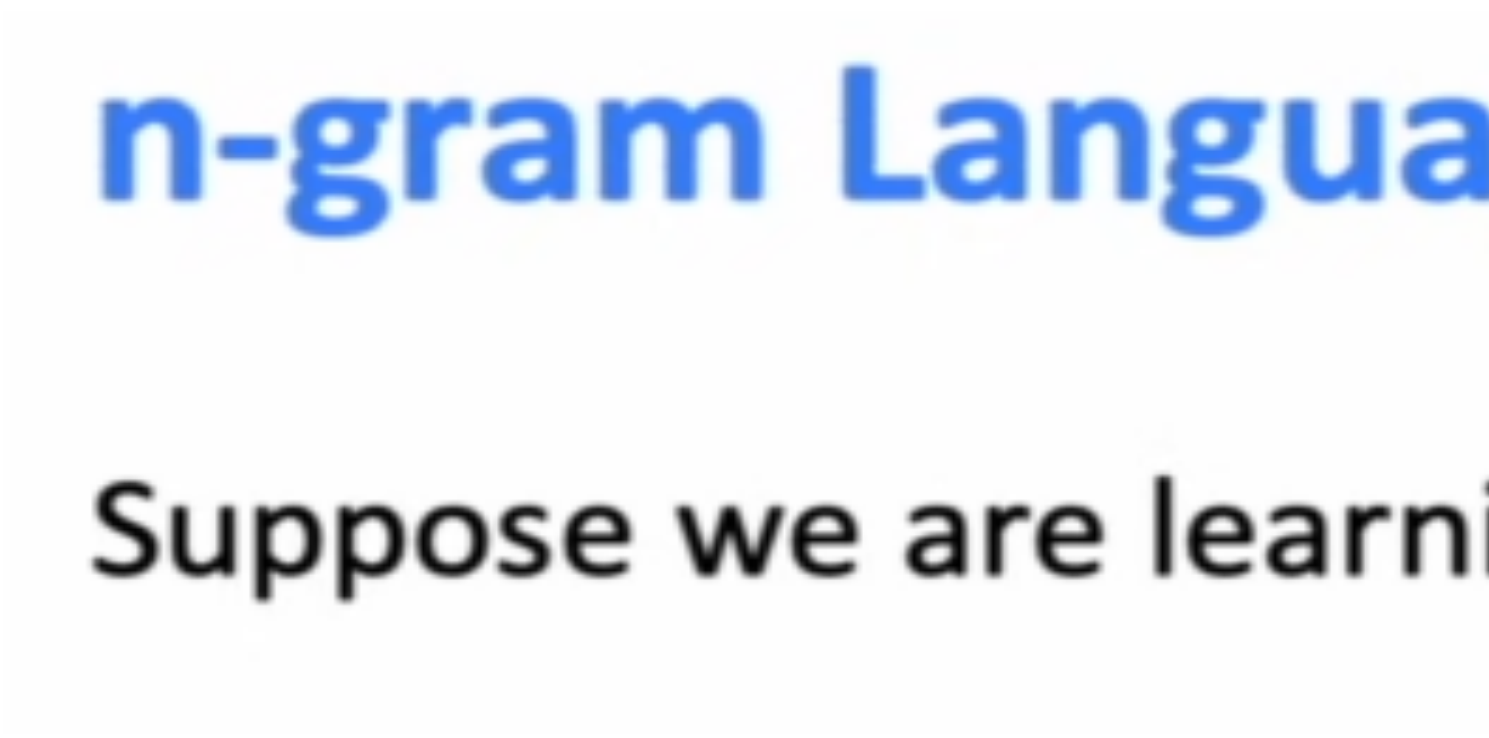
RNNs Notes Final

Sunday, October 17, 2021 7:54 AM

https://www.kaggle.com/kcsener/8-recurrent-neural-network-rnn-tutorial
https://www.kaggle.com/prashant111/comprehensive-guide-to-rnn-with-keras
https://www.kaggle.com/shivamb/beginners-guide-to-text-generation-using-lstms
https://towardsdatascience.com/beautifully-illustrated-nlp-models-from-rnn-to-transformer-8



0d69faf2109

ge Models: Exam

ing a 4-gram Language M

ple

odel.

~~as the proctor sta~~

discar

$$P(\boldsymbol{w}|\text{students op}$$

For example, suppose th

- "students opened th
- "students opened th

~~rted the clock, the~~ *studen*

rd

conc

$$\text{pened their}) = \frac{\text{count(stud}}{\text{count(stu}}$$

hat in the corpus:

eir" occurred 1000 times

eir books" occurred 400 tim

nts opened their ____

dition on this

$$\frac{\text{lents opened their } \boldsymbol{w})}{\text{udents opened their})}$$

es

- → P(books | stud
- "students opened th
  - → P(exams | stud

Sparsity Problem

Sparsity Problem 1

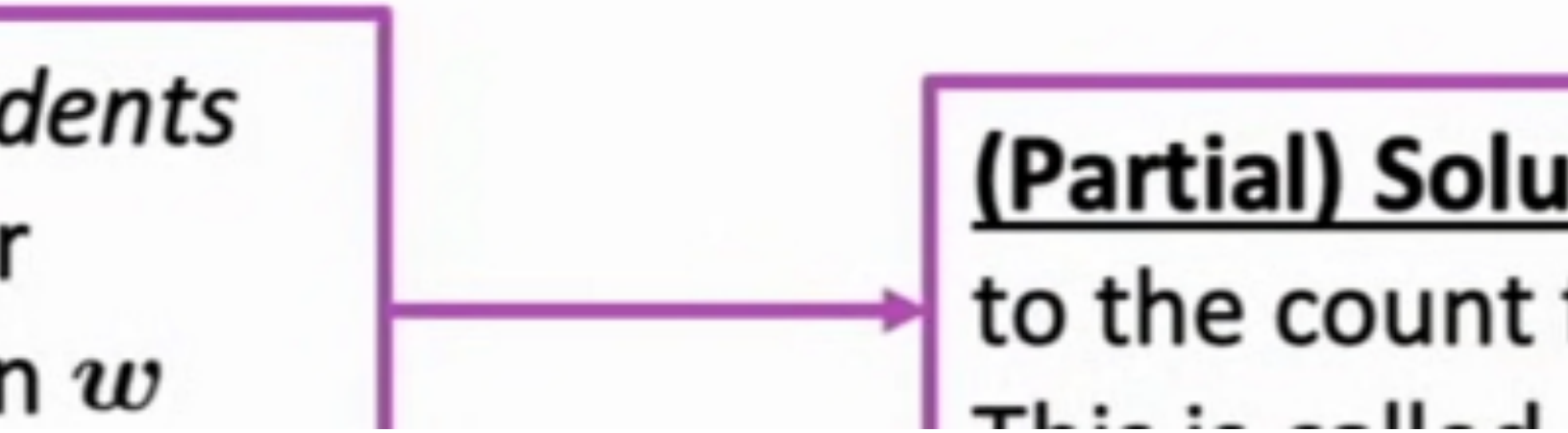
Problem: What if *"stu*
opened their w" neve
occurred in data? Ther

