Recitation 6: RNN Basics

Part 1, Miya
Part 2, Sam

RNNs process sequence data for predictions

Examples of sequences:
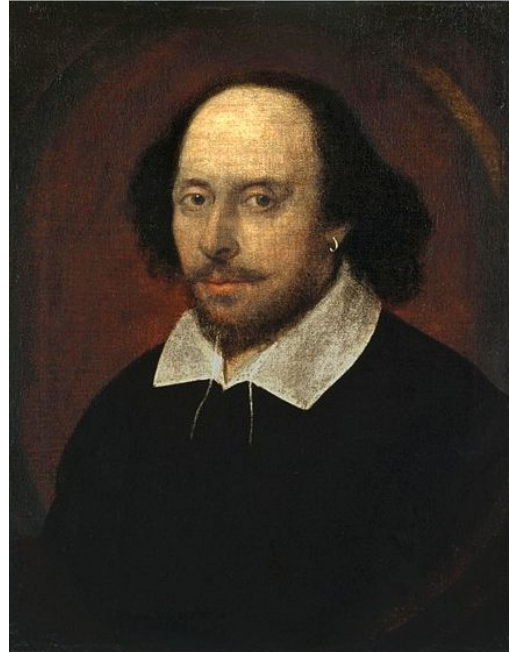
- Text
- Speech
- Physical Processes
- Video
- Audio
- …

A sequence of words

William Shakespeare

Article | Talk

Read ... nt Log in

To be, or not to be

From Wikipedia, the free encyclopedia

For other uses, see To Be or Not to Be (disambiguation).
"Perchance to Dream" redirects here. For other uses, see Percha...

"**To be, or not to be**" is the opening phrase of a soliloquy[a] uttered by Hamlet in the so-called "nunnery scene" of William Shakespeare's pla... Act 3, Scene 1. In the speech, Hamlet contemplates death and suicid... bemoaning the pain and unfairness of life but acknowledging that the... might be worse. The opening line is one of the most widely known an... lines in modern English, and the soliloquy has been referenced in innumerable works of theatre, literature and music.

Contents [hide]

A sequence of words

To be, or not to be: that is the question

William Shakespeare

to be or not to be : CNNs vs RNNs

- CNNs compute vectors for all possible phrases
  - "to be", "be or", "or not", "not to", "to be", "be :", "to be or", "be or not", "or not to", "not to be", "to be :", "to be or not", "be or not to", "or not to be", "not to be :", "to be or not to", "be or not to be", "or not to be :", "to be or not to be", "be or not to be : ", "to be or not to be : "

- RNNs process inputs sequentially
  - "to", "be", "or", "not", "to", "be", " : ", "that", "is", "the", "question"

Recurrent Neural Networks (RNNs)

Dictionary

0. Apple
1. Be
2. Example
3. Not
4. Or
5. That
6. The
7. To
8. Question
9. :

Example Phrase

Original String:

"To Be Or Not To Be:

That Is The Question"

Tokenized String:

["To", "Be", "Or",

"Not", "To", "Be", ":",

" That", "Is", "The",

"Question"]

TOKENIZED VERSION | To | be | or | not | to | · · ·
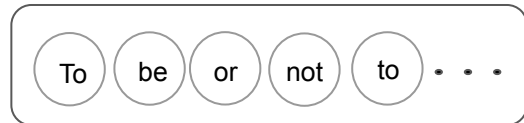
CLASSES VERSION | 7 | 1 | 4 | 3 | 7 | · · ·
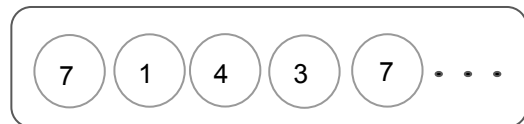
ONE-HOT ENCODING /
EMBEDDING

0	0	0	0	0
0	1	0	0	0
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0
1	0	0	0	1
0	0	0	0	0
0	0	0	0	0

