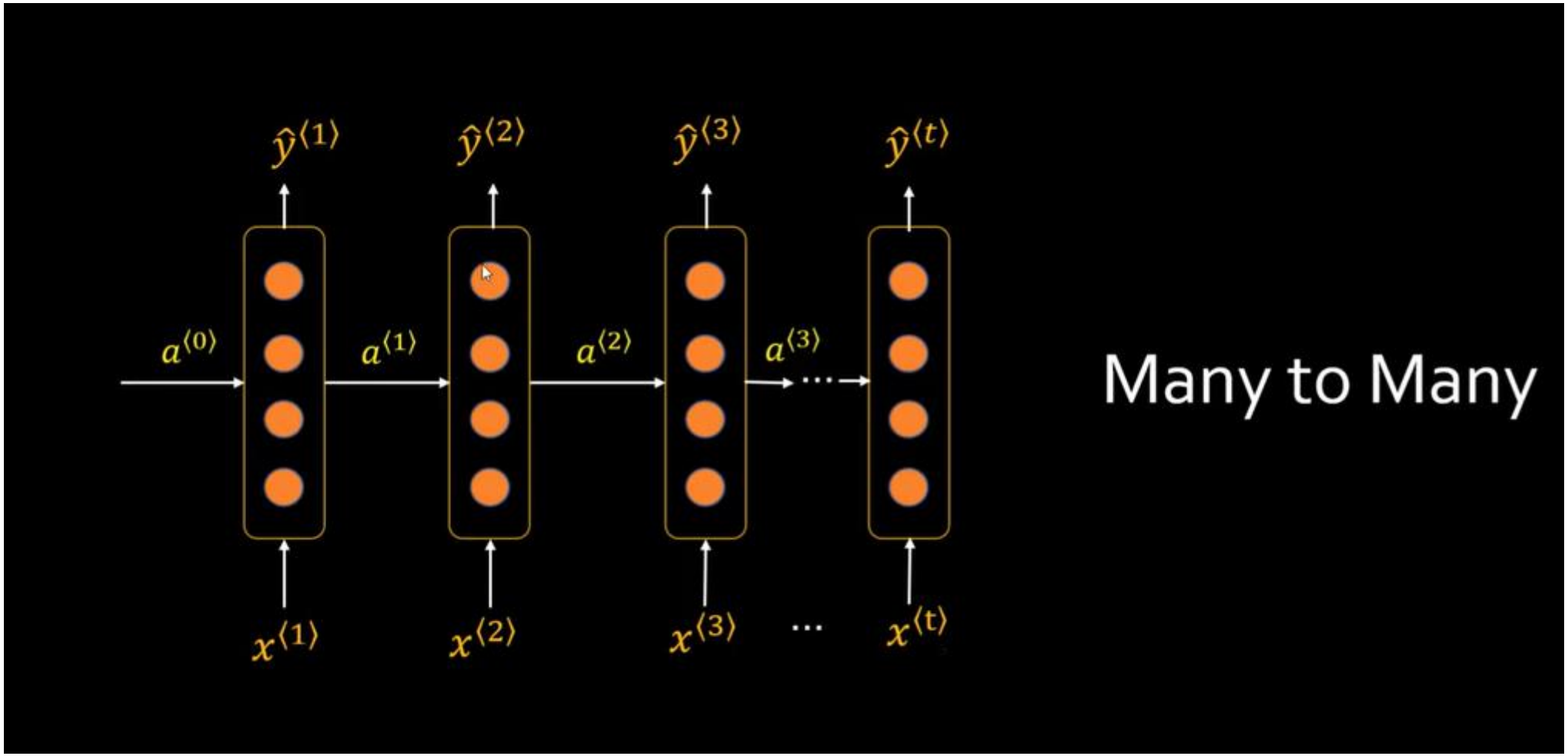
how are you? क्या हाल है?