dents opened their) = 0.4

eir exams" occurred 100 tim

dents opened their) = 0.1

ns with n-gram Lang



nes

guage Models

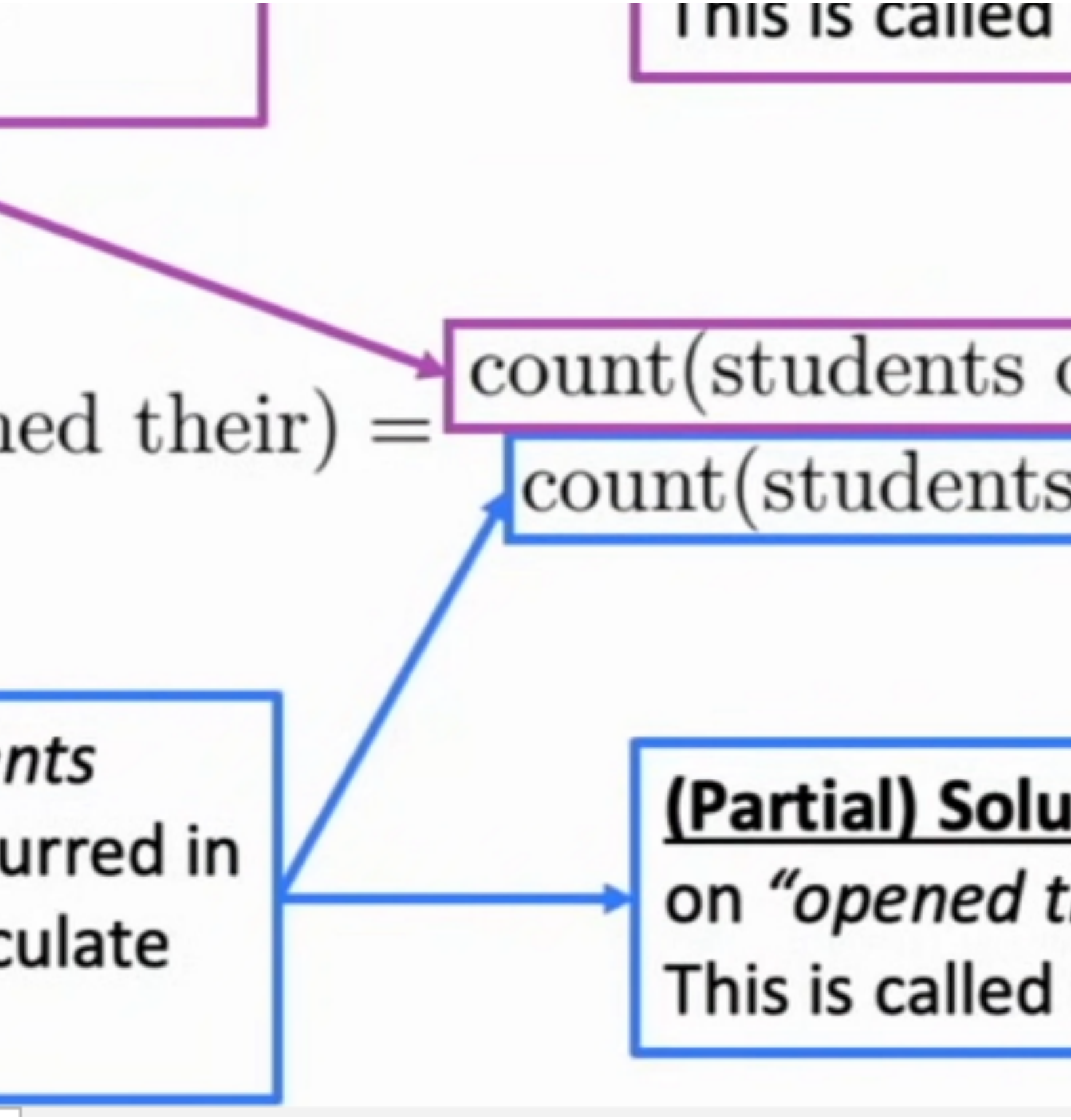
tion: Add small δ
for every $w \in V$.

has probability 0!

$$P(\boldsymbol{w}|\text{students open}$$

Sparsity Problem 2

Problem: What if *"stude opened their"* never occ data? Then we can't calc probability for *any* w!


This is called
ned their) =
count(students
count(students
nts
urred in
culate
(Partial) Solu
on "opened t
This is called

smoothing.

opened their $\boldsymbol{w}$)

s opened their)

tion: Just condition
heir" instead.
backoff.

A RNN Languag

output distribution

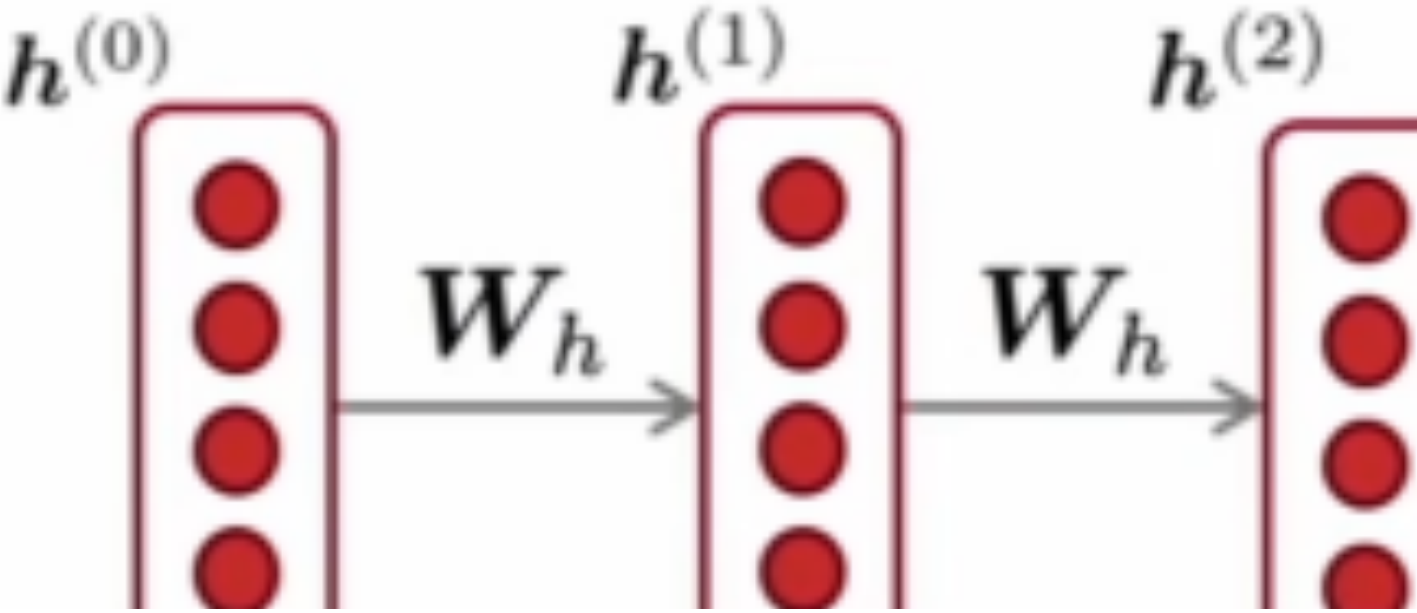
$$\hat{\boldsymbol{y}}^{(t)} = \text{softmax}\left(\boldsymbol{U}\boldsymbol{h}^{(t)} + \boldsymbol{b}_2\right) \in \mathbb{R}^{|V|}$$

hidden states

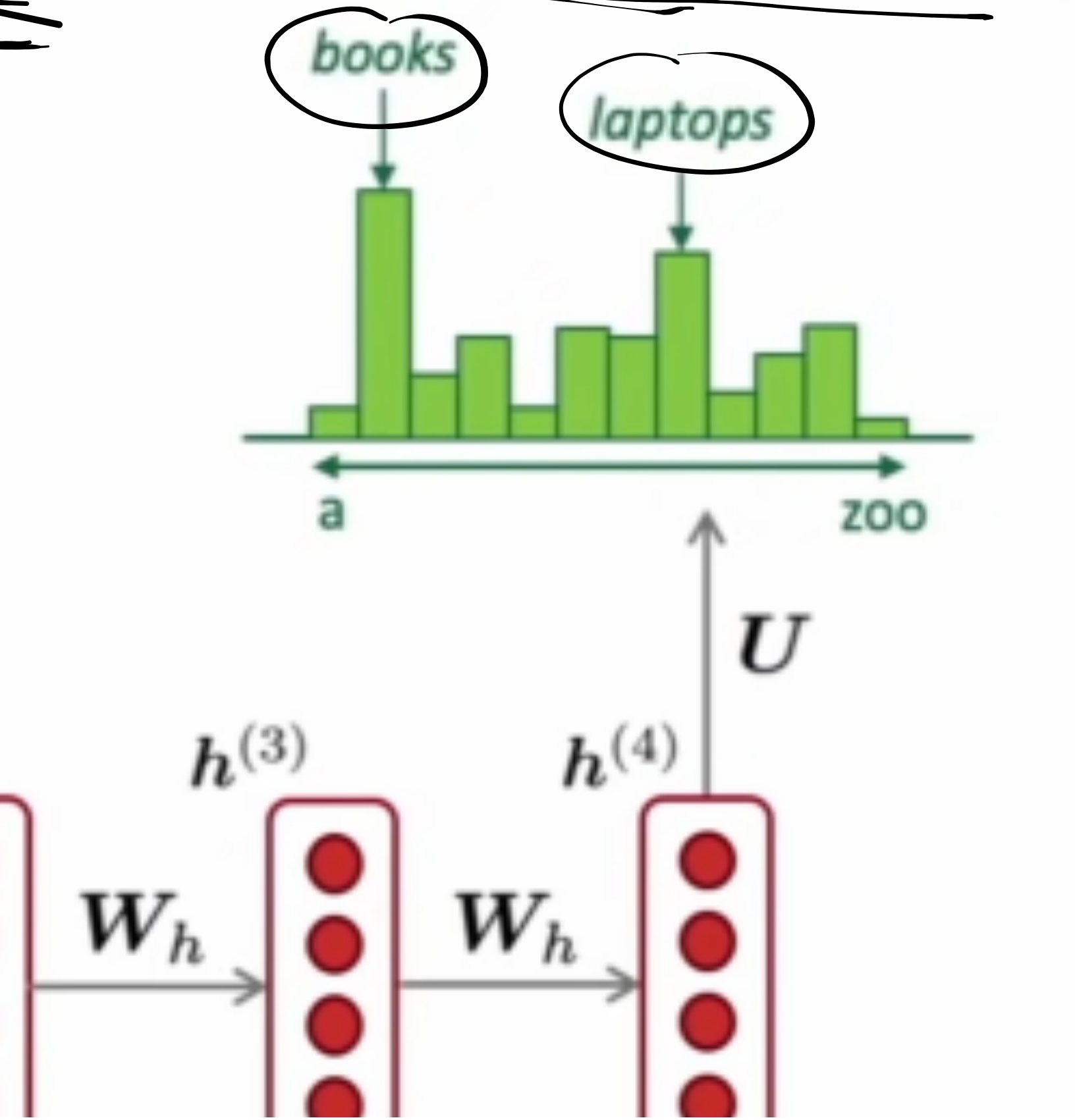
$$\boldsymbol{h}^{(t)} = \sigma\left(\boldsymbol{W}_h\boldsymbol{h}^{(t-1)} + \boldsymbol{W}_e\boldsymbol{e}^{(t)} + \boldsymbol{b}\right.$$