Recurrent Neural Networks (RNNs)

t=0

frequencies
in frame 0

0
0
0
0
0
0
0
1
0
0

Recurrent Neural Networks (RNNs)

t=0

frequencies
in frame 0

0.03
0.01
0.04
0.15
0.09
0.31
0.04
0.87
0.01
0.26

Recurrent Neural Networks (RNNs)

t=0

frequencies
in frame 0

0.03
0.01
0.04
0.15
0.09
0.31
0.04
0.87
0.01
0.26

7

Recurrent Neural Networks (RNNs)

t=0

t=1

frequencies in frame 0

frequencies in frame 1

0
0
0
0
0
0
0
0
1
0
0

0
1
0
0
0
0
0
0
0
0
0
0
0

7

1
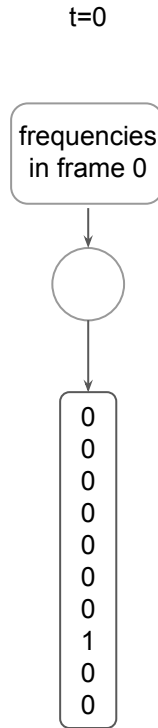
Recurrent Neural Networks (RNNs)

Recurrent Neural Network (RNNs)

NEW INFORMATION

PREDICTION

tanh