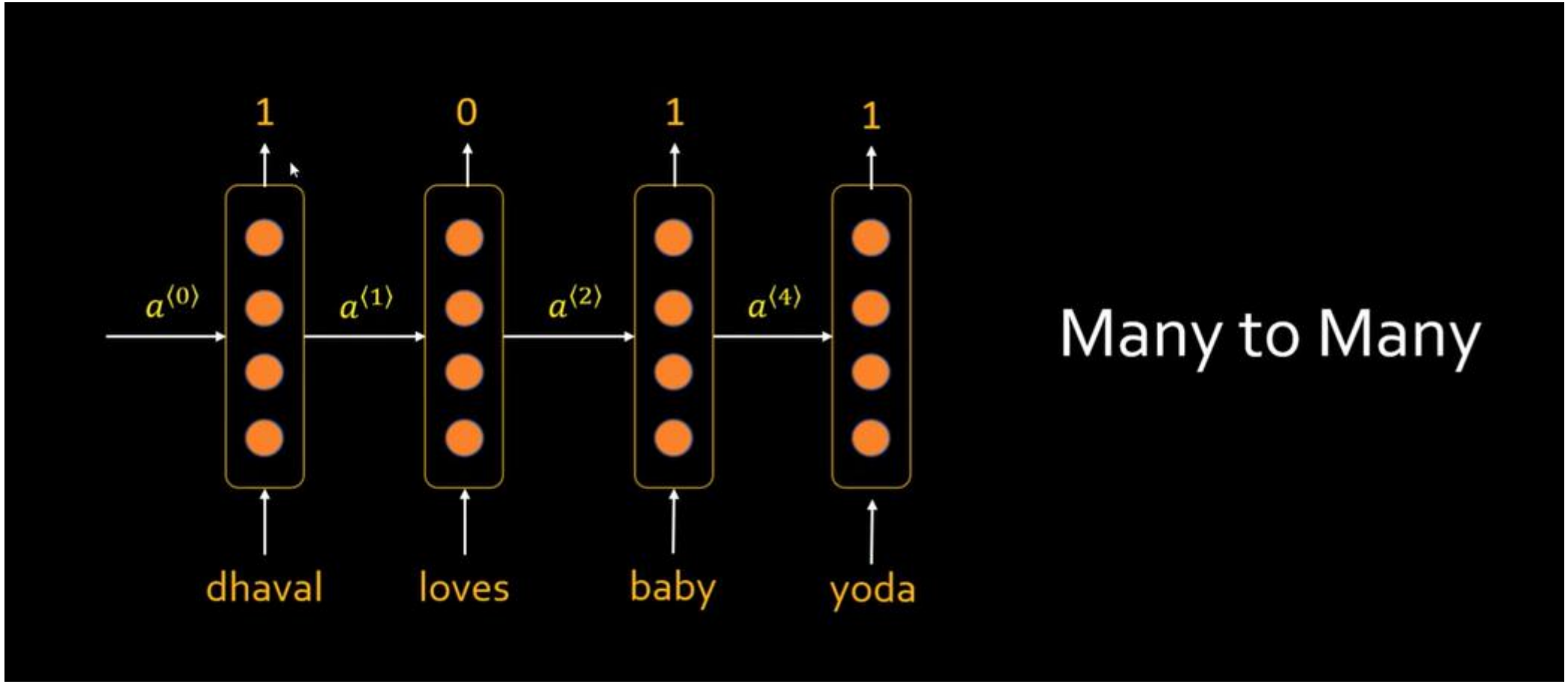


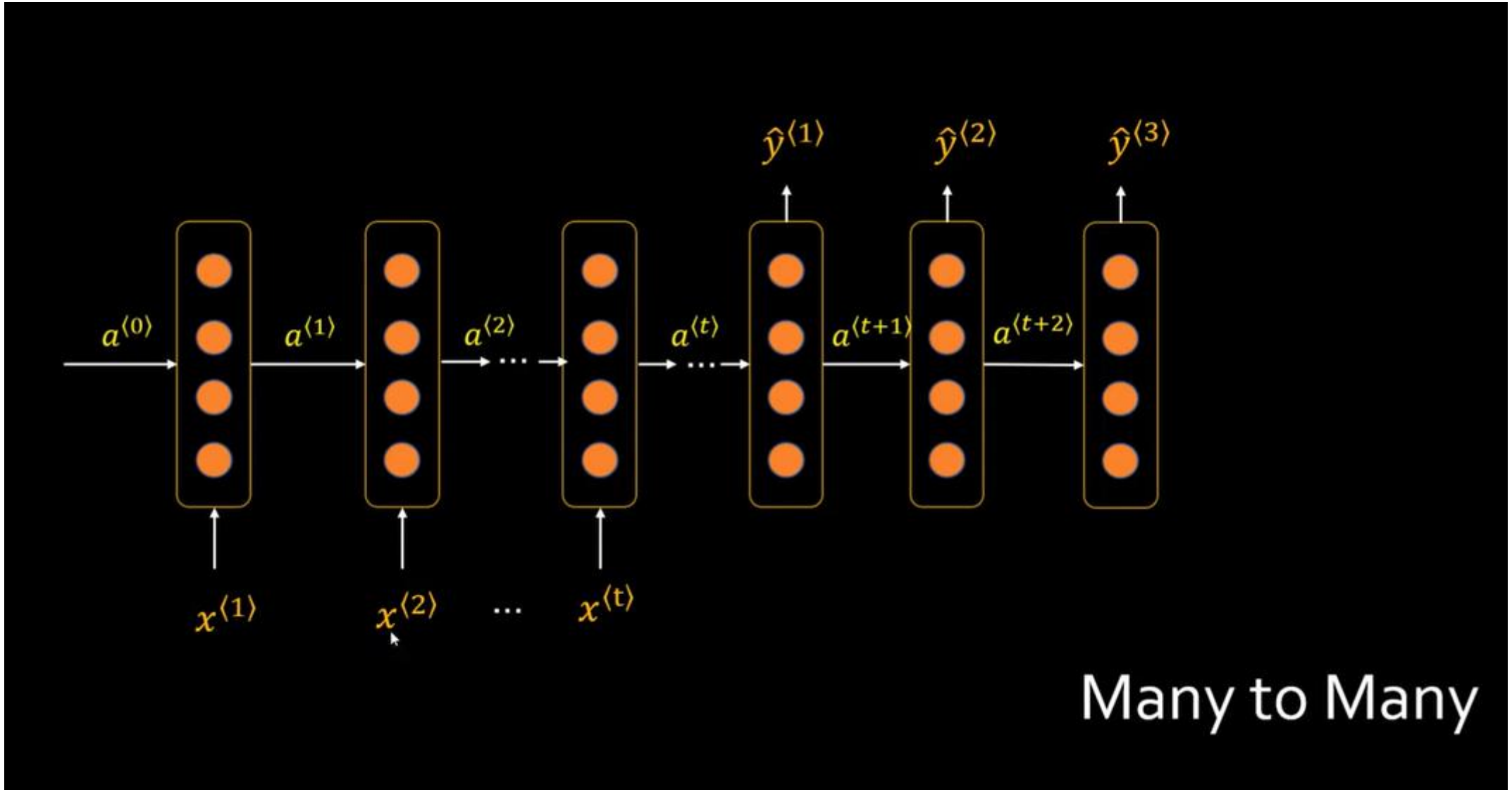
Deep RNN


$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 4 \rangle}$
dhaval
loves
baby
yoda

Types of RNN


$\hat{y}^{\langle 1 \rangle}$
$\hat{y}^{\langle 2 \rangle}$
$\hat{y}^{\langle 3 \rangle}$
$\hat{y}^{\langle t \rangle}$
$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 3 \rangle}$
$x^{\langle 1 \rangle}$
$x^{\langle 2 \rangle}$
$x^{\langle 3 \rangle}$
$x^{\langle t \rangle}$
Many to Many