$\boldsymbol{h}^{(0)}$ is the initial hidden state

ge Model



$$\hat{y}^{(4)} = P(\boldsymbol{x}^{(5)}|\text{the students opened their})$$



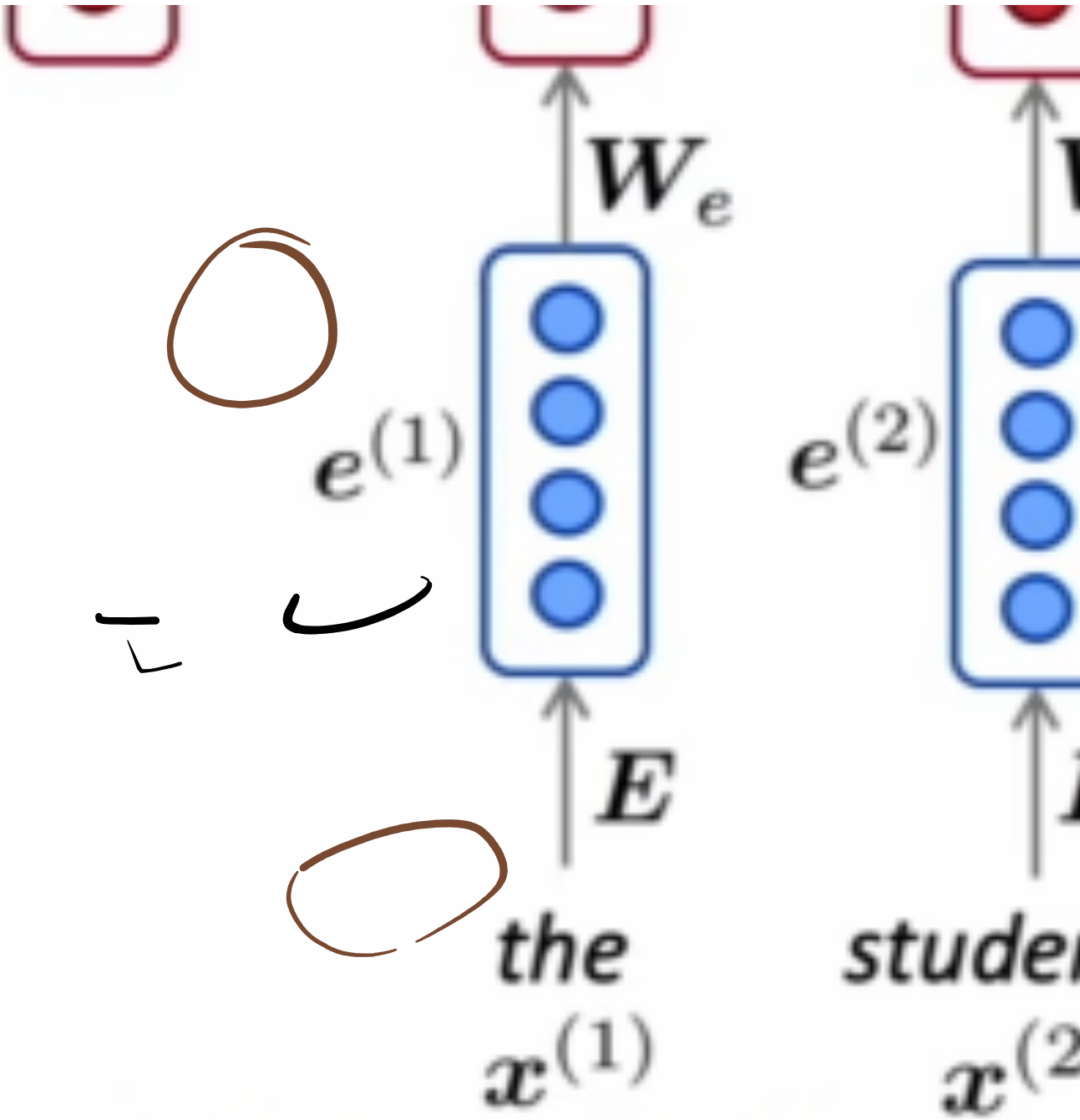
word embeddings

$$e^{(t)} = Ex^{(t)}$$

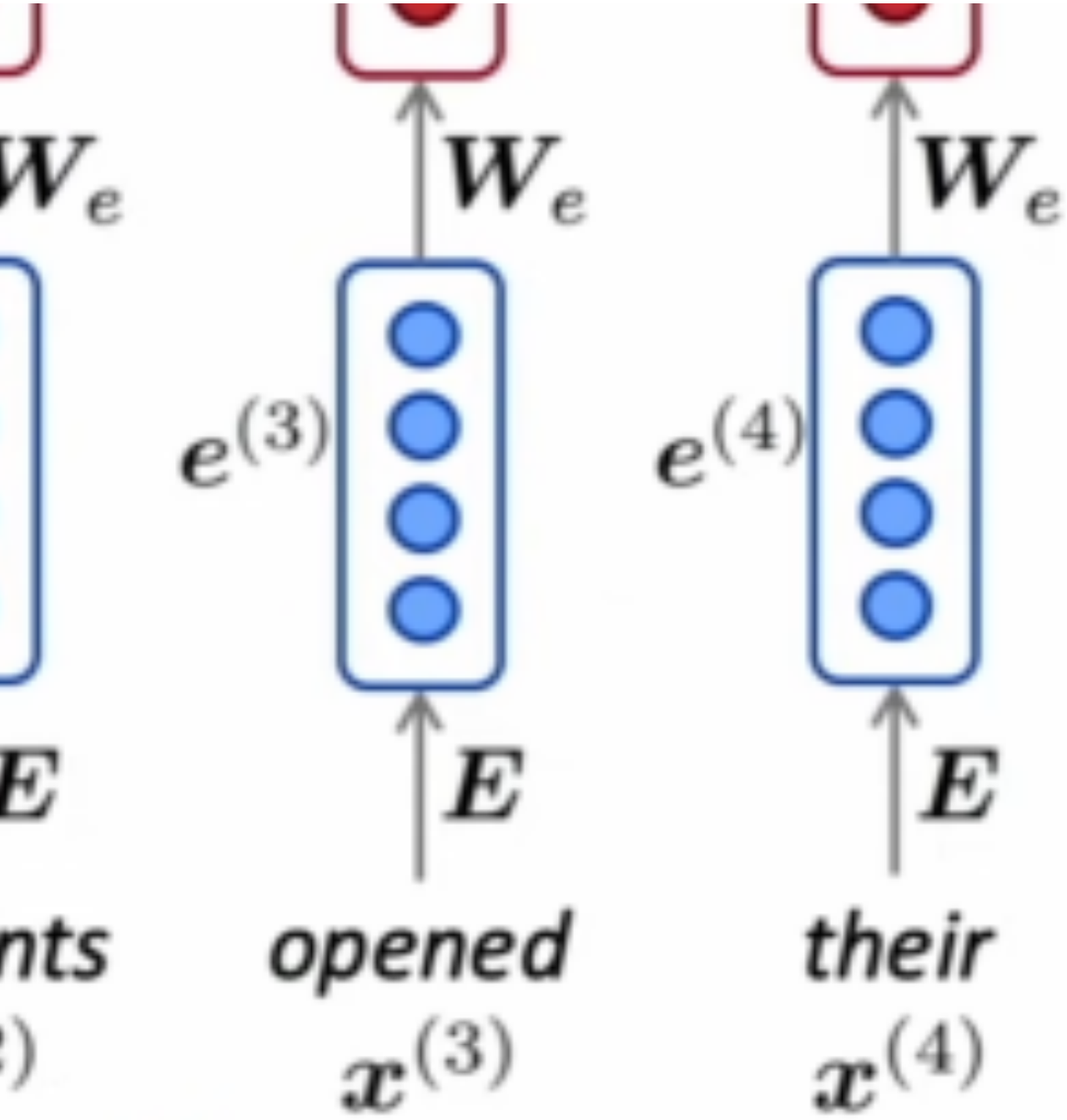
words / one-hot vectors

$$x^{(t)} \in \mathbb{R}^{|V|}$$

Note:
longe



this input sequence could be much
r, but this slide doesn't have space!

