Recurrent Neural Networks (RNNs)

Why do we need *Sequential Modeling*?

Examples of Sequence data	Input Data	Output
Speech Recognition		*This is RNN*
Machine Translation	Hello, I am Pankaj.	Hallo, ich bin Pankaj. हैलो, म�पंकज 5ं।
Language Modeling	Recurrent neural ? based ? model	**network** **language**
Named Entity Recognition	Pankaj lives in Munich	**Pankaj** lives in **Munich** *person* *location*
Sentiment Classification	There is nothing to like in this movie.	
Video Activity Analysis		Punching

x

y

not interested at

x

y

not interested at → this time

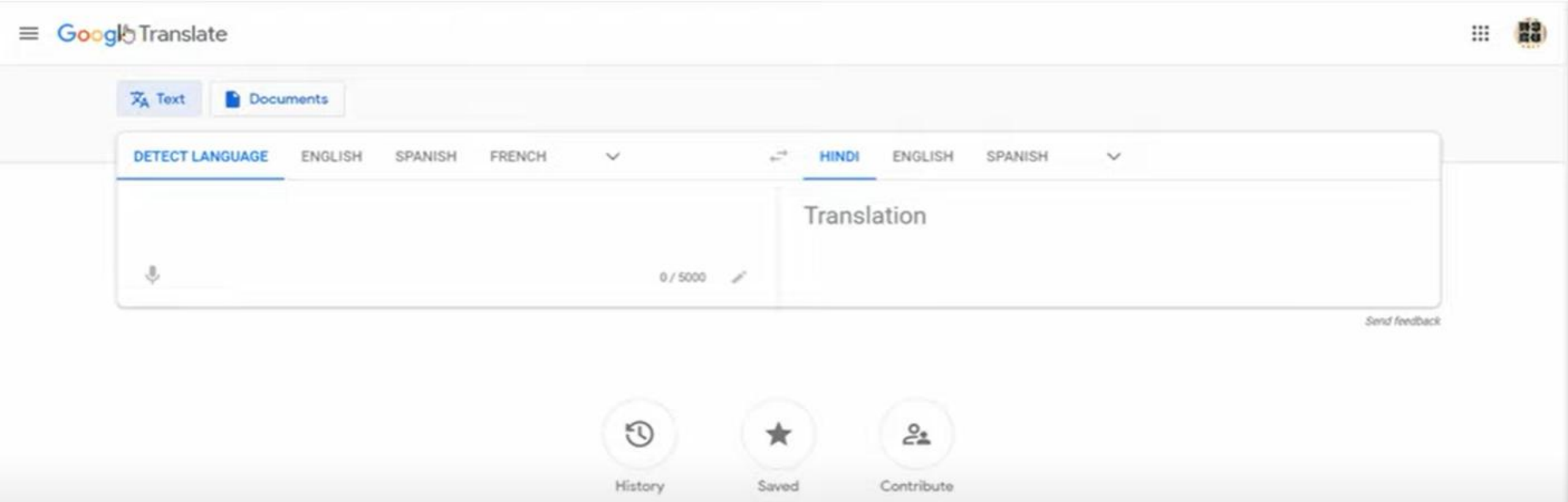
x		y
not interested at	→	this time
will let you know if it changes	→	in the future

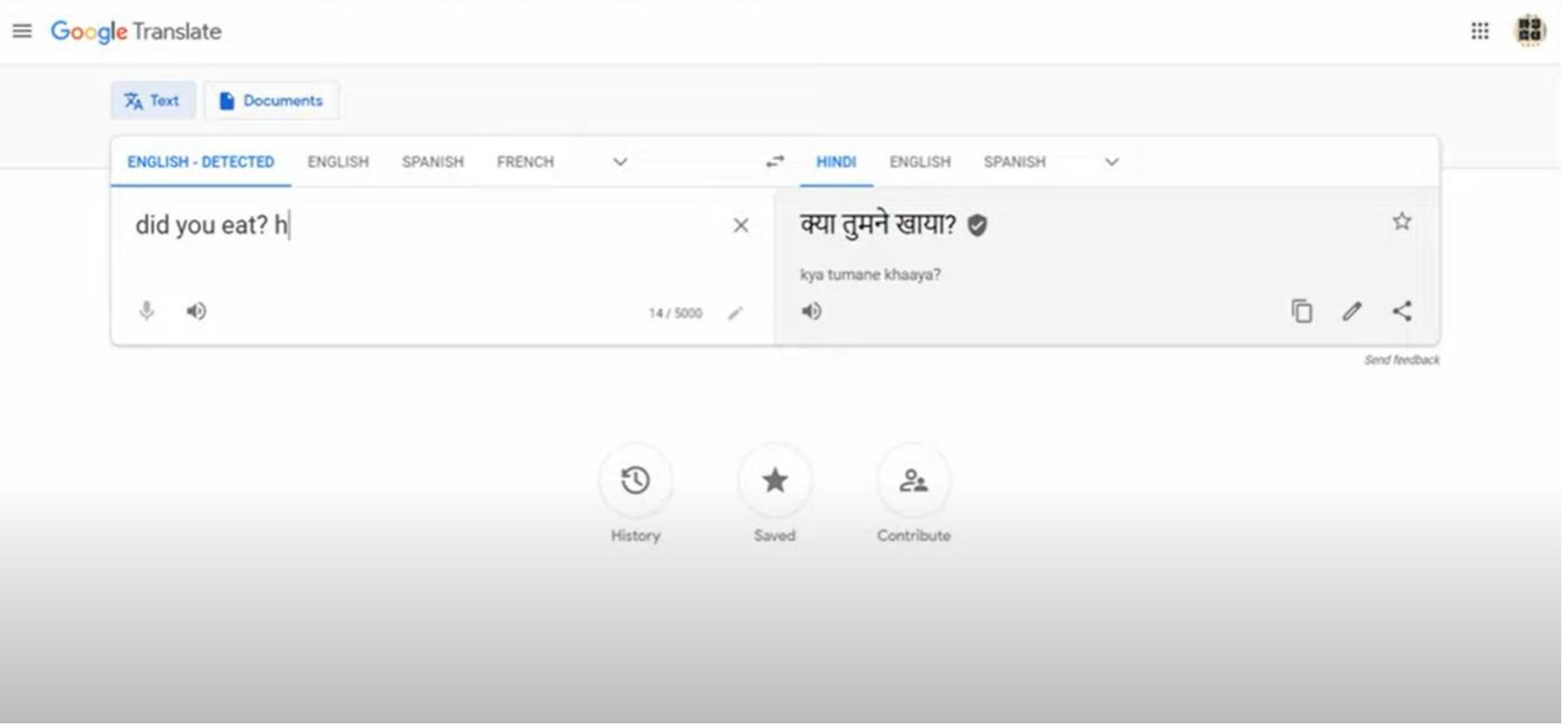
GALGOTIAS UNIVERSITY

NAAC GRADE A+ Accredited University

School of Computing Science and Engineering

Course Code : R1UC604C Course Name: Deep Learning




Google Translate
Text
Documents
ENGLISH - DETECTED
ENGLISH
SPANISH
FRENCH
HINDI
ENGLISH
SPANISH
did you eat? h
14 / 5000
क्या तुमने खाया?
kya tumane khaaya?
Send feedback
History
Saved
Contribute

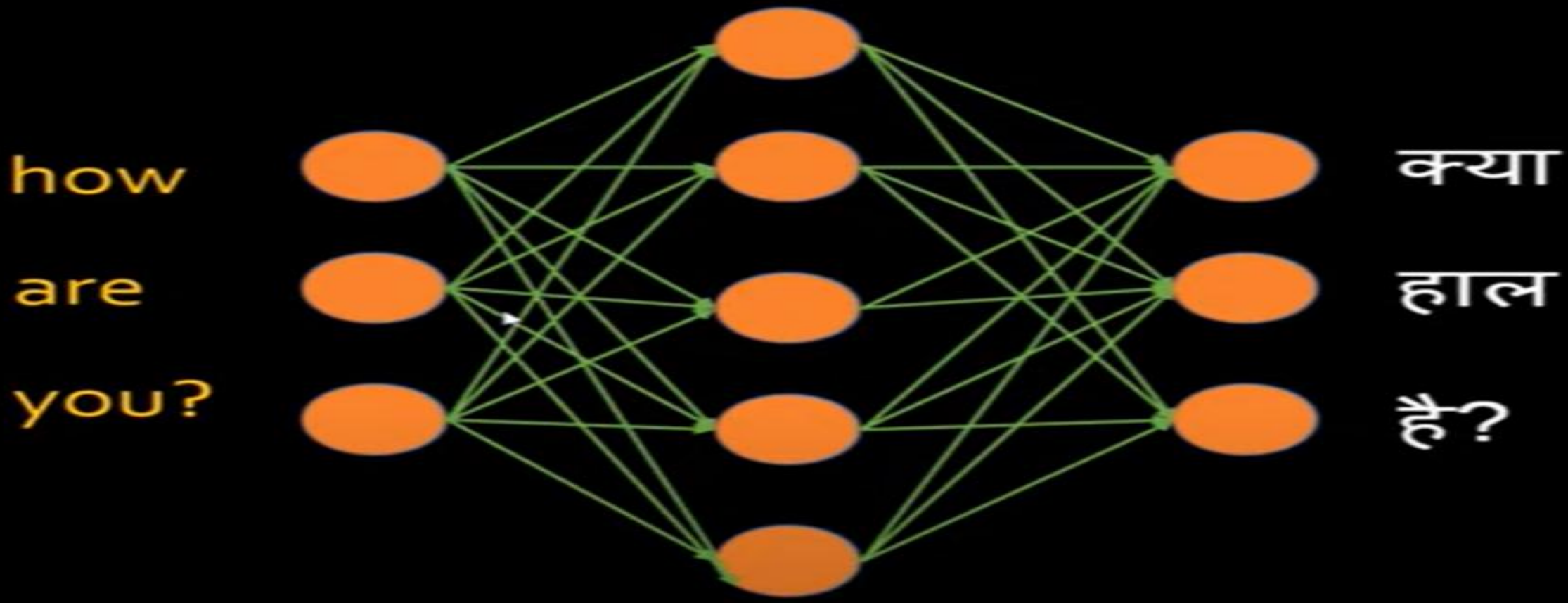
x Rudolph Smith bought 1000 shares of tesla Inc. in March 2020