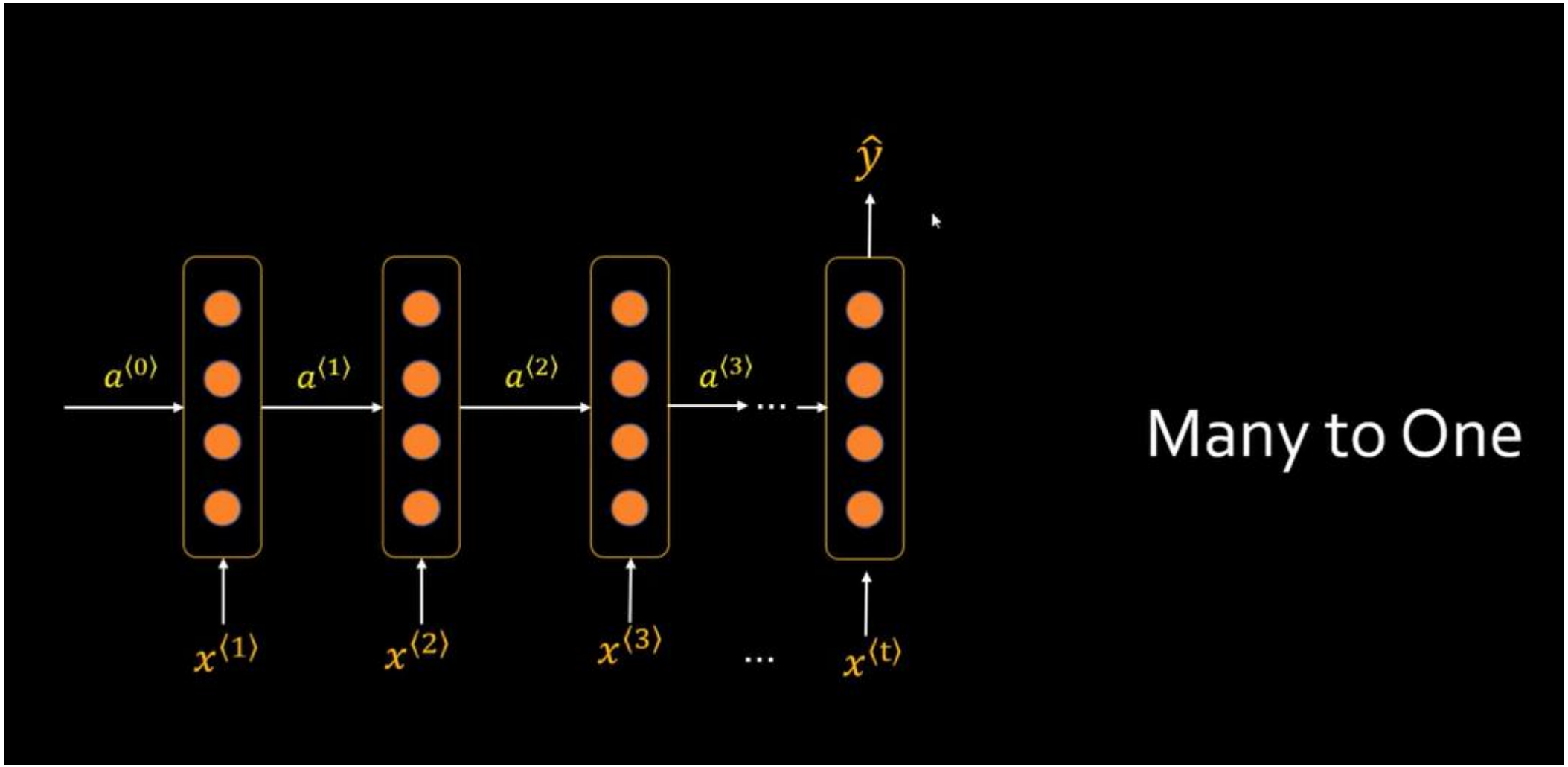

1
0
1
1
$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 4 \rangle}$
dhaval
loves
baby
yoda
Many to Many


$\hat{y}^{\langle 1 \rangle}$
$\hat{y}^{\langle 2 \rangle}$
$\hat{y}^{\langle 3 \rangle}$
$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle t \rangle}$
$a^{\langle t+1 \rangle}$
$a^{\langle t+2 \rangle}$
$x^{\langle 1 \rangle}$
$x^{\langle 2 \rangle}$
$x^{\langle t \rangle}$
Many to Many