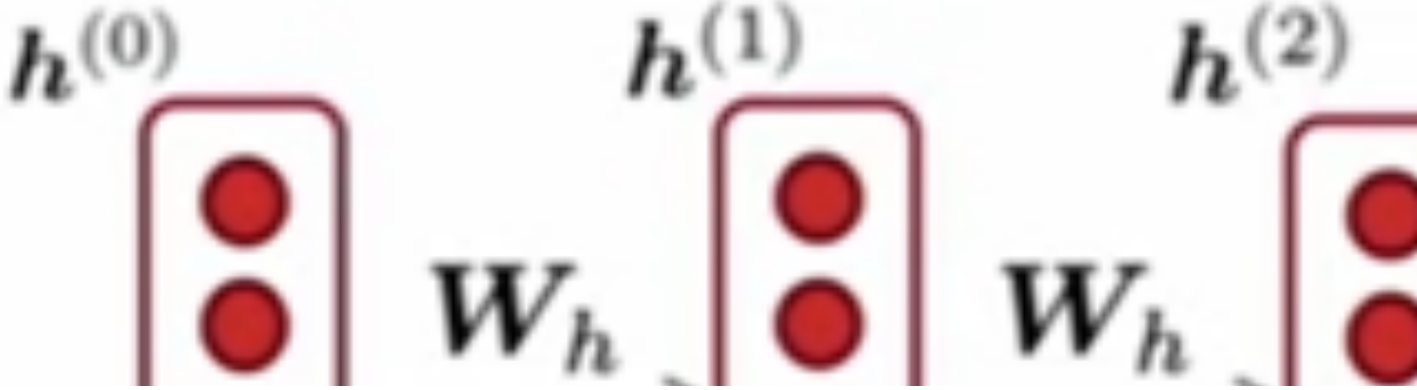

W_e
W_e
W_e
$e^{(3)}$
$e^{(4)}$
E
E
E
nts
opened
their
$x^{(3)}$
$x^{(4)}$

A RNN Languag

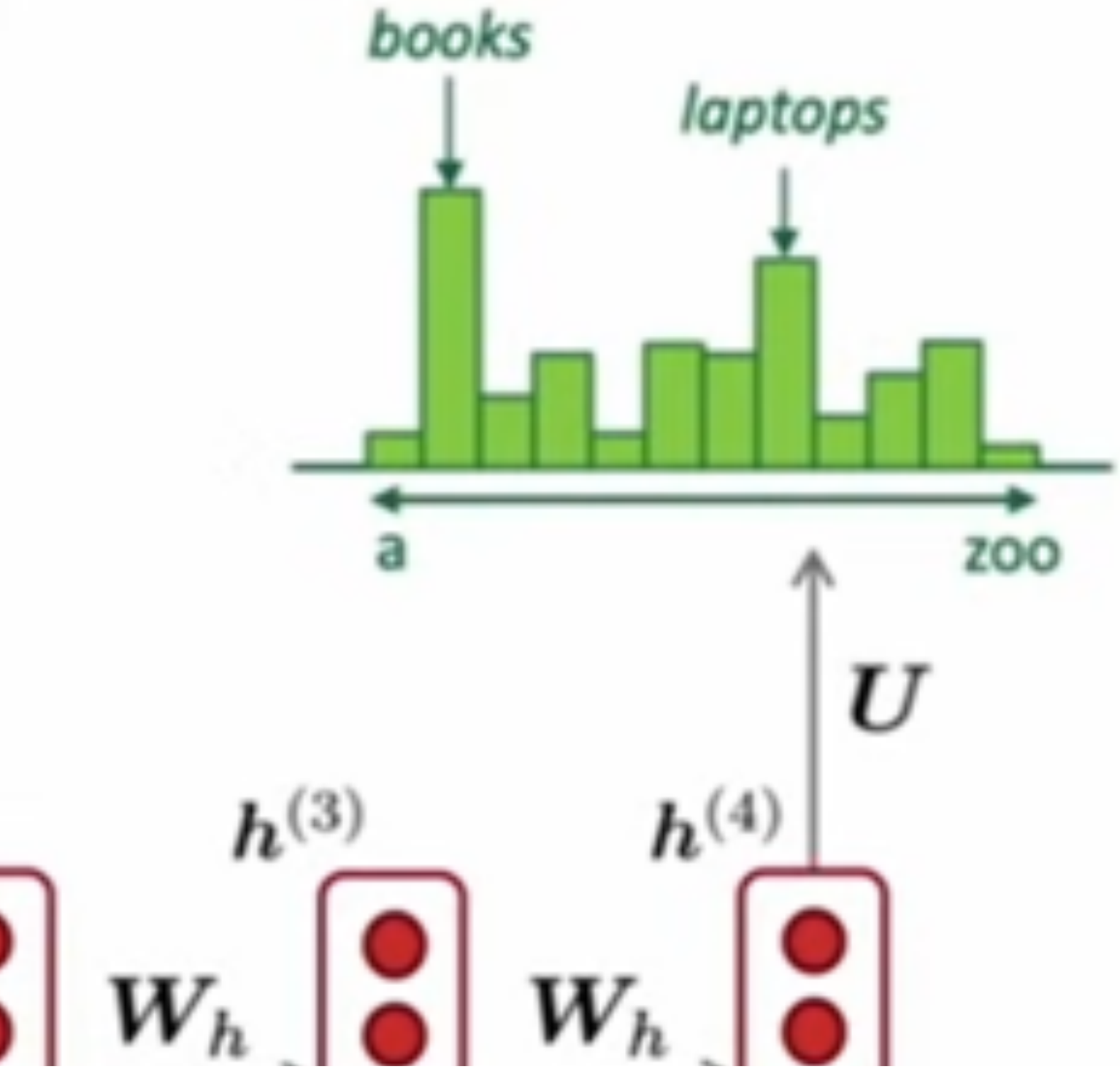
RNN **Advantages**:

- Can process any length input
- Computation for step *t* can (in theory) use information from many steps back
- Model size doesn't