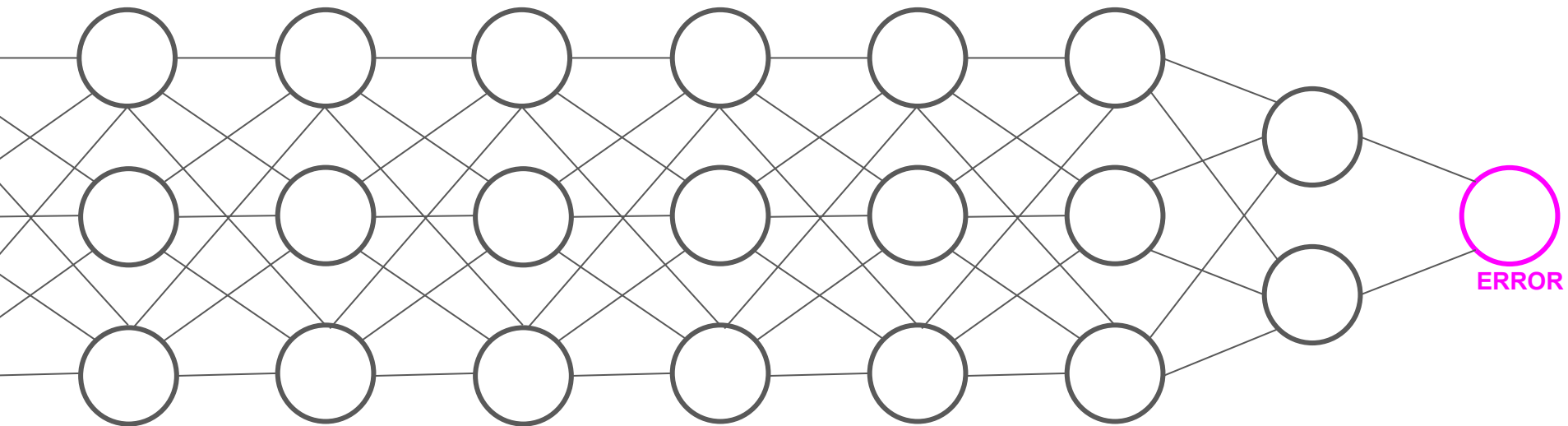$\tanh(x)$
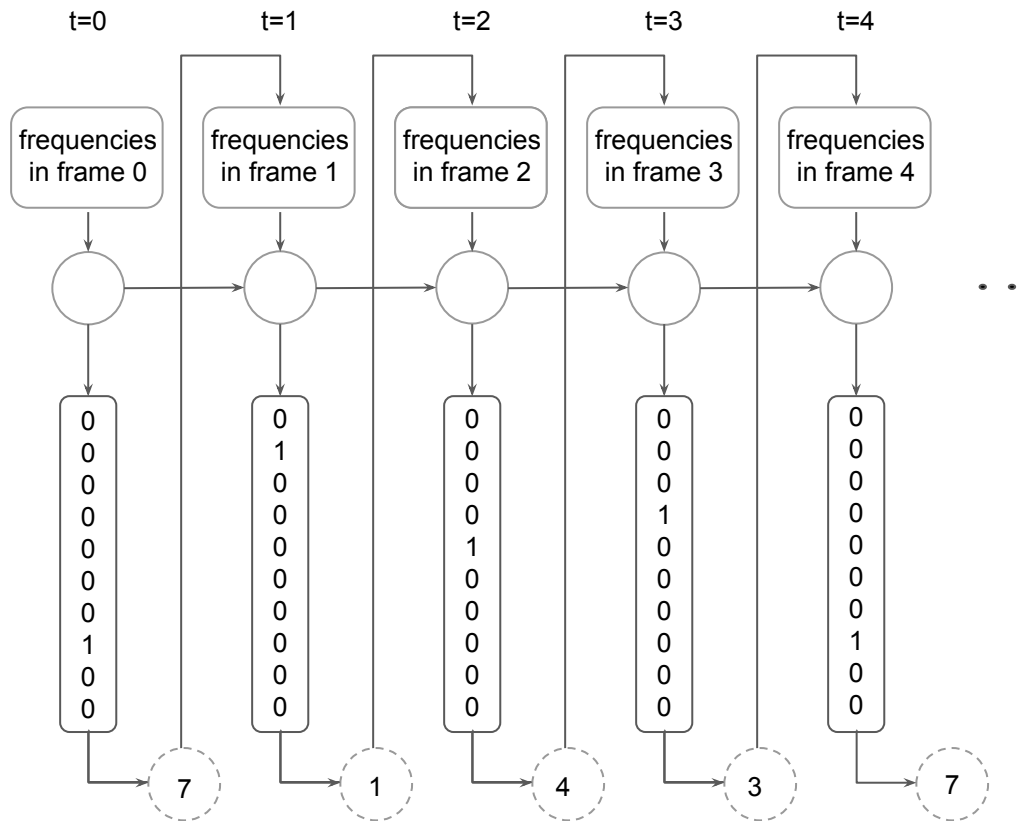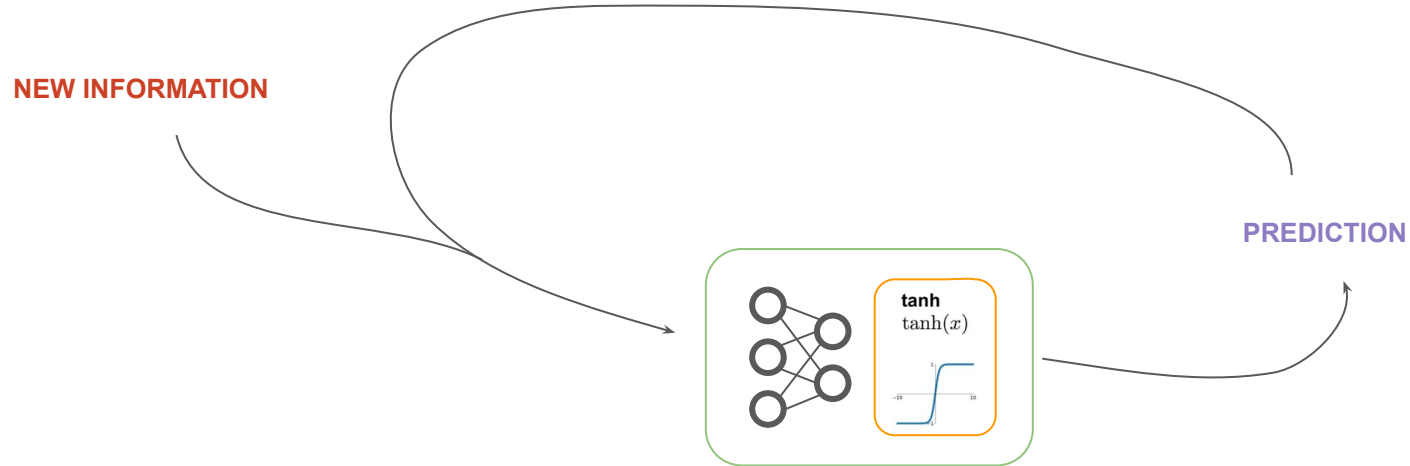
Short-term Memory… vanishing gradient
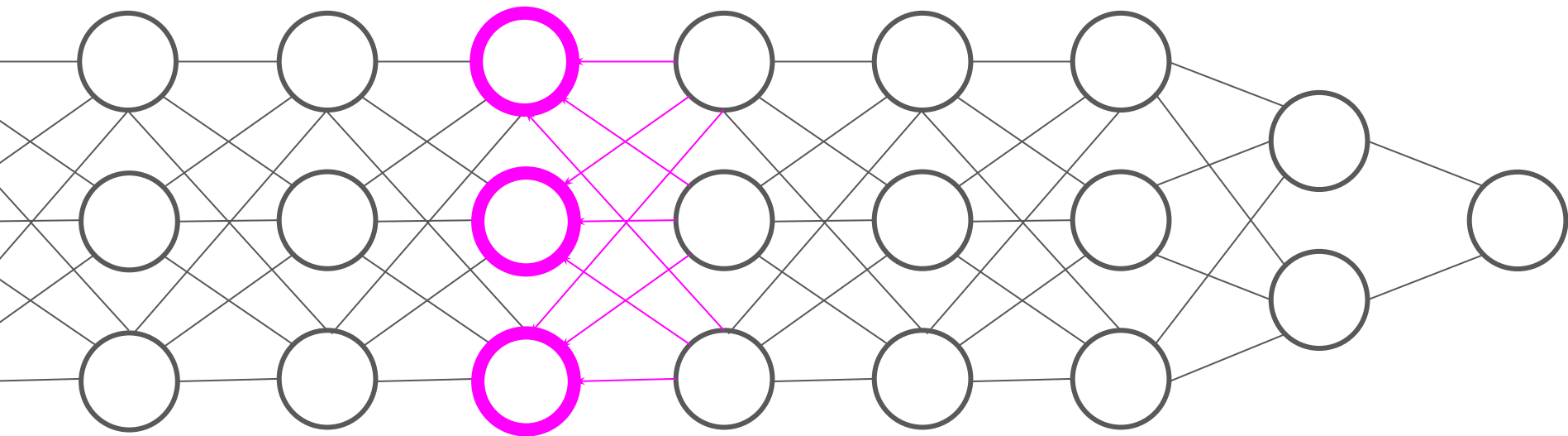
to be or not to be :
to be :
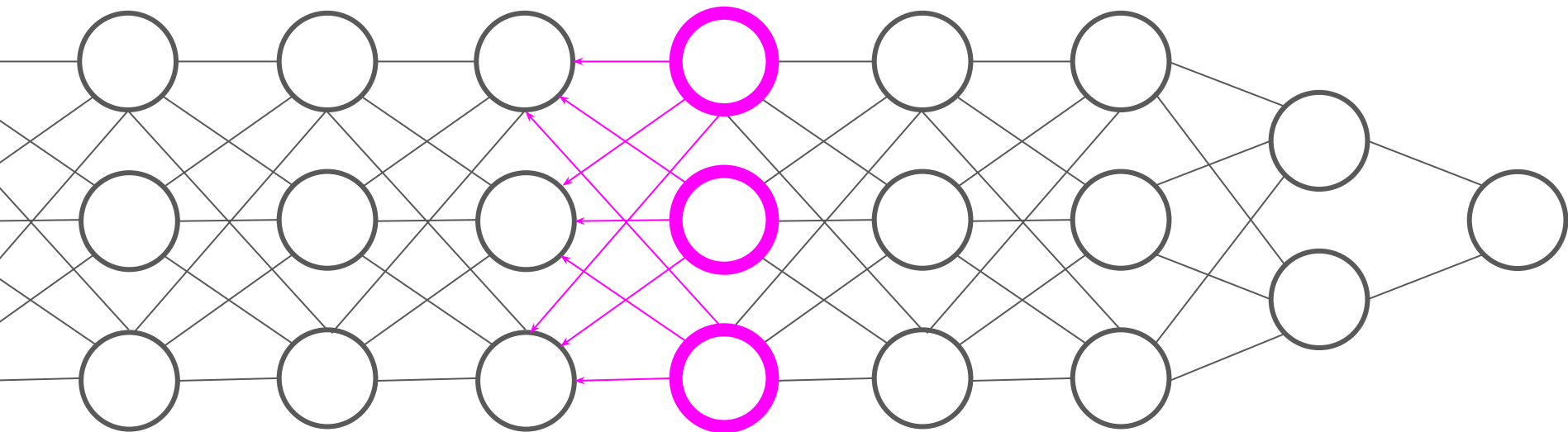to be or not to be or not to be or not to be or not …