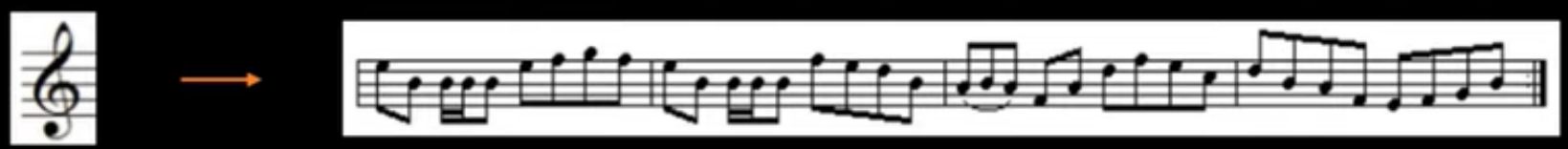
Sentiment Analysis

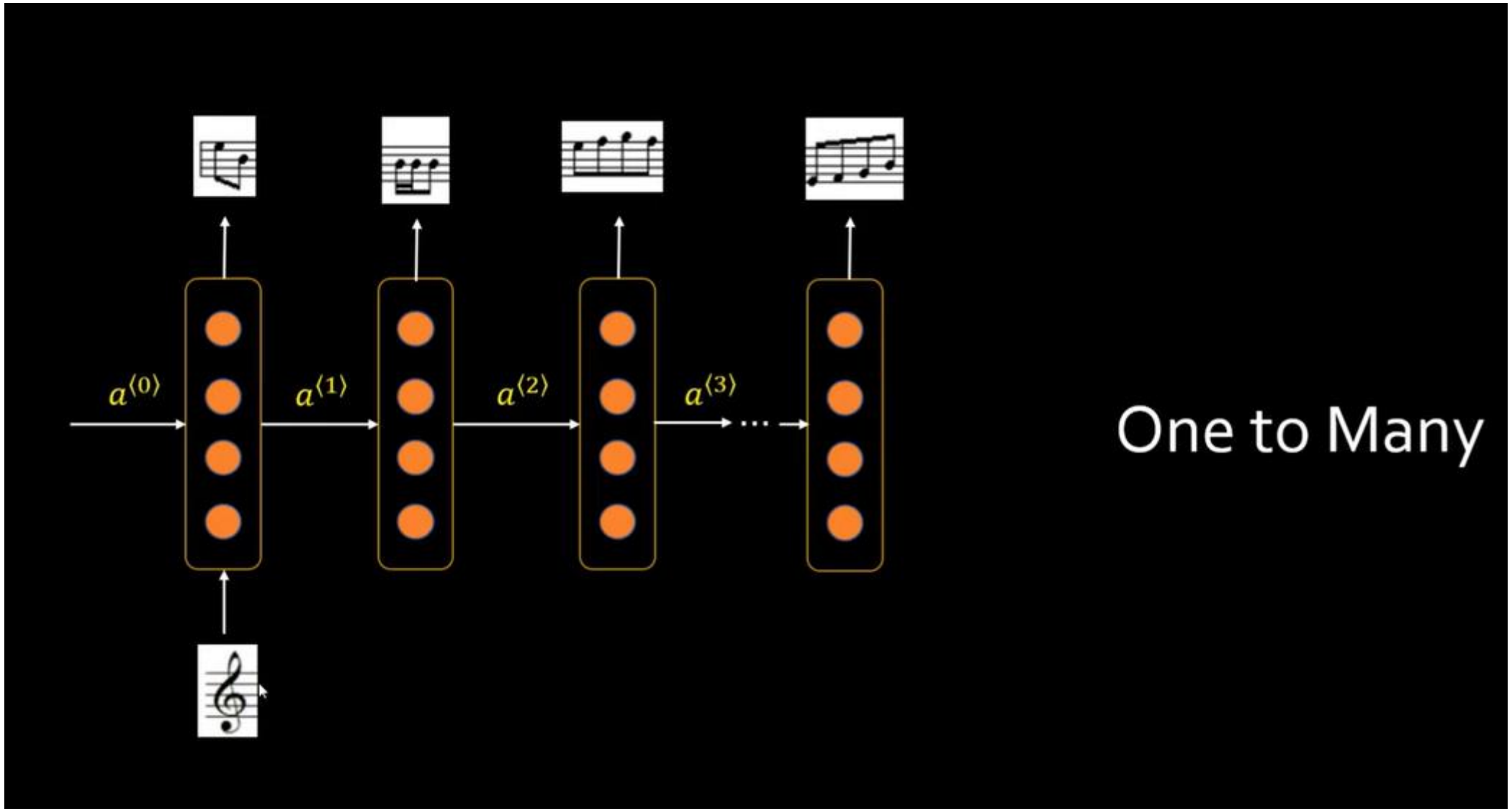
Not only the fan was expensive, but it was broken when it arrived. → ★☆☆☆☆