ge Model



$\hat{y}^{(4)} = P(\boldsymbol{x}^{(5)}|\text{the students opened their}$

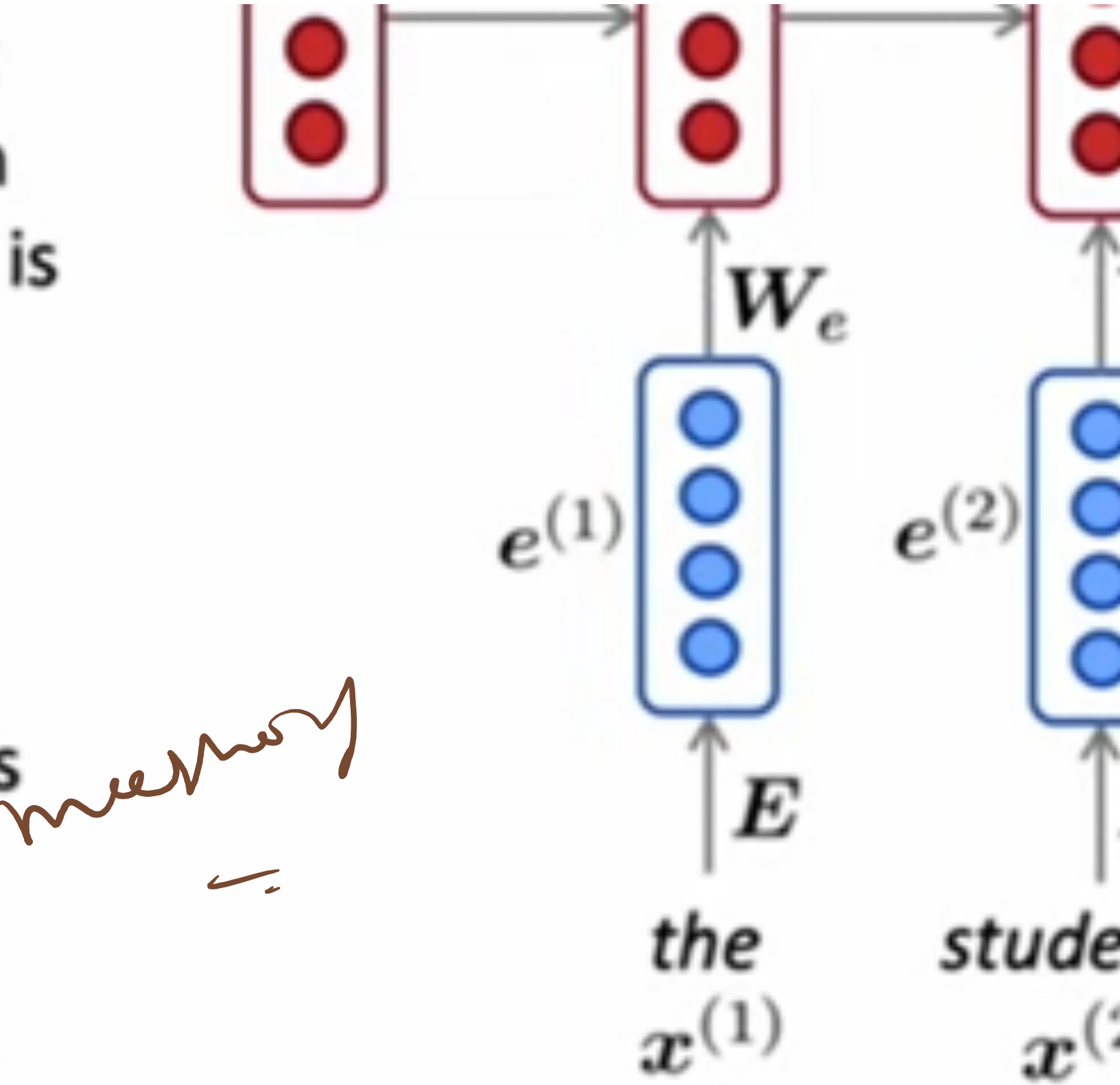


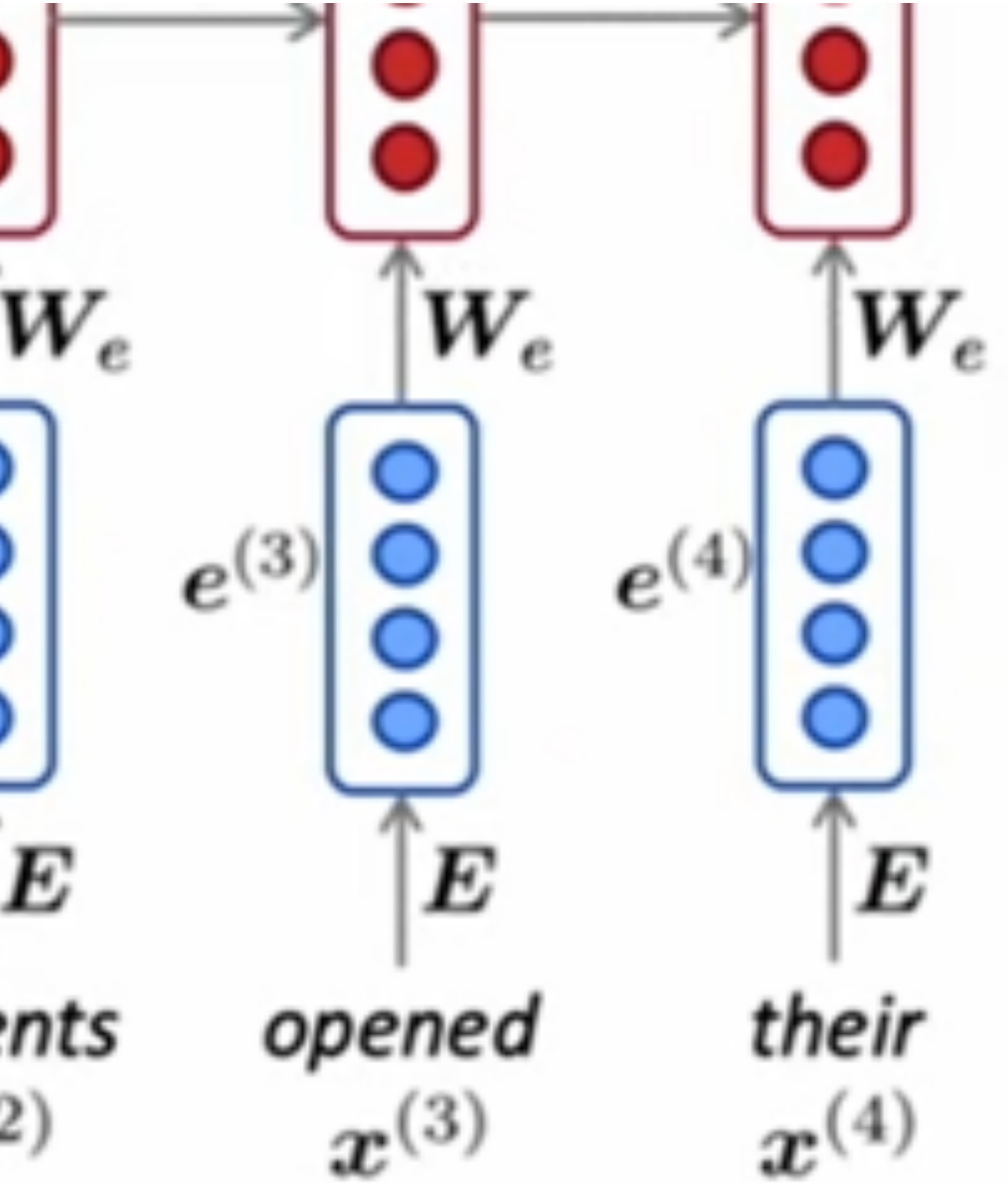
increase for longer input

- Same weights applied on every timestep, so there symmetry in how inputs are processed.

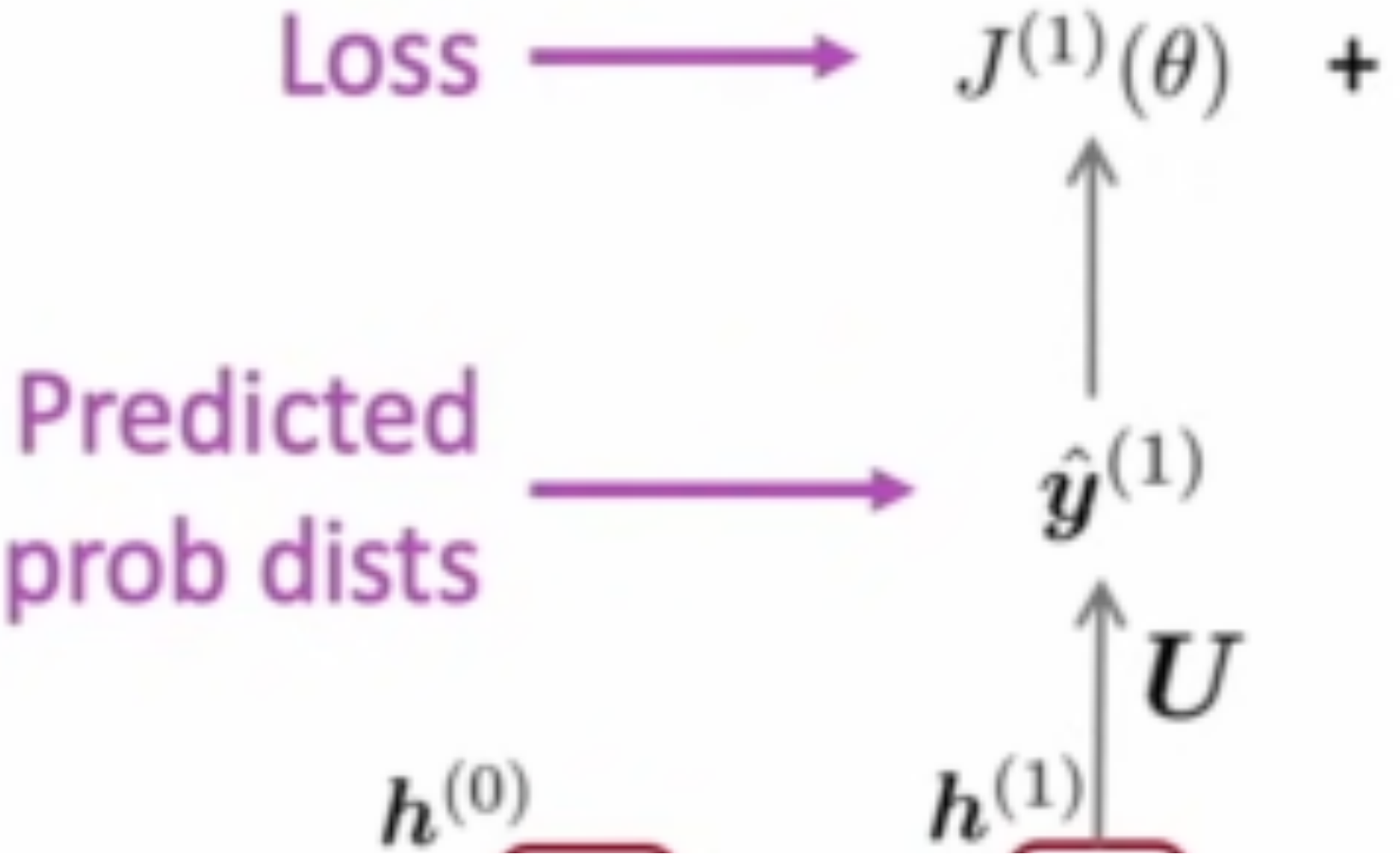
RNN **Disadvantages**:

- Recurrent computation is slow
- In practice, difficult to access information from many steps back