y Rudolph Smith (Person) bought 1000 shares of tesla Inc. (Company) in March 2020 (time)

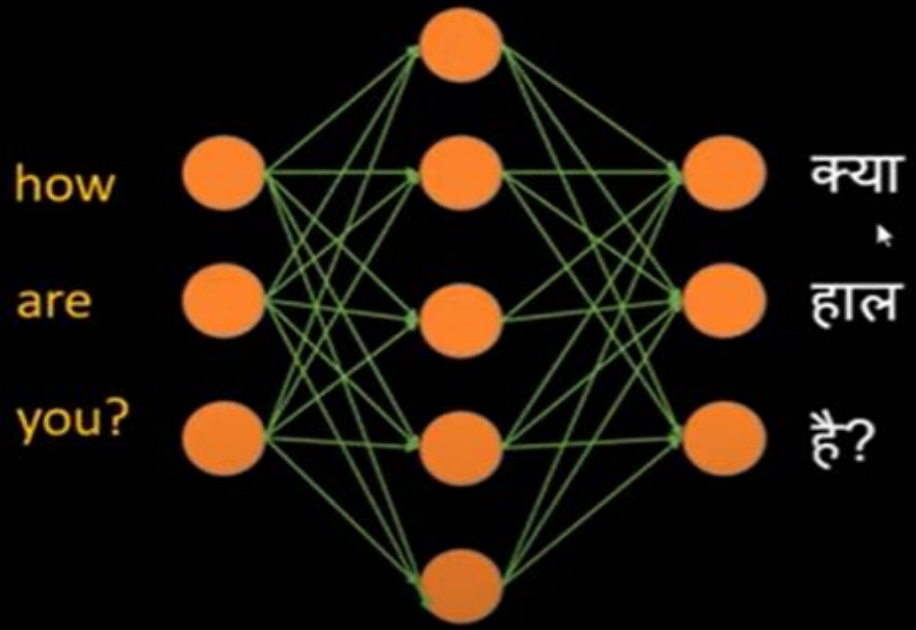
NER: Named Entity Recognition

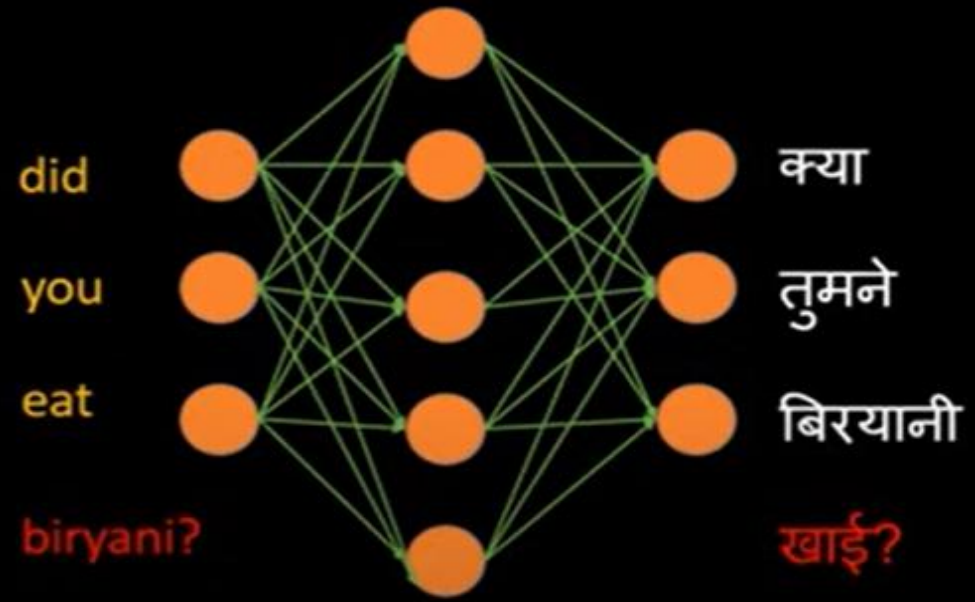
auto complete
not interested at
this time
translation
how are you?
क्या हाल है?
NER
Rudolph Smith bought 1000 shares of tesla Inc. in March 2020
Person
Company
time
Rudolph Smith bought 1000 shares of tesla Inc. in March 2020
Sentiment Analysis
Not only the fan was expensive, but it was broken when it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

Simple Artificial Neural Network?



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





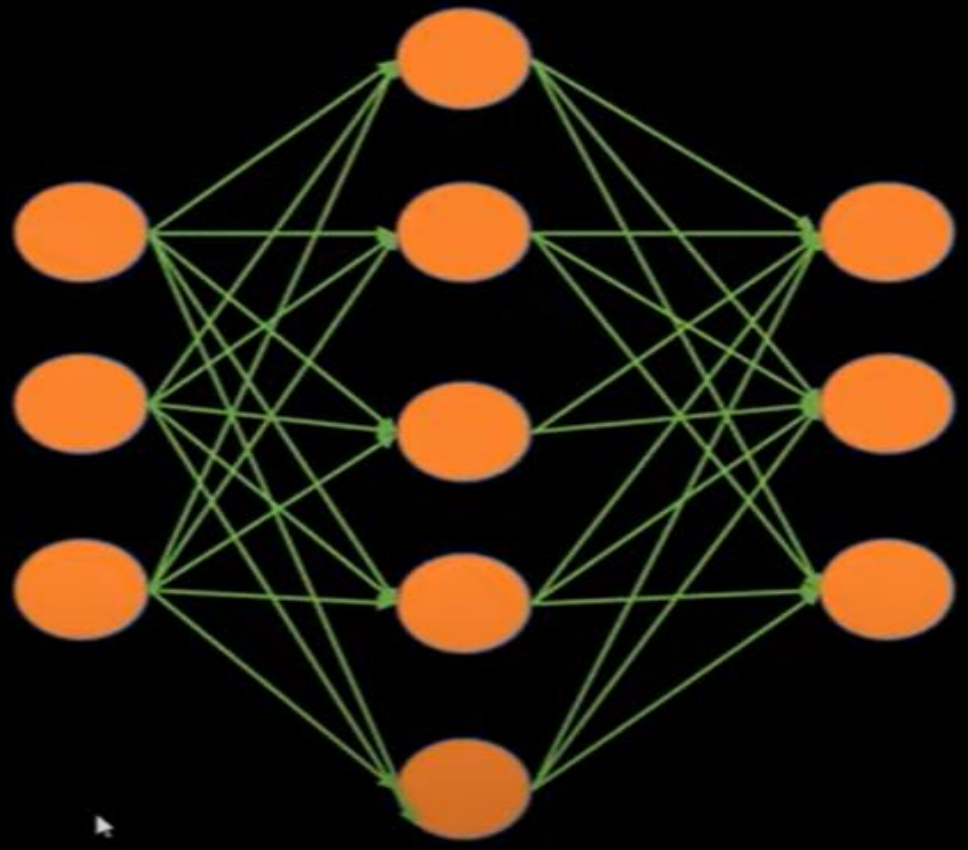
Issue # 2: Too much computation

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]