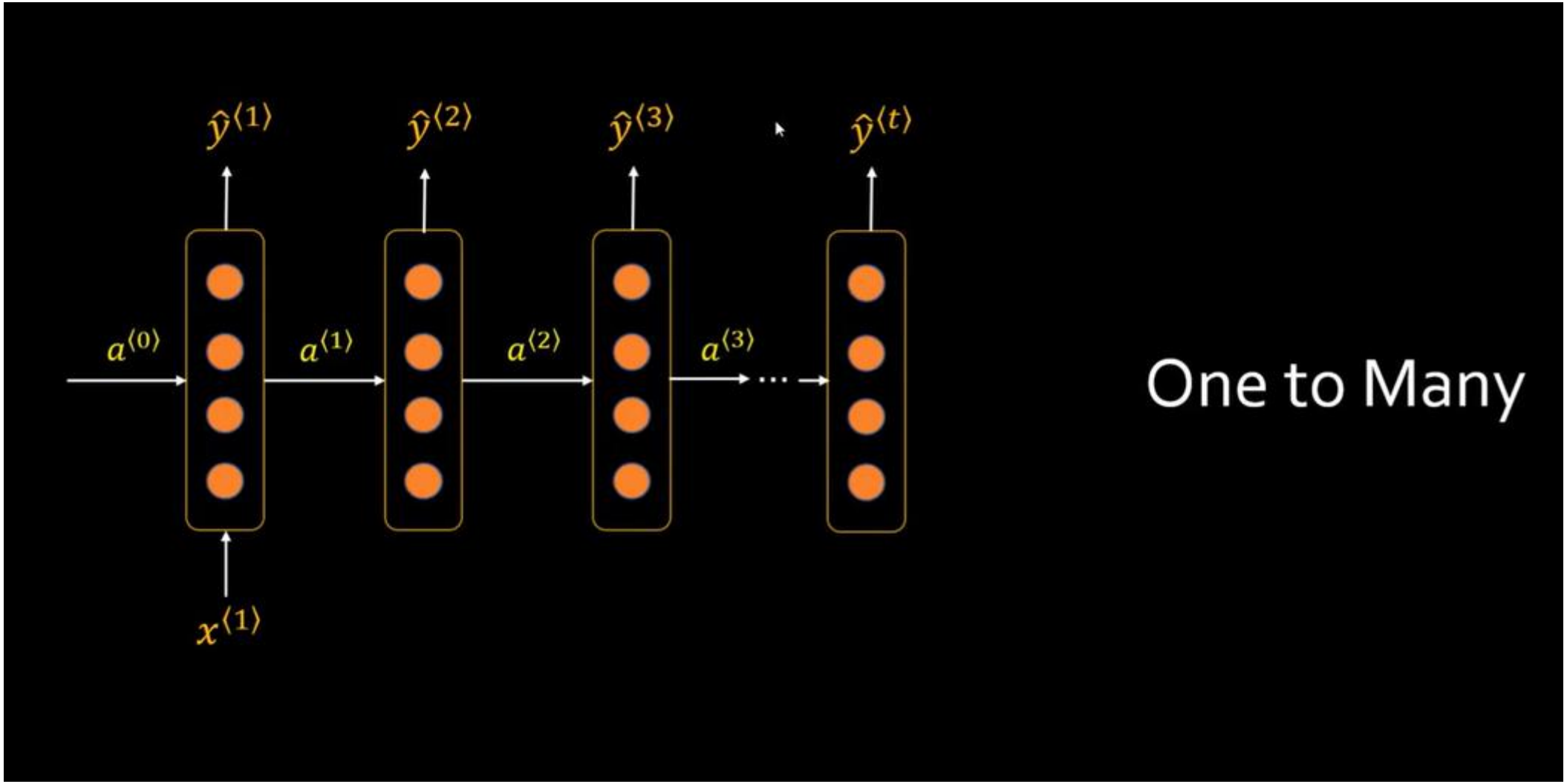
The fan works like a charm, I wasn't expecting such a good quality at this cheap price → ★★★★★