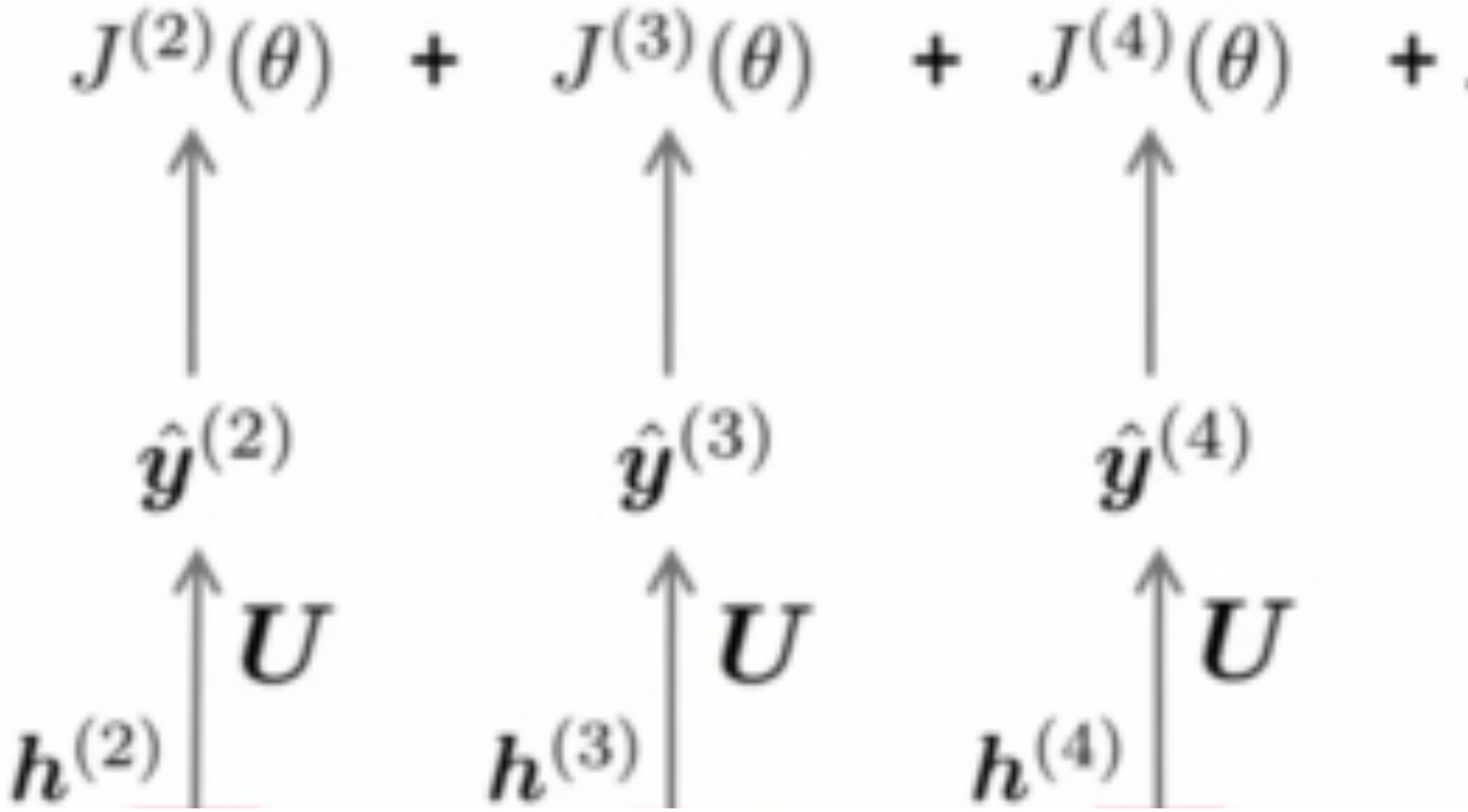

is
W_e
$e^{(1)}$
$e^{(2)}$
E
the
stude
$x^{(1)}$


W_e
W_e
W_e
$e^{(3)}$
$e^{(4)}$
E
E
E
ents
opened
their
$x^{(3)}$
$x^{(4)}$

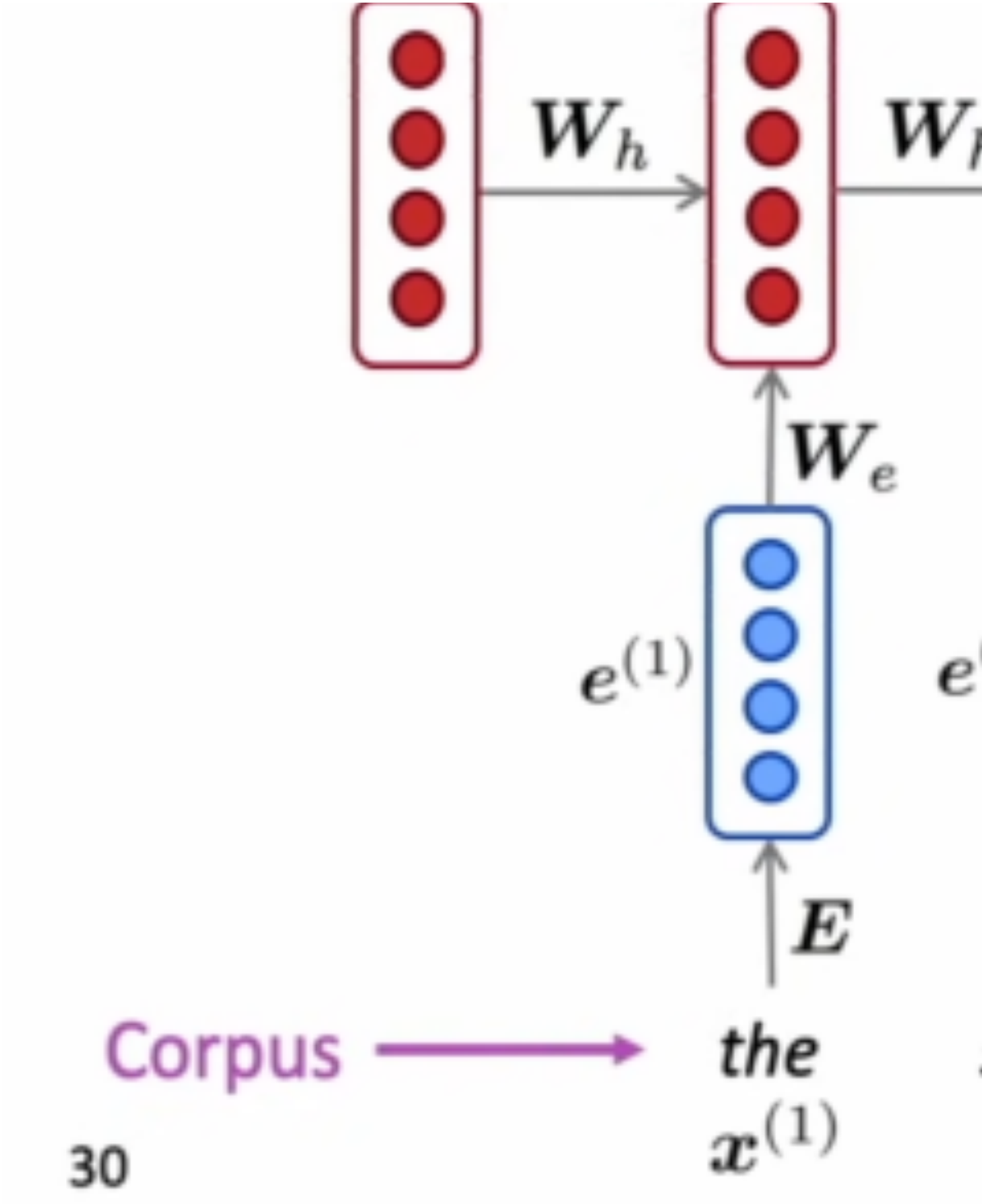
Training a RNN L

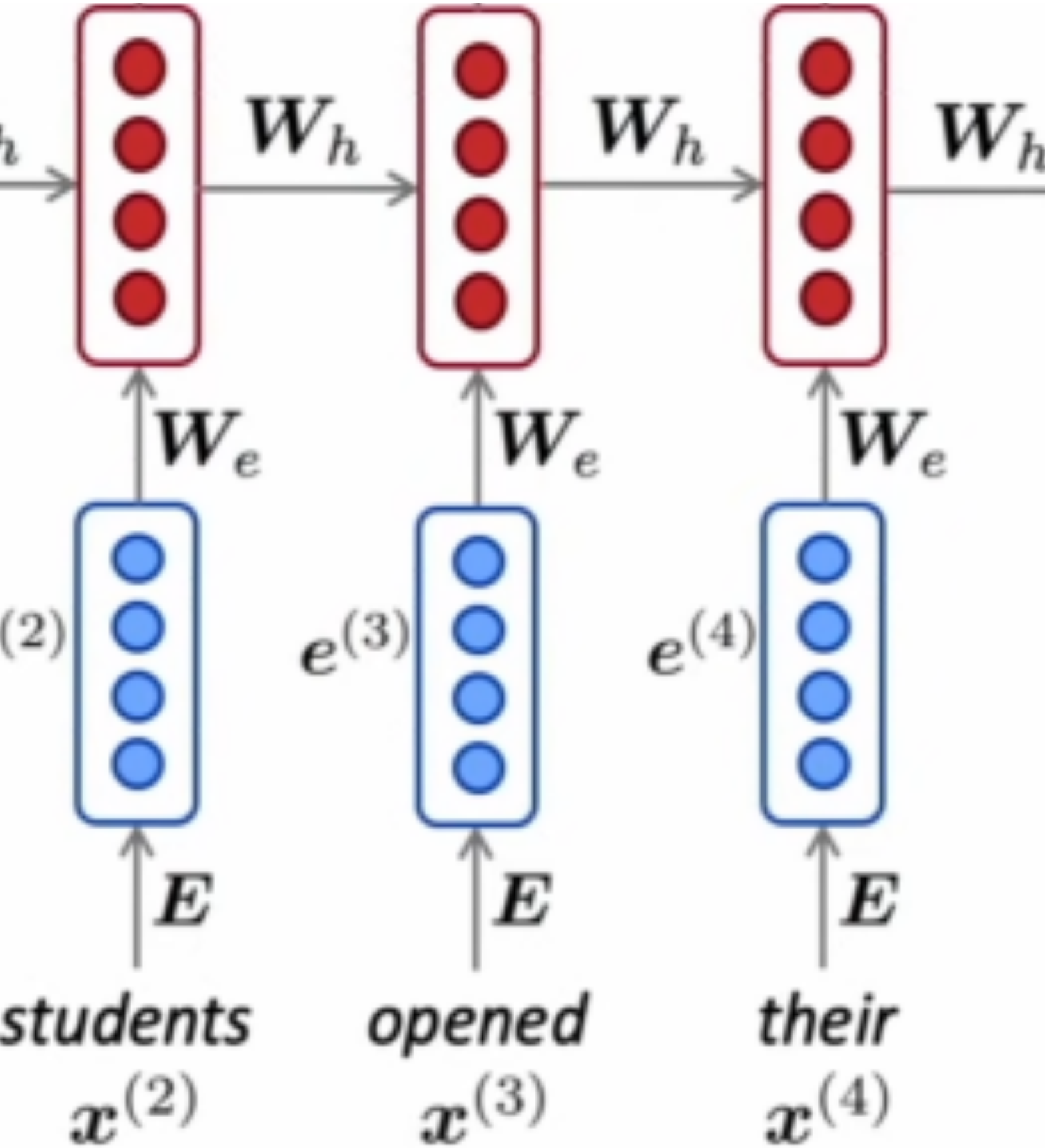


Language Model



... = $$J(\theta) = \frac{1}{T}\sum_{t=1}^{T} J^{(t)}(\theta)$$


W_h
W_e
$e^{(1)}$
E
Corpus
the
$x^{(1)}$

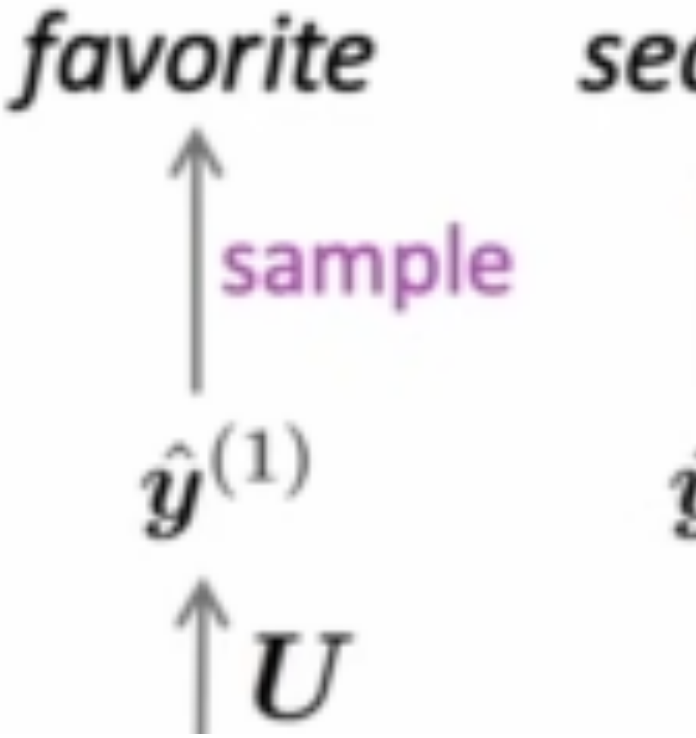

W_h
W_h
W_h
W_e
W_e
W_e
$e^{(2)}$
$e^{(3)}$
$e^{(4)}$
E
E
E
students
opened
their
$x^{(2)}$
$x^{(3)}$
$x^{(4)}$

exams ...