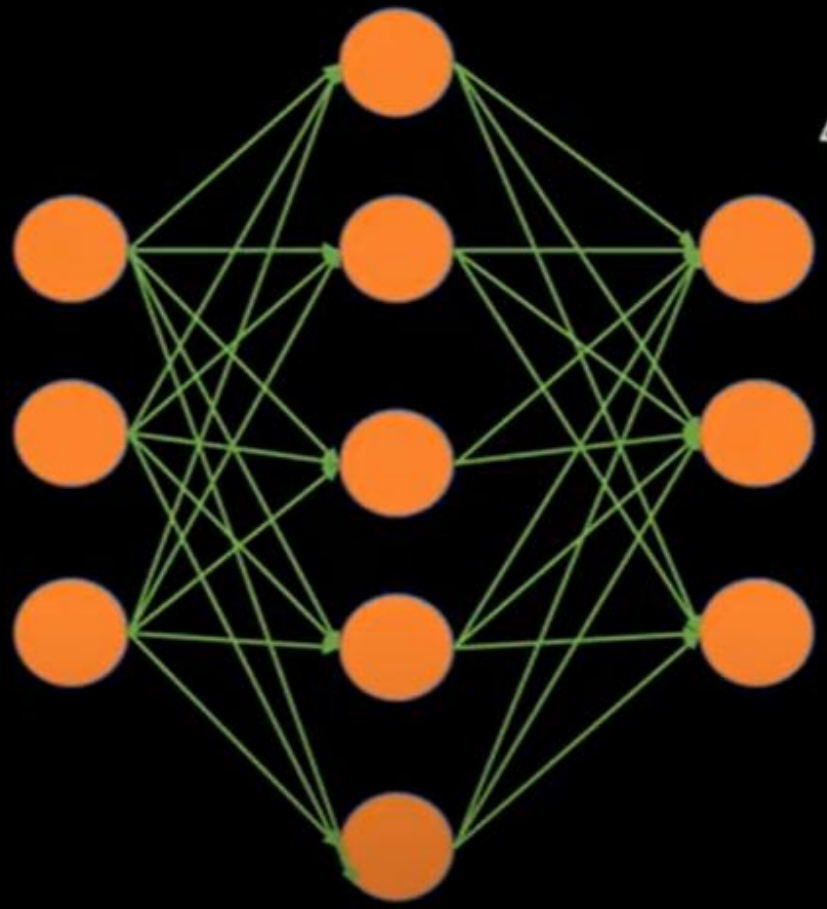
Issue # 2: Too much computation

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?