$\hat{y}$
$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 3 \rangle}$
...
$x^{\langle 1 \rangle}$
$x^{\langle 2 \rangle}$
$x^{\langle 3 \rangle}$
...
$x^{\langle t \rangle}$
Many to One

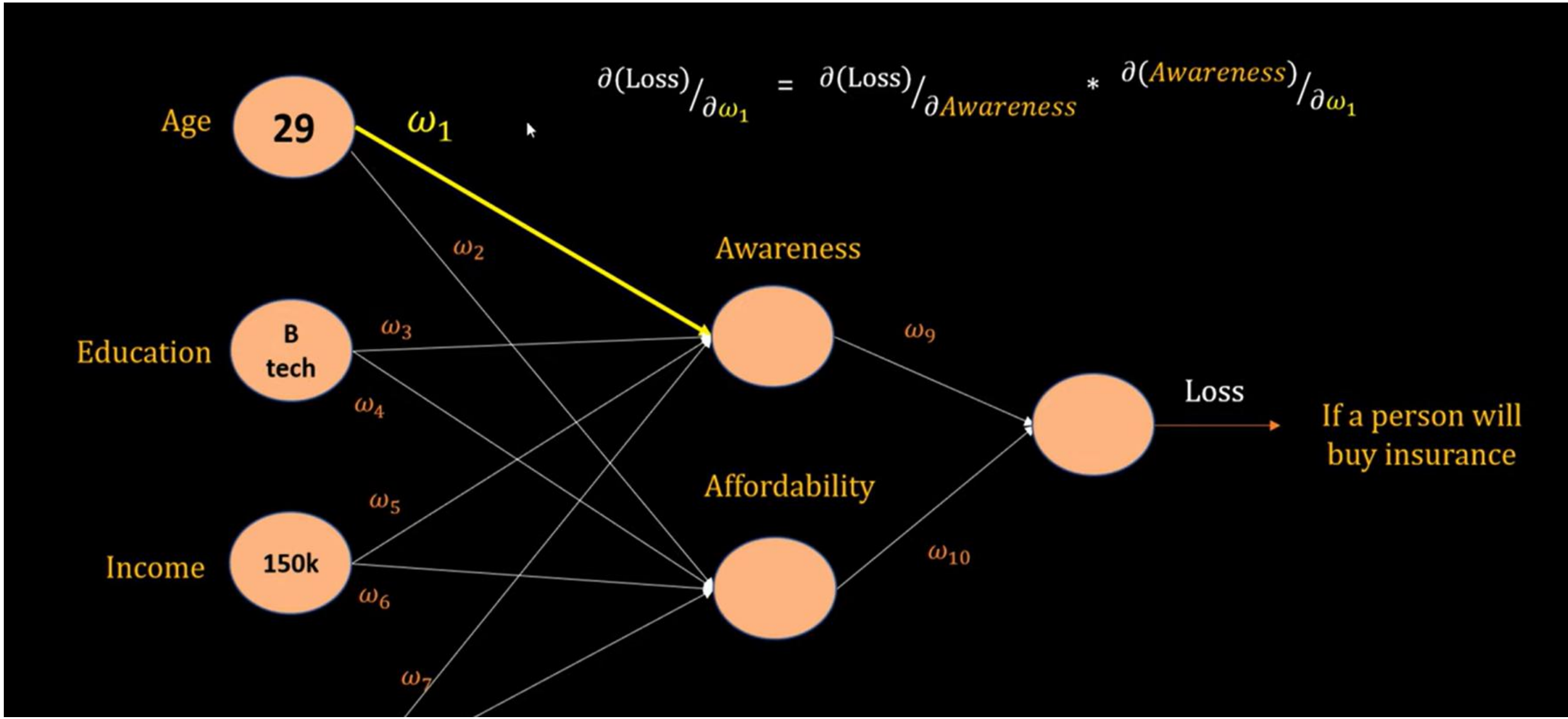
Music Generation




$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 3 \rangle}$
One to Many


$\hat{y}^{\langle 1 \rangle}$
$\hat{y}^{\langle 2 \rangle}$
$\hat{y}^{\langle 3 \rangle}$
$\hat{y}^{\langle t \rangle}$
$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 3 \rangle}$
$x^{\langle 1 \rangle}$
One to Many

Vanishing Gradient

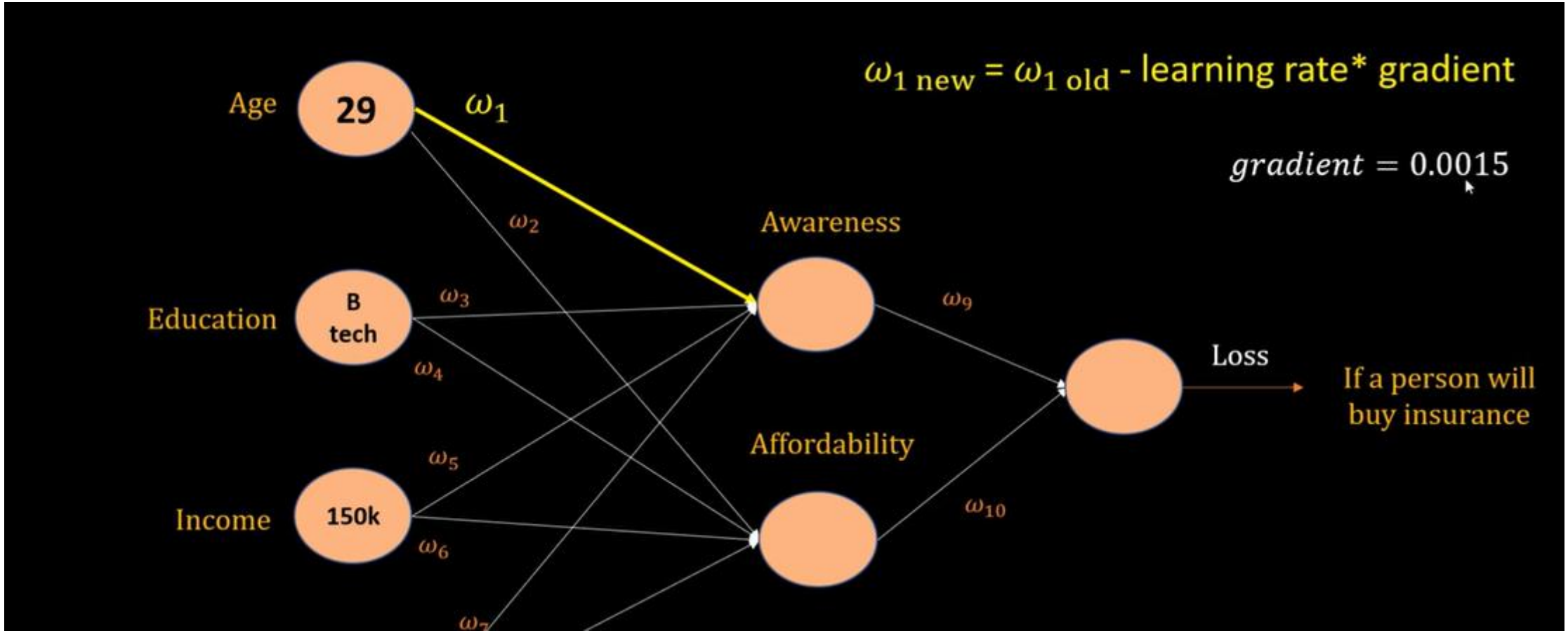

$\partial(\text{Loss})/\partial\omega_1 = \partial(\text{Loss})/\partial Awareness * \partial(Awareness)/\partial\omega_1$
Age
29
ω_1
ω_2
Awareness
Education
B tech
ω_3
ω_4
ω_9
Loss
If a person will buy insurance
ω_5
Affordability
Income
150k
ω_6
ω_{10}
ω_7

$$\partial(\text{Loss})/\partial\omega_1 = \partial(\text{Loss})/\partial Awareness * \partial(Awareness)/\partial\omega_1$$

$$gradient = d1 * d2$$

$$gradient = 0.03 * 0.05$$

$$gradient = 0.0015$$


$\omega_{1\ new} = \omega_{1\ old}$ - learning rate* gradient
$gradient = 0.0015$
Age
29
ω_1
ω_2
Awareness
Education
B tech
ω_3
ω_4
ω_9
Loss
If a person will buy insurance
Affordability
ω_5
Income
150k
ω_6
ω_{10}
ω_7

As number of hidden layers grow, gradient becomes very small and weights will hardly change . This will hamper the learning process.