Vanishing Gradient

ERROR
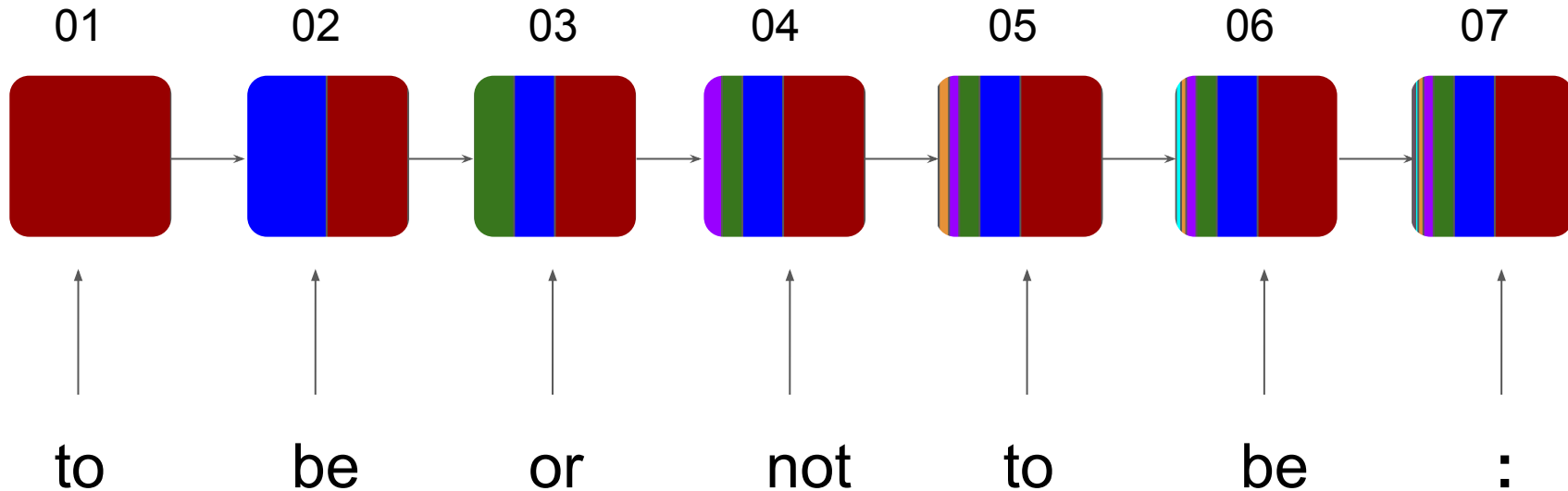
Vanishing Gradient

Vanishing Gradient

Reducing Short-term Memory

- Long Short-term Memory (LSTM)