3 Issues using ANN for sequence problems

Variable size of input/output neurons

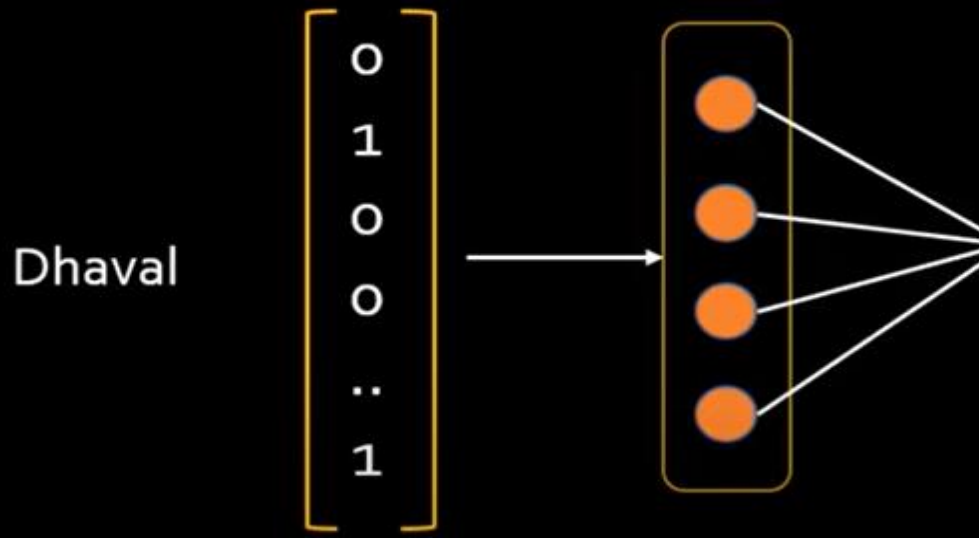
Too much computation

No parameter sharing

Dhaval loves baby yoda
person
person
Dhaval
loves
baby yoda
1
0
1
1

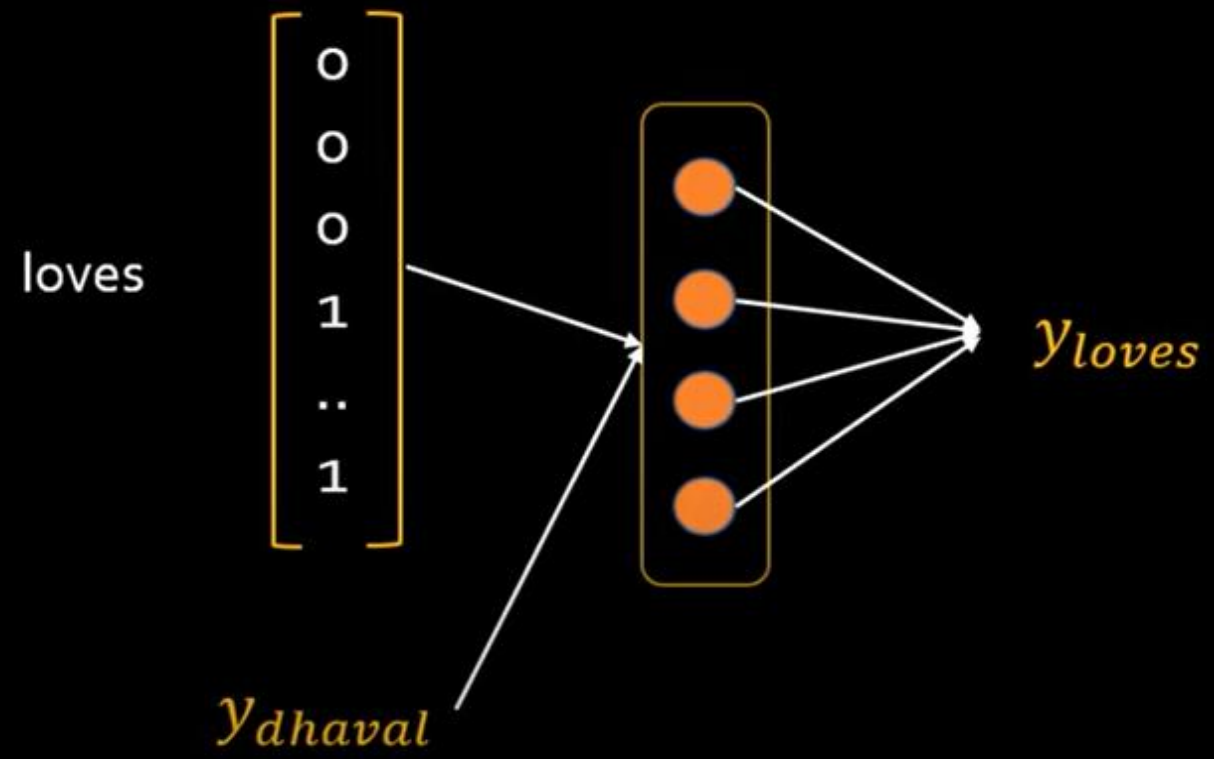
Named Entity Recognition

Dhaval loves baby yoda

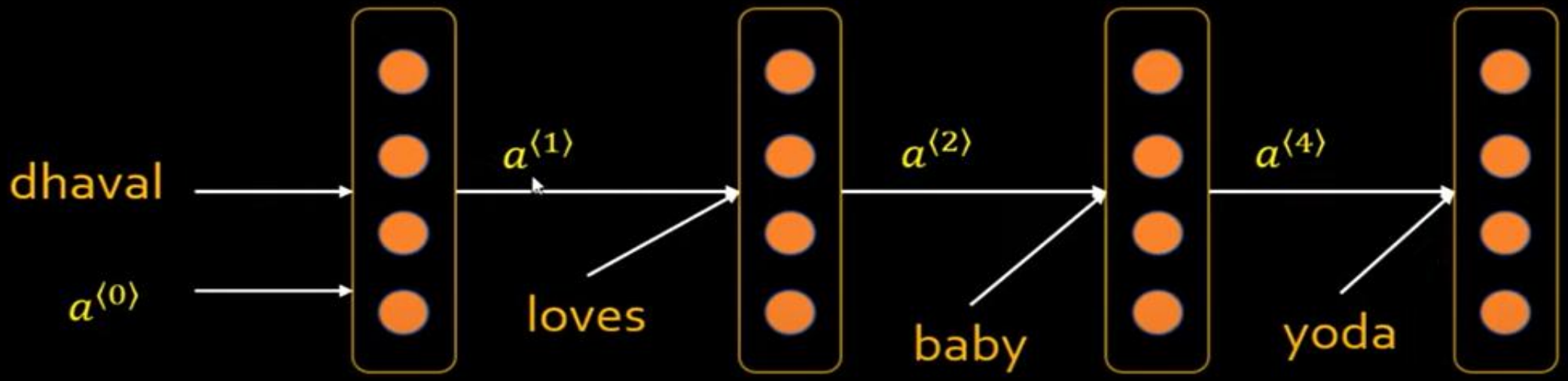


Named Entity Recognition

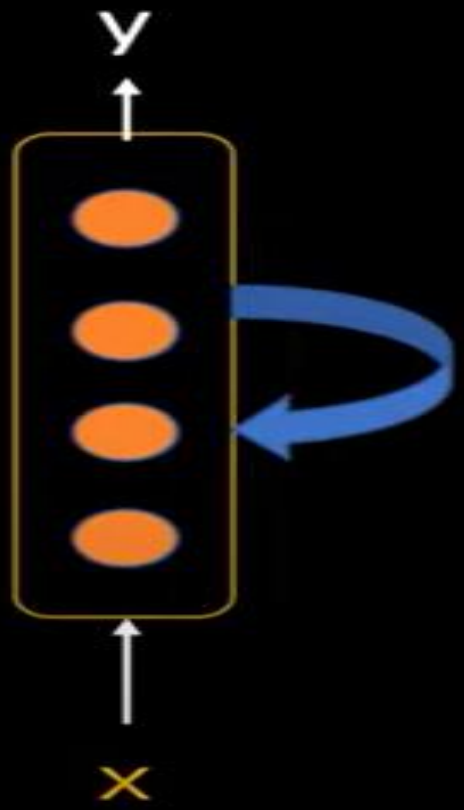
Dhaval loves baby yoda



Named Entity Recognition



Generic Representation of RNN

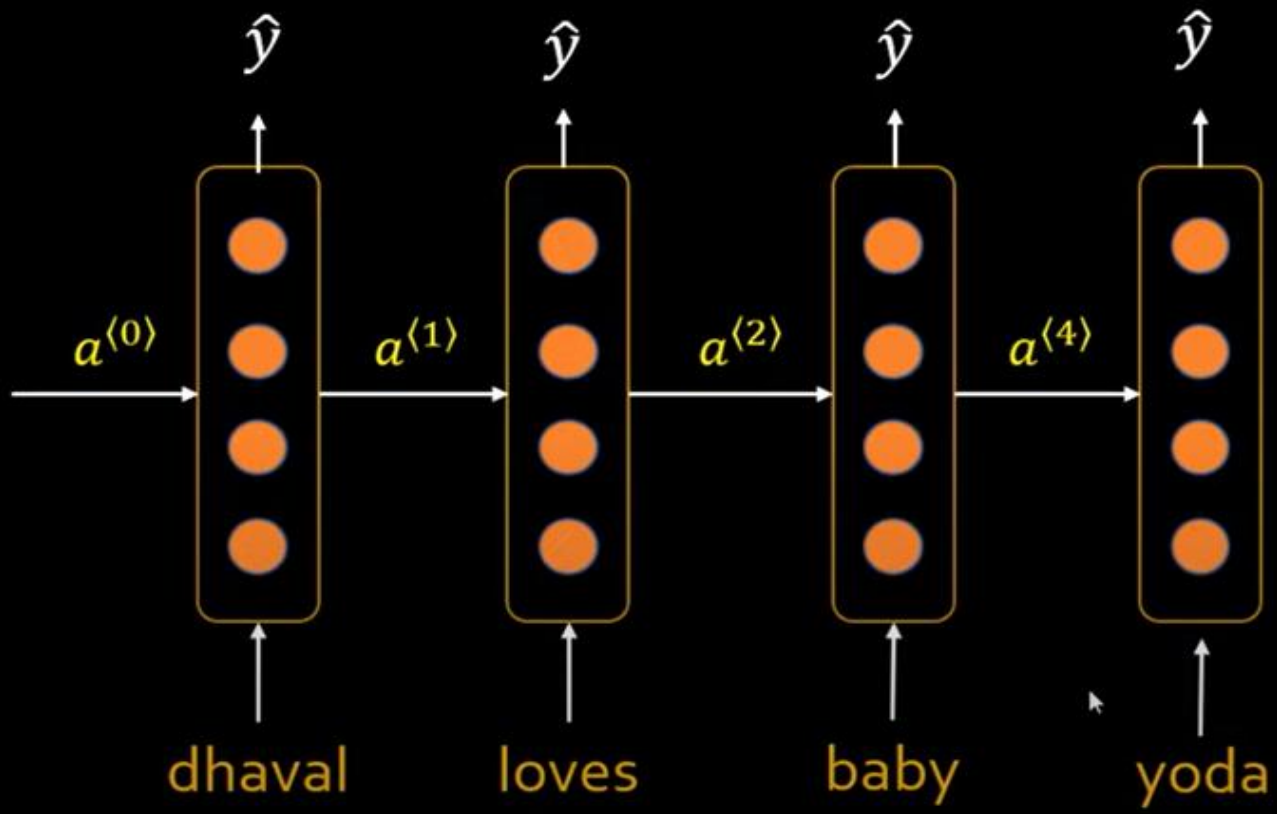


Training : Named Entity Recognition (NER)