Generating text

Just like a n-gram Langua

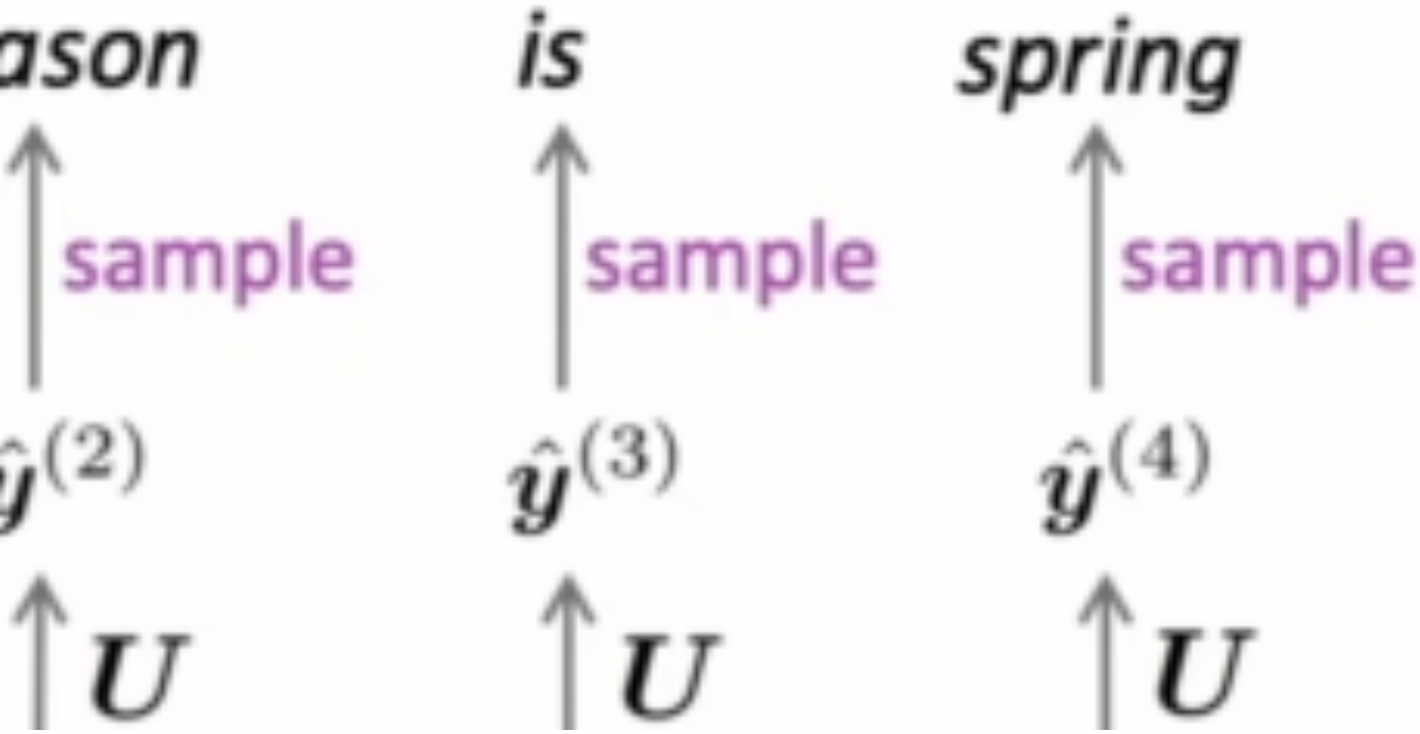
generate text by repeate



t with a RNN Langua

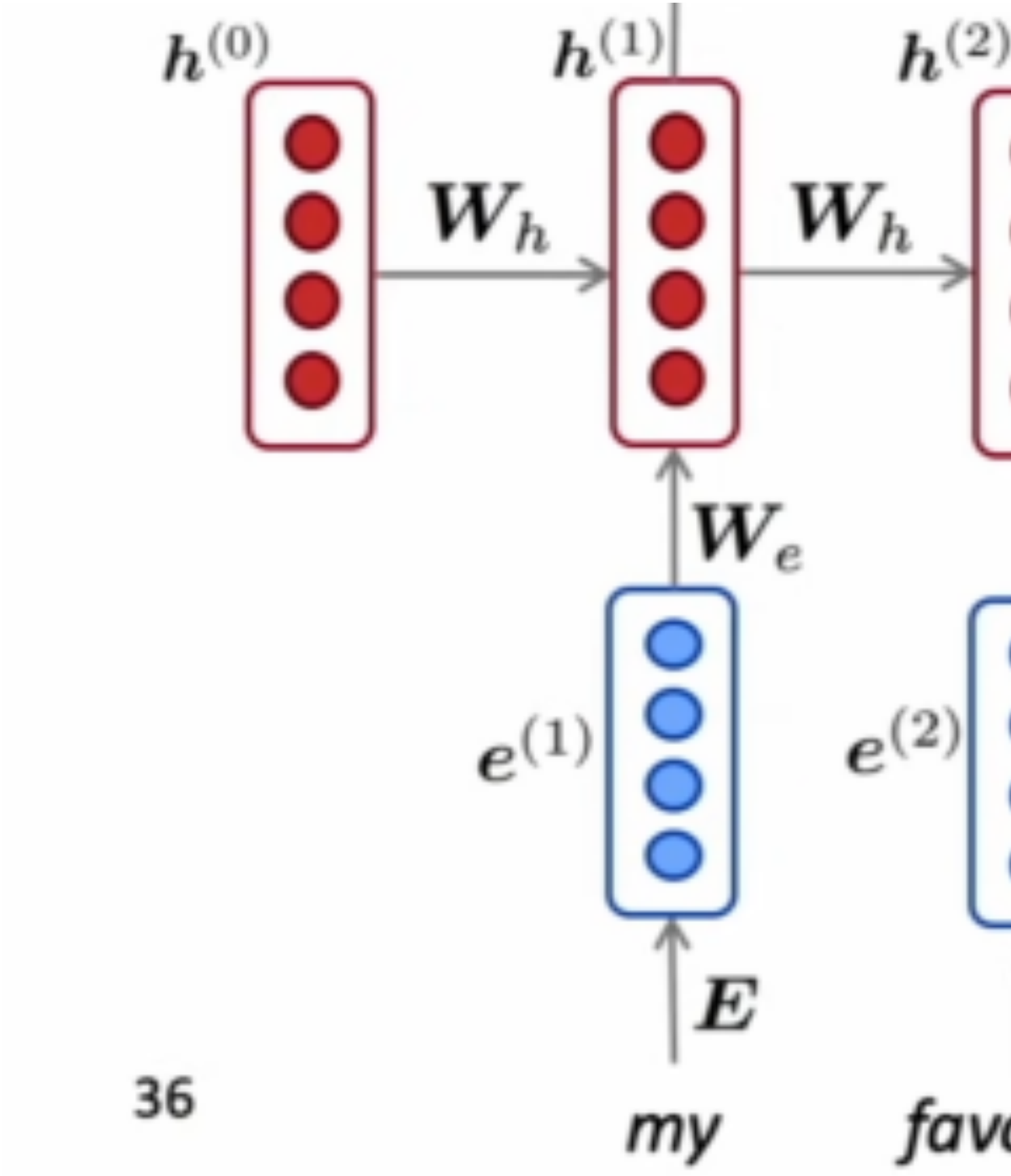
ge Model, you can use a RNN La

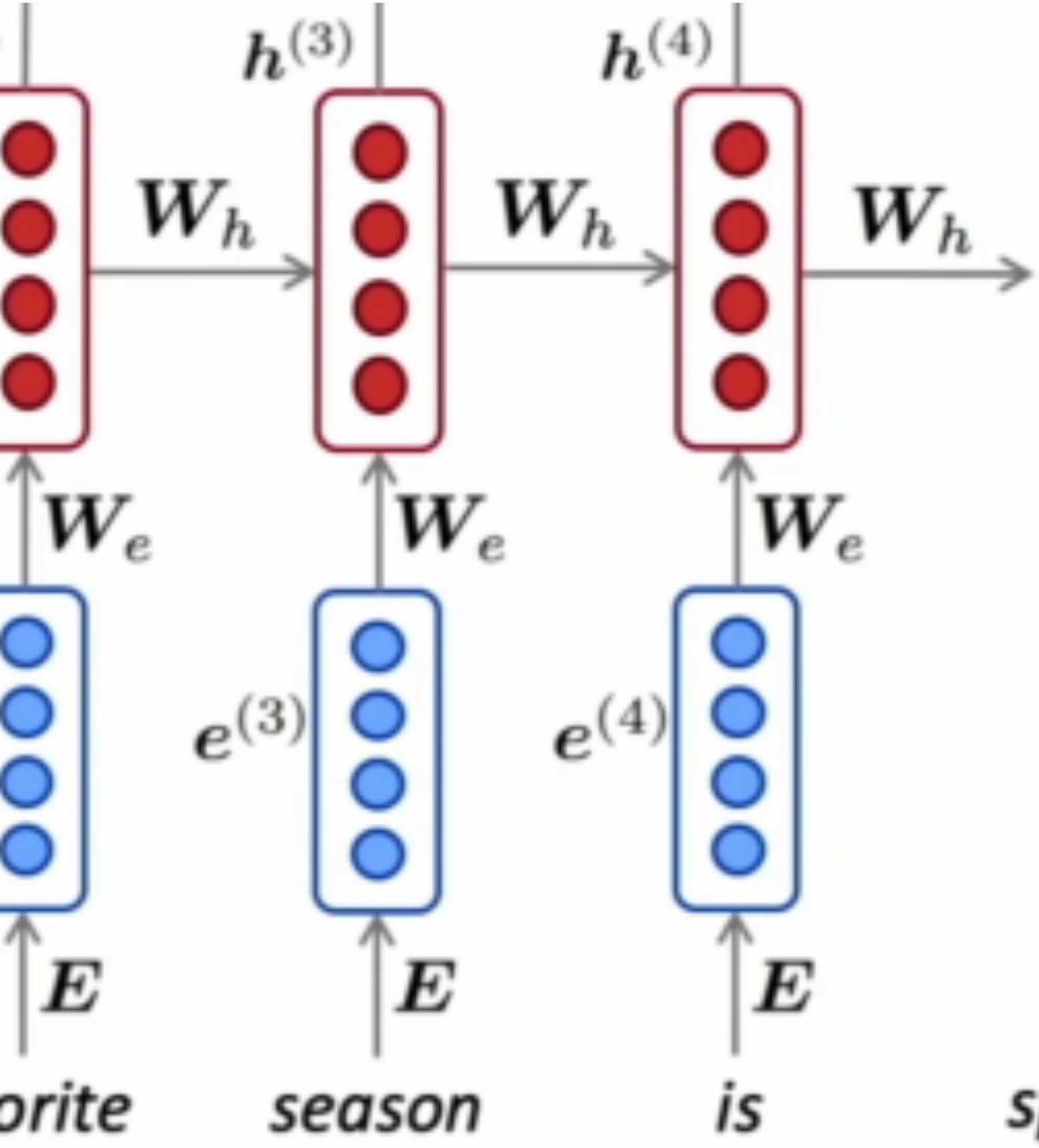
d sampling. Sampled output is n



age Model

anguage Model to
next step's input.


$h^{(0)}$
$h^{(1)}$
$h^{(2)}$
W_h
W_h
W_e
$e^{(1)}$
$e^{(2)}$
E
my
fav


$h^{(3)}$
$h^{(4)}$
W_h
W_h
W_h
W_e
W_e
W_e
$e^{(3)}$
$e^{(4)}$
E
E
E
orite
season
is
sp

...

ring

Generating text

- Let's have some fu
- You can train a RN
  in that style.
- RNN-LM trained o

"Sorry," Harry shouted,

with a RNN Langua

n!

N-LM on any kind of text, t

n *Harry Potter*:

, panicking—"I'll leave those broom

ge Model

hen generate text

ns in London, are

they?”

“No idea,” said Nearly H
last bit of treacle Char
common room perche
spider hadn’t felt it see

Source: ht

Generating text

Headless Nick, casting low close by
ms, from Harry's shoulder, and to a
d upon it, four arms held a shining
emed. He reached the teams too.

tps://medium.com/deep-writing/harry-potter-written

with a RNN Languag

Cedric, carrying the
nswer him the
knob from when the

-by-artificial-intelligence-8a9431803da6

ge Model

- Let's have some fun
- You can train a RNN
  in that style.
- RNN-LM trained on

```
      Title: CHOCOLATE RANCH
Categories: Game, Casseroles
      Yield: 6 Servings

      2 tb Parmesan cheese --
      1 c  Coconut milk
      3    Eggs, beaten