- Gated Recurrent Unit (GNU)
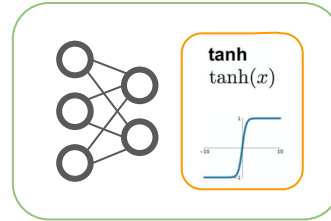
Recurrent Neural Network (RNNs)
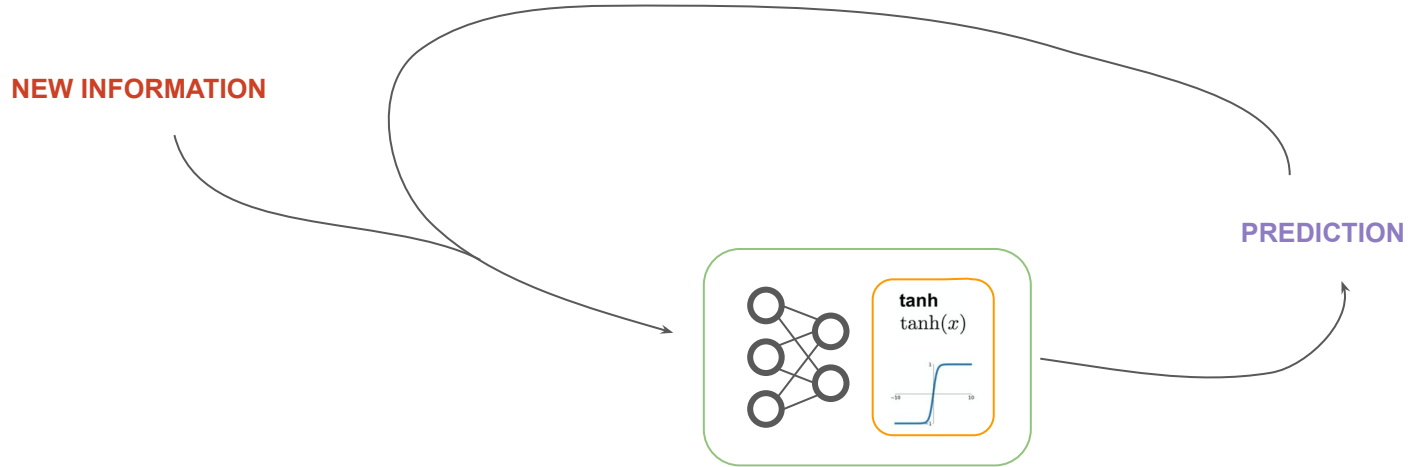
to be or not to be **:**
to be :
to be or not to be or not to be or not to be or not …
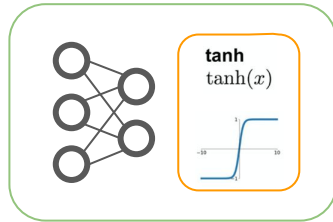
NEW INFORMATION