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

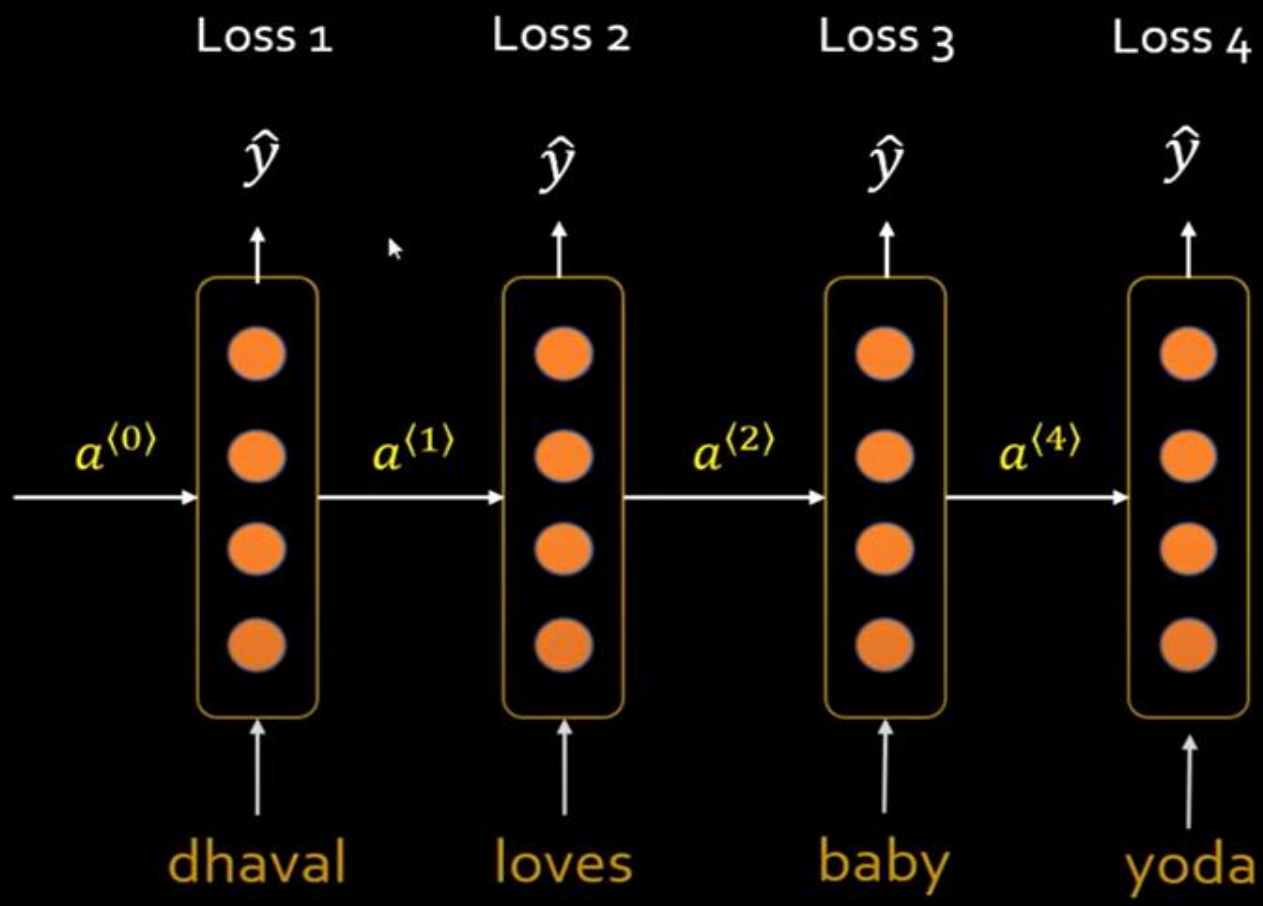
Training

Dhaval loves baby yoda → 1 0 1 1



Training

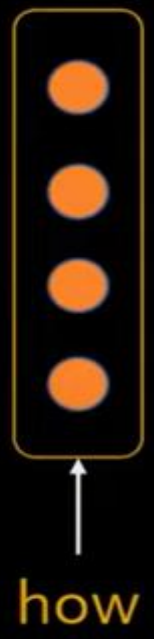
Dhaval loves baby yoda → 1 0 1 1



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

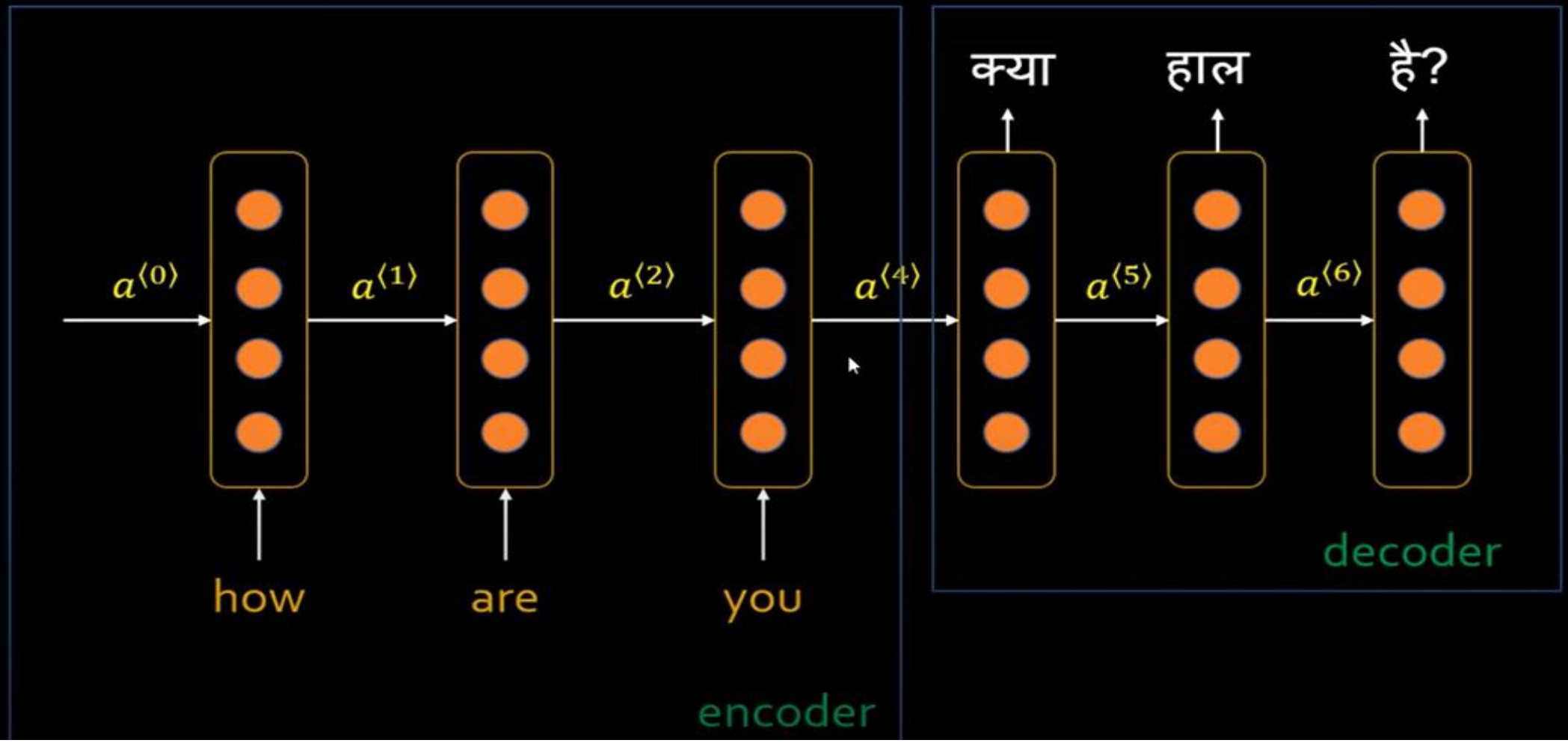
Language translation

how are you? क्या हाल है?

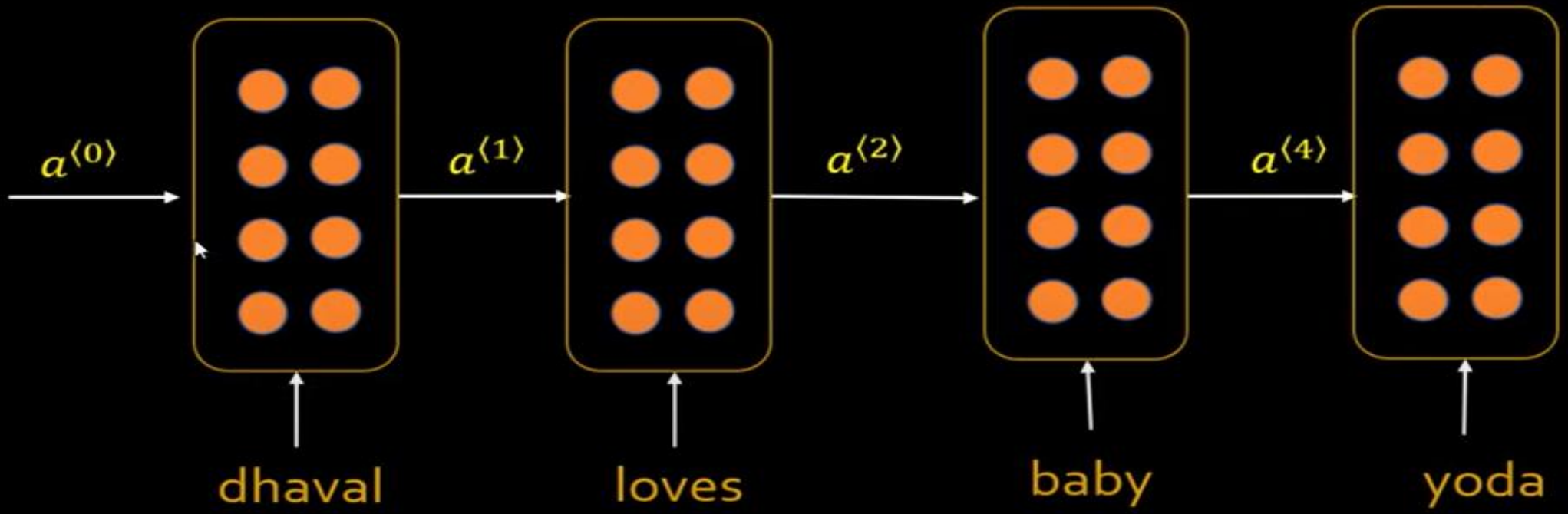


Language translation

how are you? क्या हाल है?