Vanishing Gradients

$$\partial(\text{Loss})/_{\partial \omega_1} = \partial(\text{Loss})/_{\partial Awareness} * \partial(Awareness)/_{\partial \omega_1}$$

$$gradient = d1 * d2$$

$$gradient = 100 * 500$$

$$gradient = 50000$$

$$gradient = d1 * d2 * d3 * d4 * \ldots * dn$$

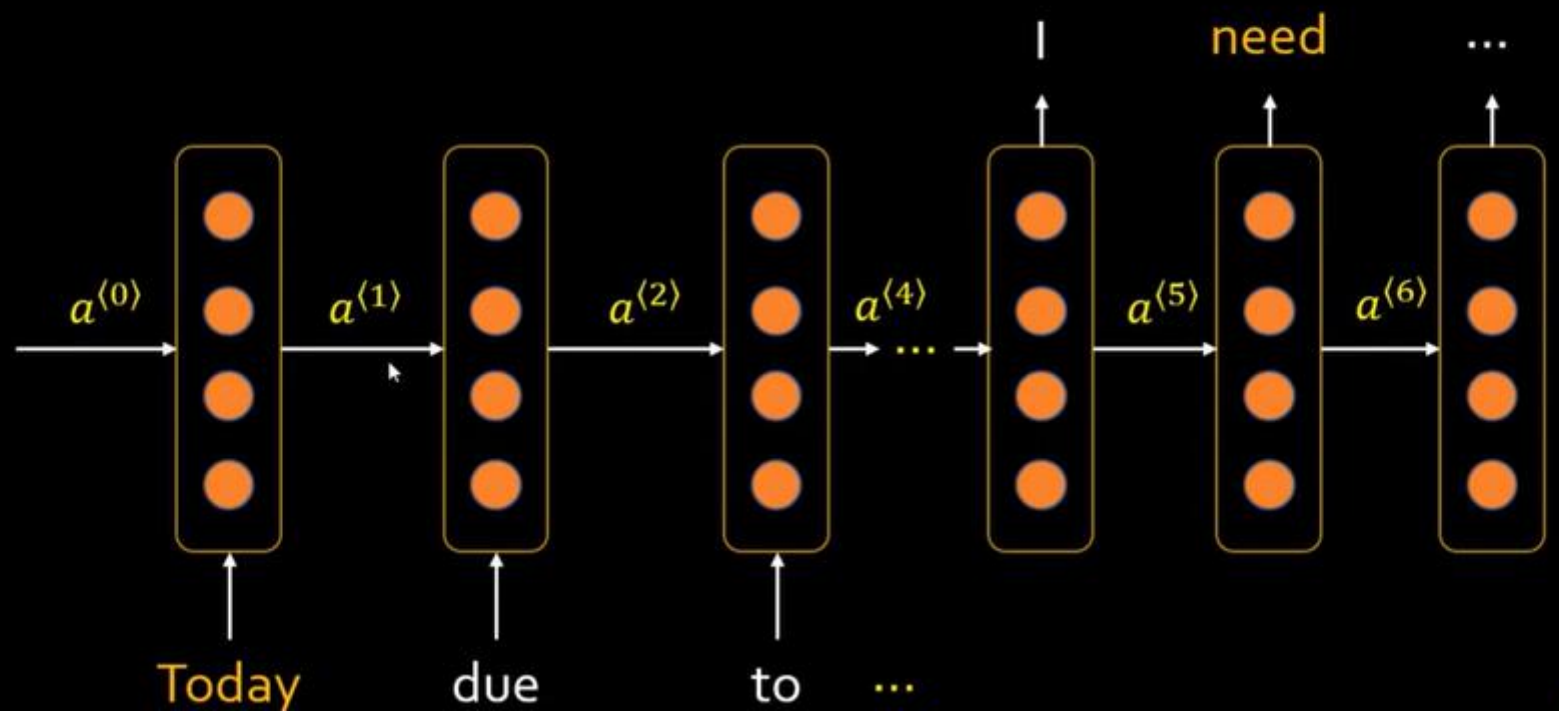
Vanishing gradient problem is more prominent in very deep neural networks.

Vanishing gradient problem in **RNN**

Today, due to my current job situation and family conditions, I need to take a loan.

Last year, due to my current job situation and family conditions, I had to take a loan.

Today, due to my current job situation and family conditions, I need to take a loan.



Solutions?

GRU

LSTM