```

n!

N-LM on any kind of text, th

recipes:

```
BARBECUE
, Cookies, Cookies
```

```
chopped
```

hen generate text

```
3    Eggs, beaten

Place each pasta over layer
until firm. Serve hot in bo

Combine the cheese and salt
and stir in the chocolate a
```

39

Generating text w

rs of lumps. Shape mixture into the mo
odied fresh, mustard, orange and chees

t together the dough in a large skille
nd pepper.

Source: https://gist.gith

with a RNN Languag

```
derate oven and simmer
e.

t; add the ingredients
```

ub.com/nylki/1efbaa36635956d35bcc

e Model

- Let's have some fun
- You can train a RNN
  in that style.
- RNN-LM trained on

!

-LM on any kind of text, the

paint color names:

31 137 165

51 124 112

9 173 140

ite 221 215 236

178 181 196

wn 133 104 85

144 106 74

237 217 177

232 223 215

Sand Dan 201 1

Grade Bat 48 94

Light Of Blast 1

Grass Bat 176 9

Sindis Poop 204

Dope 219 209 1

Testing 156 101

Stoner Blue 152

Burble Simp 226

n generate text

72 143
83
75 150 147
9 108
205 194
79
106
165 159
6 181 132

Burf Pink 223

Rose Hork 23

This is an example of a char

40

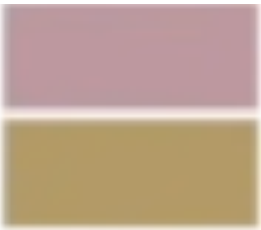
Source:

Why should we

- Language Modeling

173 179

0 215 198



Stanky Bean 197

Turdly 190 164

acter-level RNN-LM (predicts what ch

http://aiweirdness.com/post/160776374467/new-paint

care about Language

is a benchmark task that he

7 162 171

116

haracter comes next)

-colors-invented-by-neural-network

e Modeling?

elps us