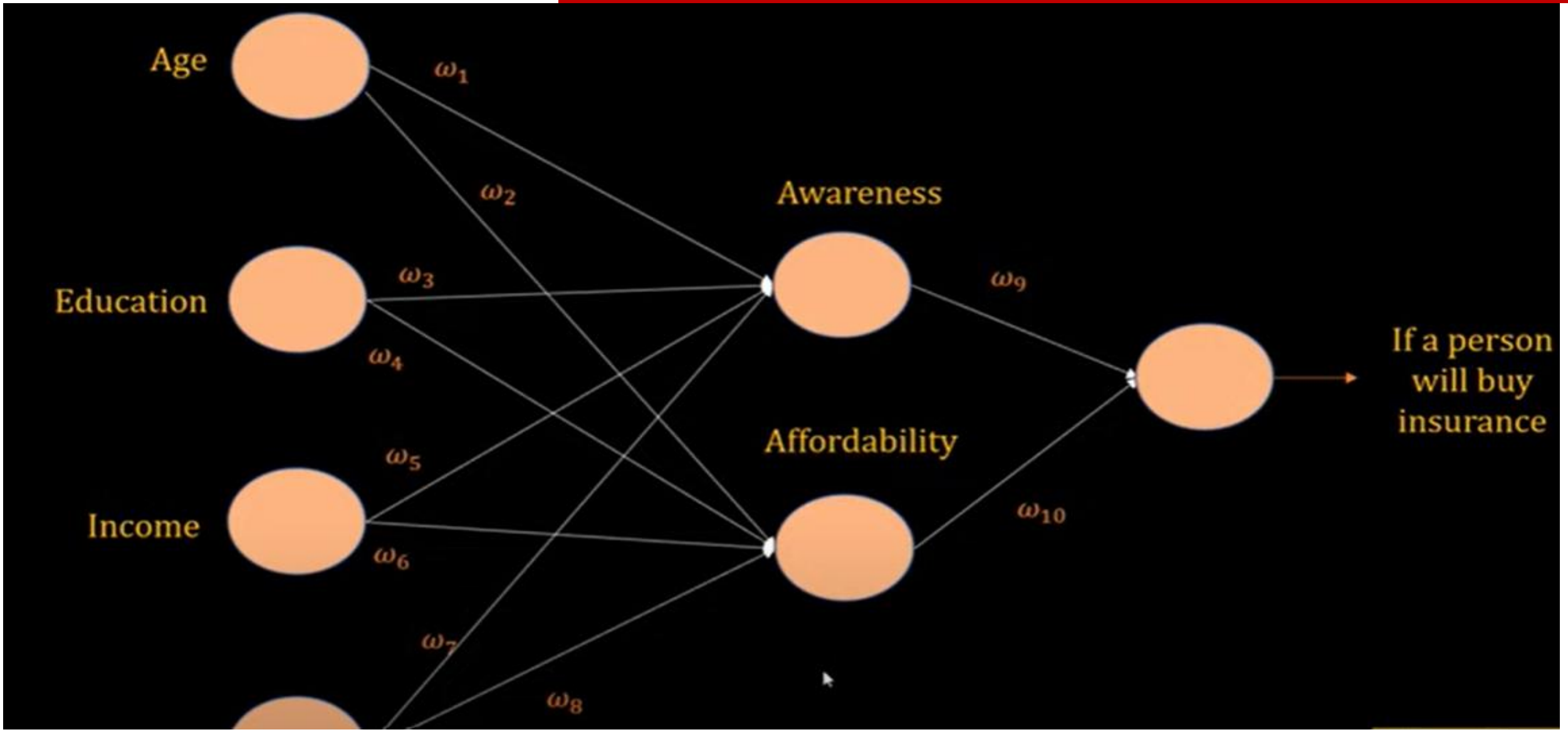


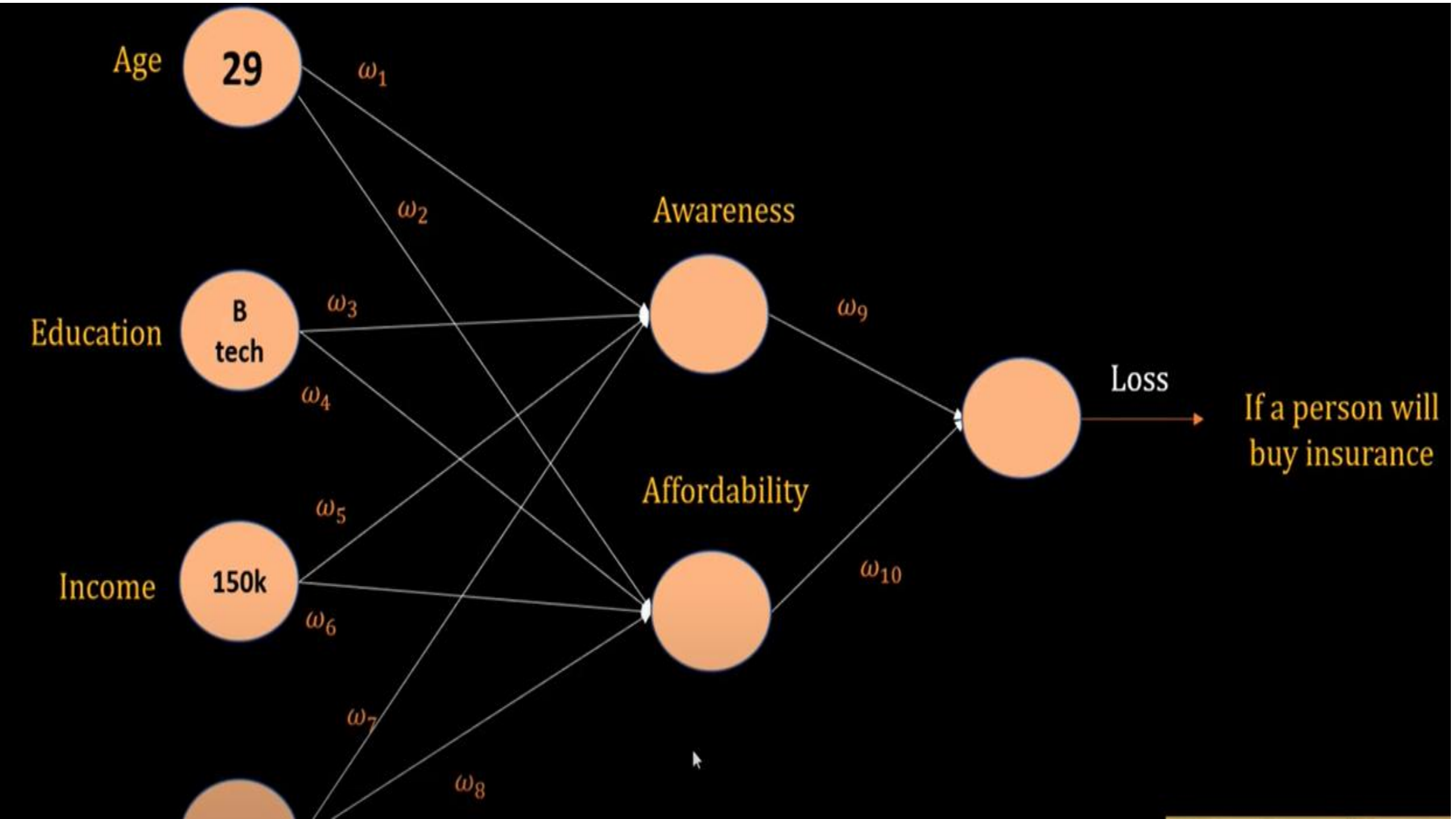
Deep RNN

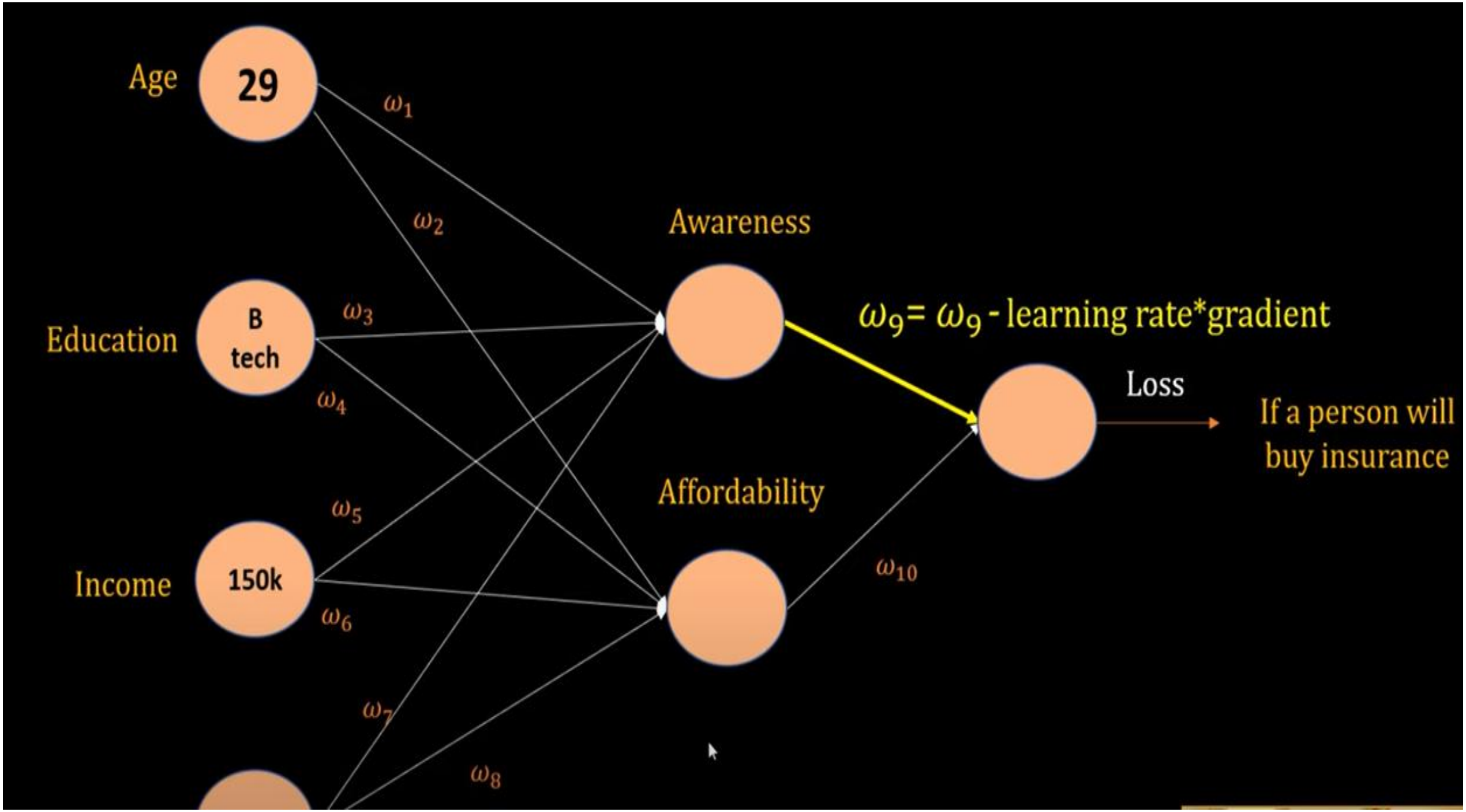

$a^{\langle 0 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle 4 \rangle}$
dhaval
loves
baby
yoda

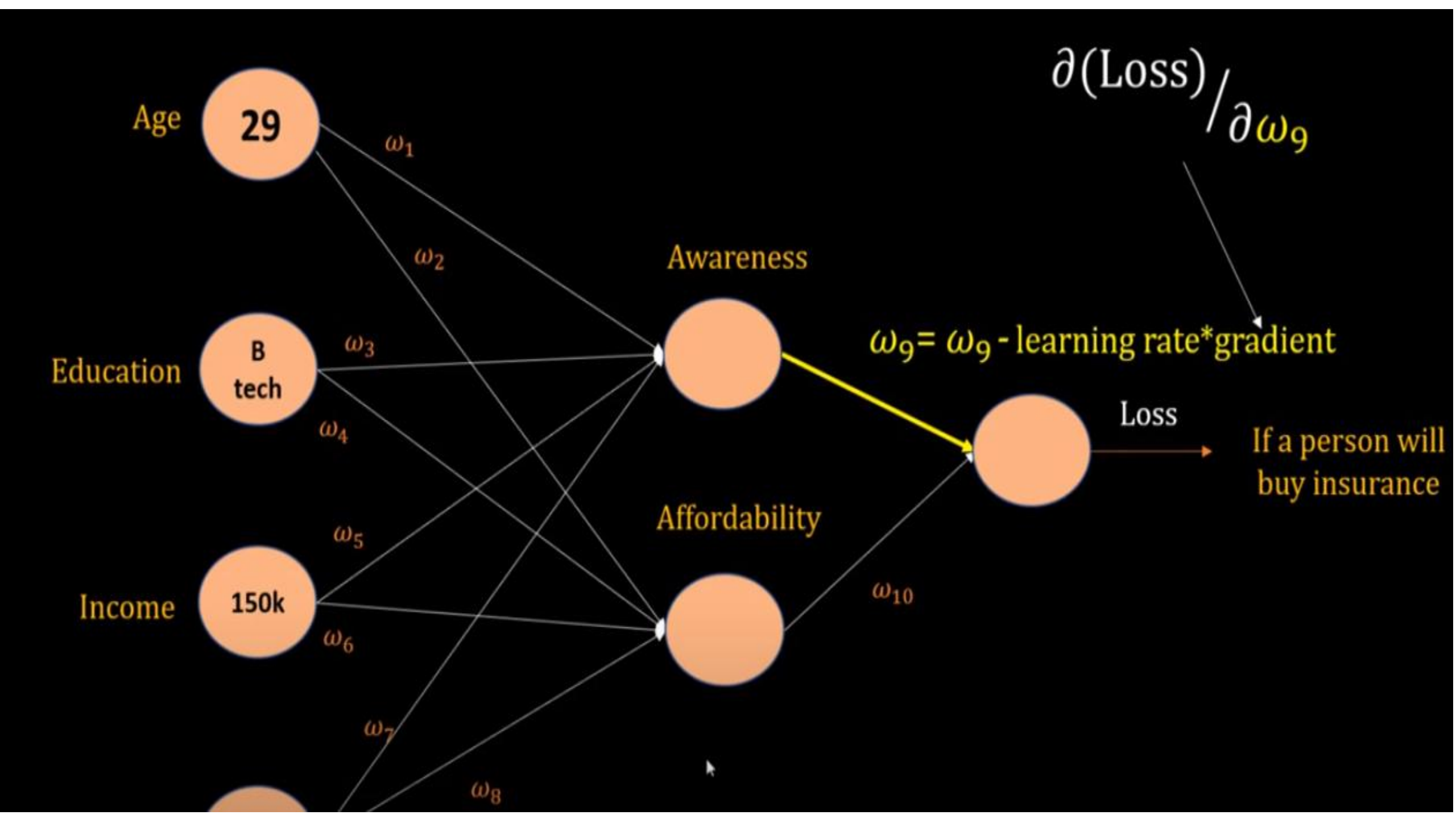
Long Shot-Term Memory

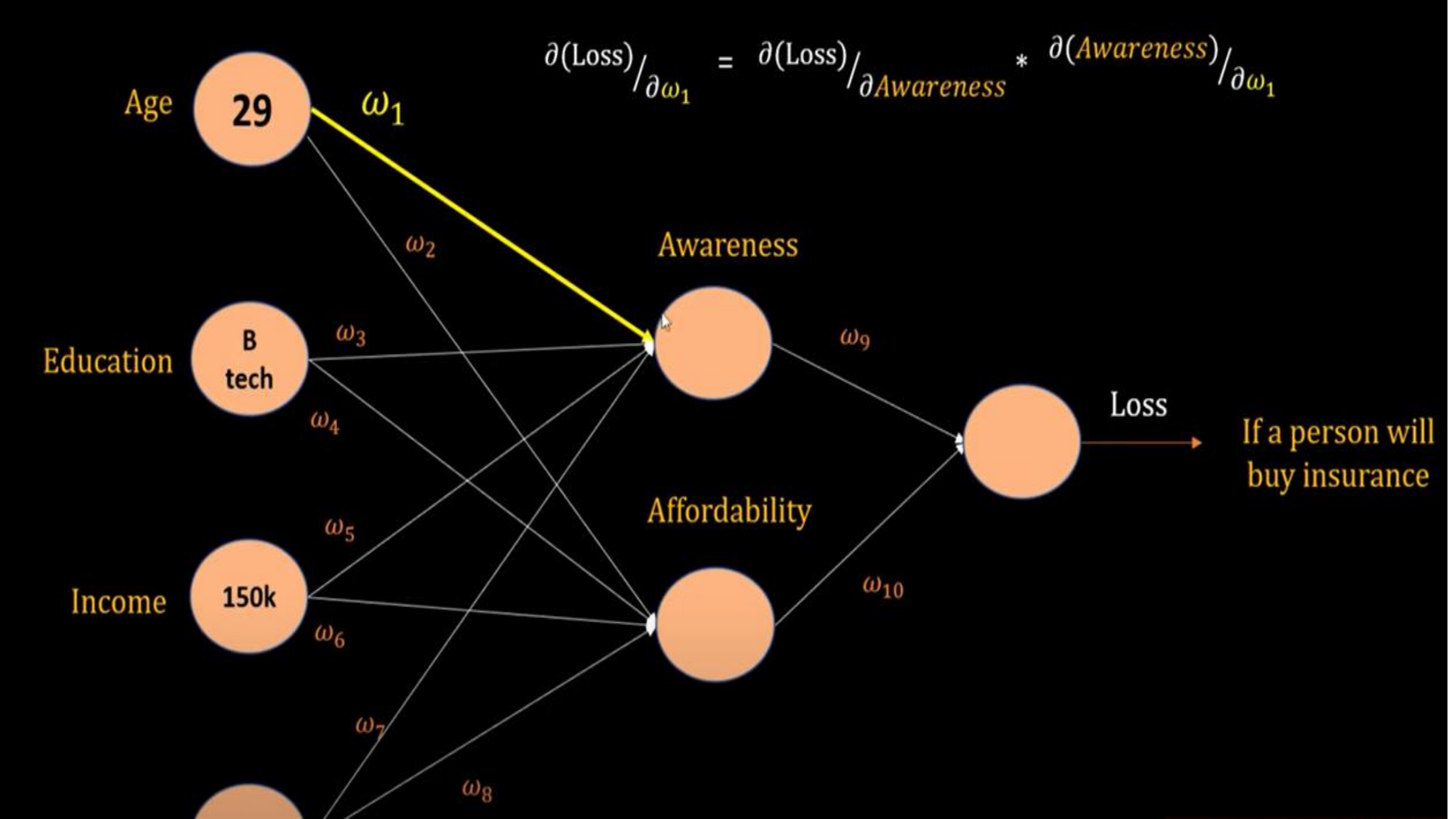
Vanishing Gradient Issue


Age
Education
Income
ω_1
ω_2
ω_3
ω_4
ω_5
ω_6
ω_7
ω_8
Awareness
Affordability
ω_9
ω_{10}
If a person will buy insurance


Age
29
Education
B tech
Income
150k
ω_1
ω_2
ω_3
ω_4
ω_5
ω_6
ω_7
ω_8
Awareness
Affordability
ω_9
ω_{10}
Loss
If a person will buy insurance


Age
29
Education
B tech
Income
150k
ω_1
ω_2
ω_3
ω_4
ω_5
ω_6
ω_7
ω_8
Awareness
Affordability
ω_9= ω_9 - learning rate*gradient
ω_{10}
Loss
If a person will buy insurance


$\partial(\text{Loss})/\partial\omega_9$
Age
29
ω_1
ω_2
Awareness
Education
B tech
ω_3
ω_4
$\omega_9 = \omega_9$ - learning rate*gradient
Loss
If a person will buy insurance
Affordability
ω_5
Income
150k
ω_6
ω_{10}
ω_7
ω_8


$\partial(\text{Loss})/\partial\omega_1 = \partial(\text{Loss})/\partial Awareness * \partial(Awareness)/\partial\omega_1$
Age
29
ω_1
ω_2
Awareness
Education
B tech
ω_3
ω_4
ω_9
Loss
If a person will buy insurance
Affordability
ω_5
Income
150k
ω_6
ω_{10}
ω_7
ω_8

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

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

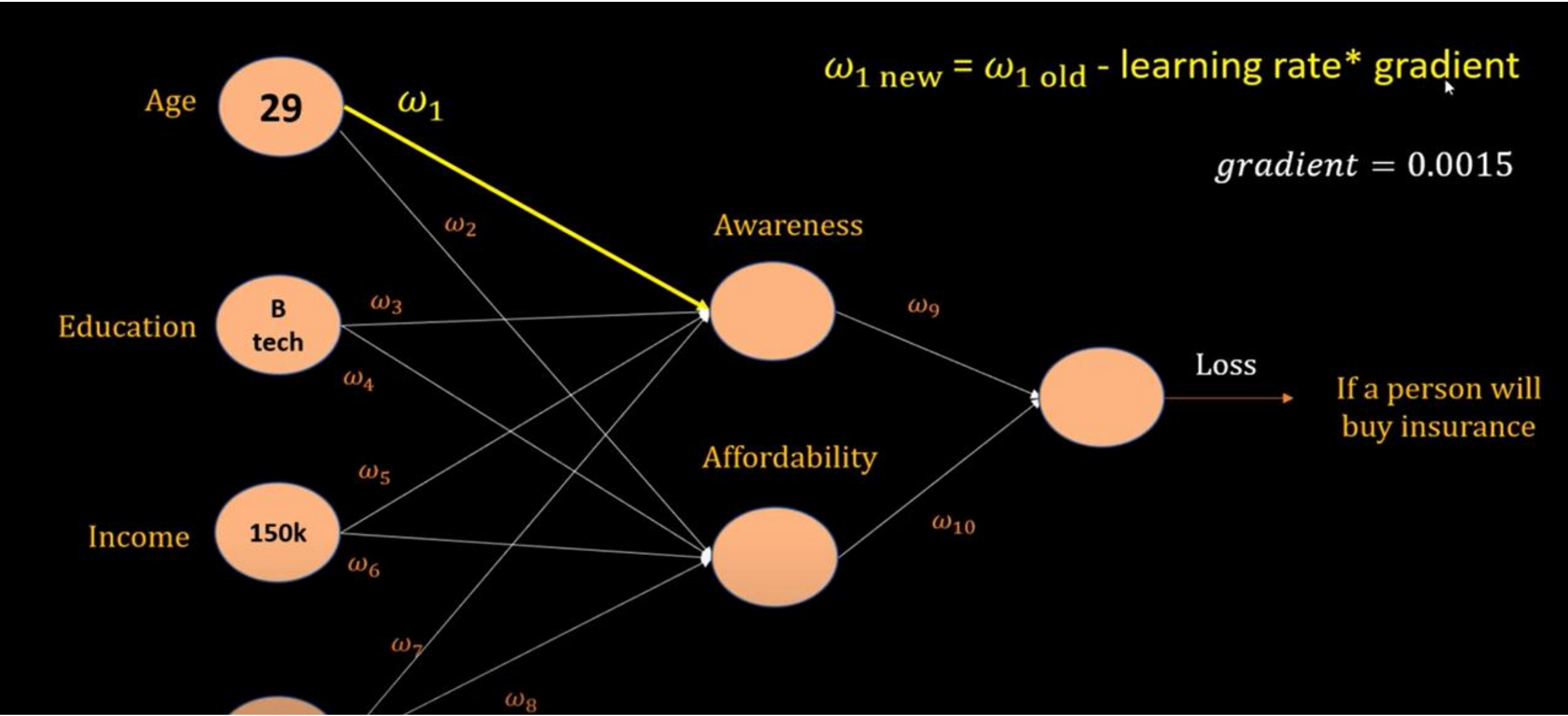
$$gradient = d1 * d2$$

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

$$gradient = d1 * d2$$

$$gradient = 0.03 * 0.05$$

$$gradient = 0.0015$$


$\omega_{1\,new} = \omega_{1\,old}$ - learning rate* gradient
$gradient = 0.0015$
Age
29
ω_1
ω_2
Education
B tech
ω_3
ω_4
Income
150k
ω_5
ω_6
ω_7
ω_8
Awareness
Affordability
ω_9
ω_{10}
Loss
If a person will buy insurance

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

Vanishing Gradients

When individual derivatives are large, the final derivate will also become huge and weights would change drastically.

Exploding Gradients

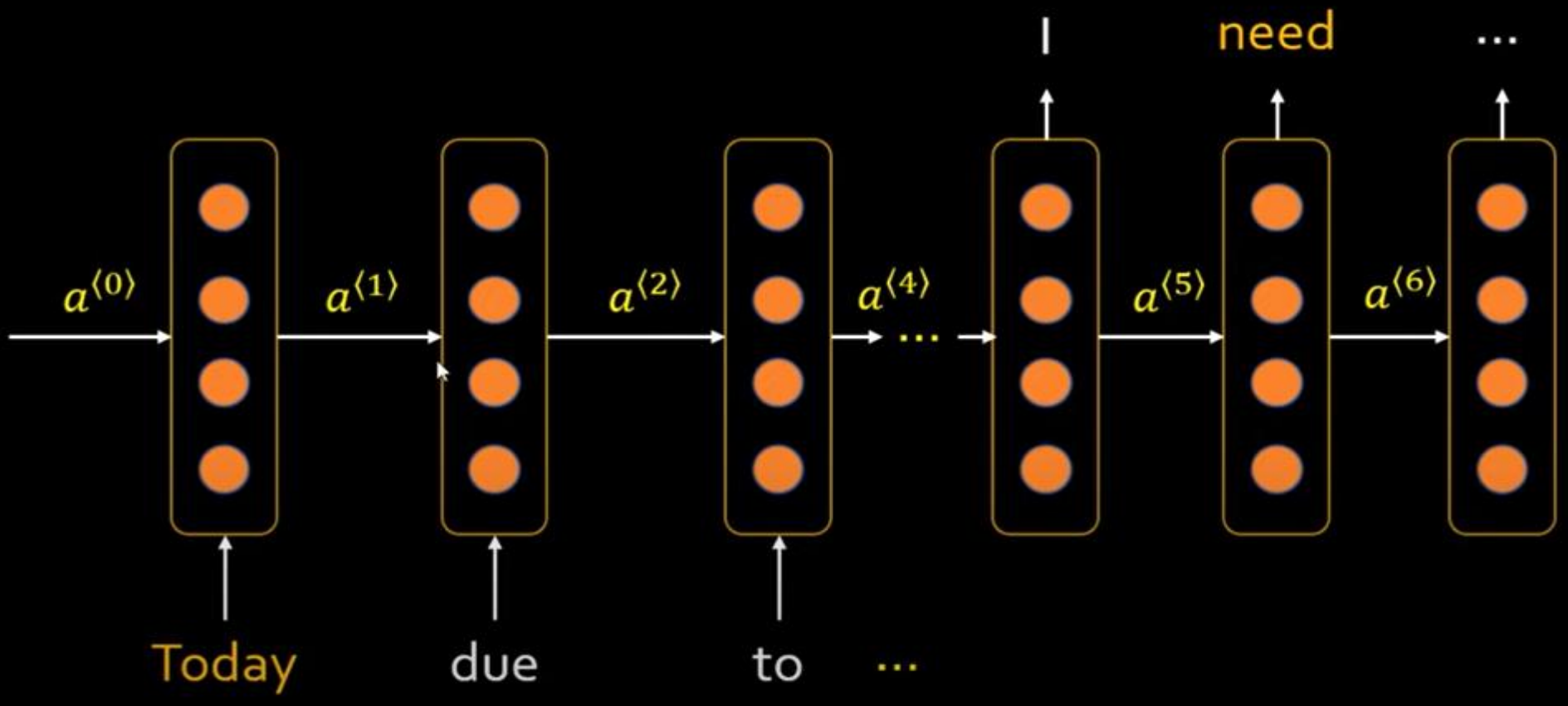
$$gradient = d1 * d2 * d3 * d4 * \ldots * dn$$

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

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.



Thank You