PREDICTION

tanh
$\tanh(x)$

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

NEW INFORMATION
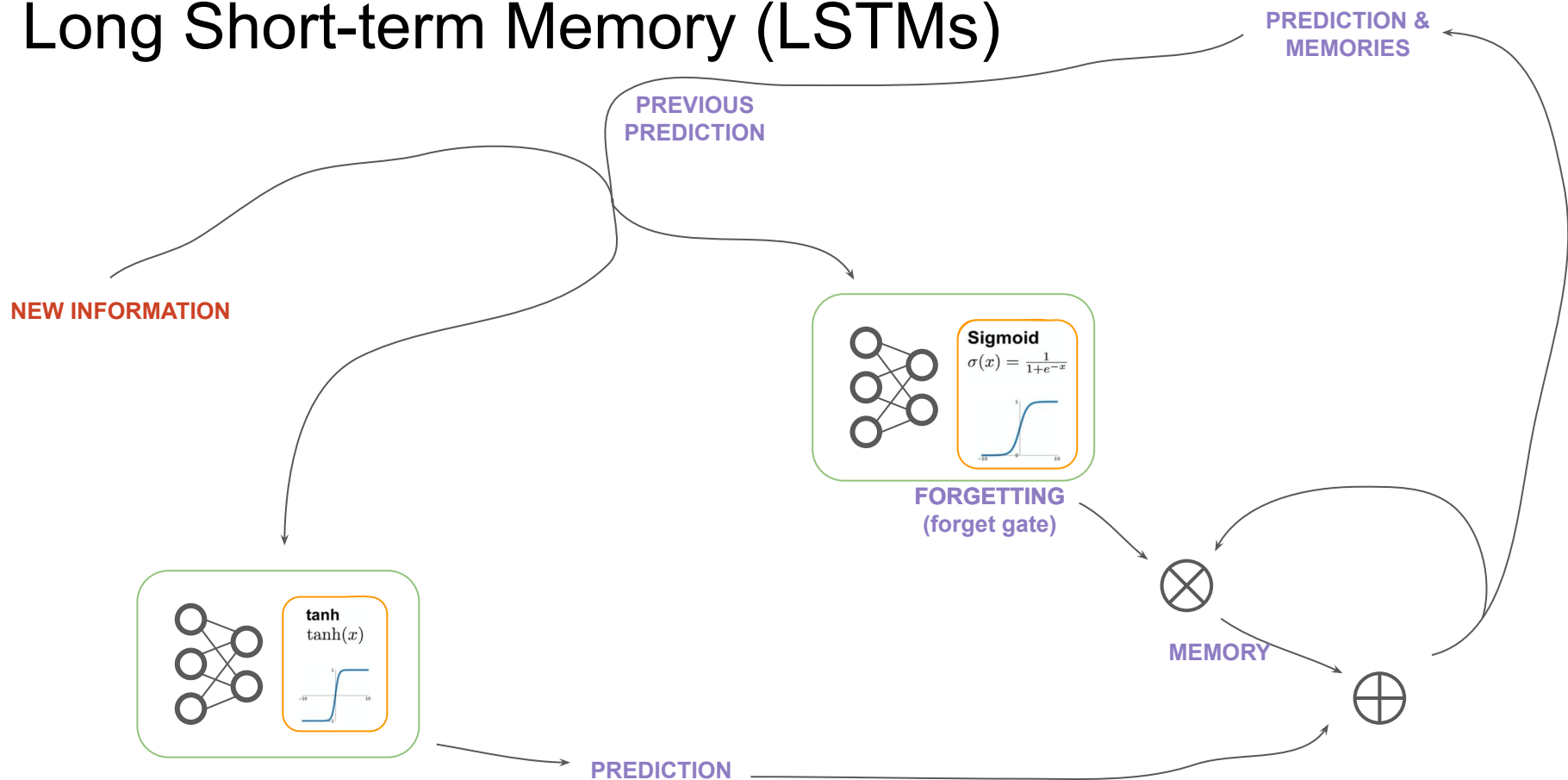
tanh
$\tanh(x)$

PREDICTION

Long Short-term Memory (LSTMs)

PREDICTION & MEMORIES

PREVIOUS PREDICTION

NEW INFORMATION

Sigmoid

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

FORGETTING (forget gate)

tanh

$\tanh(x)$

MEMORY

PREDICTION
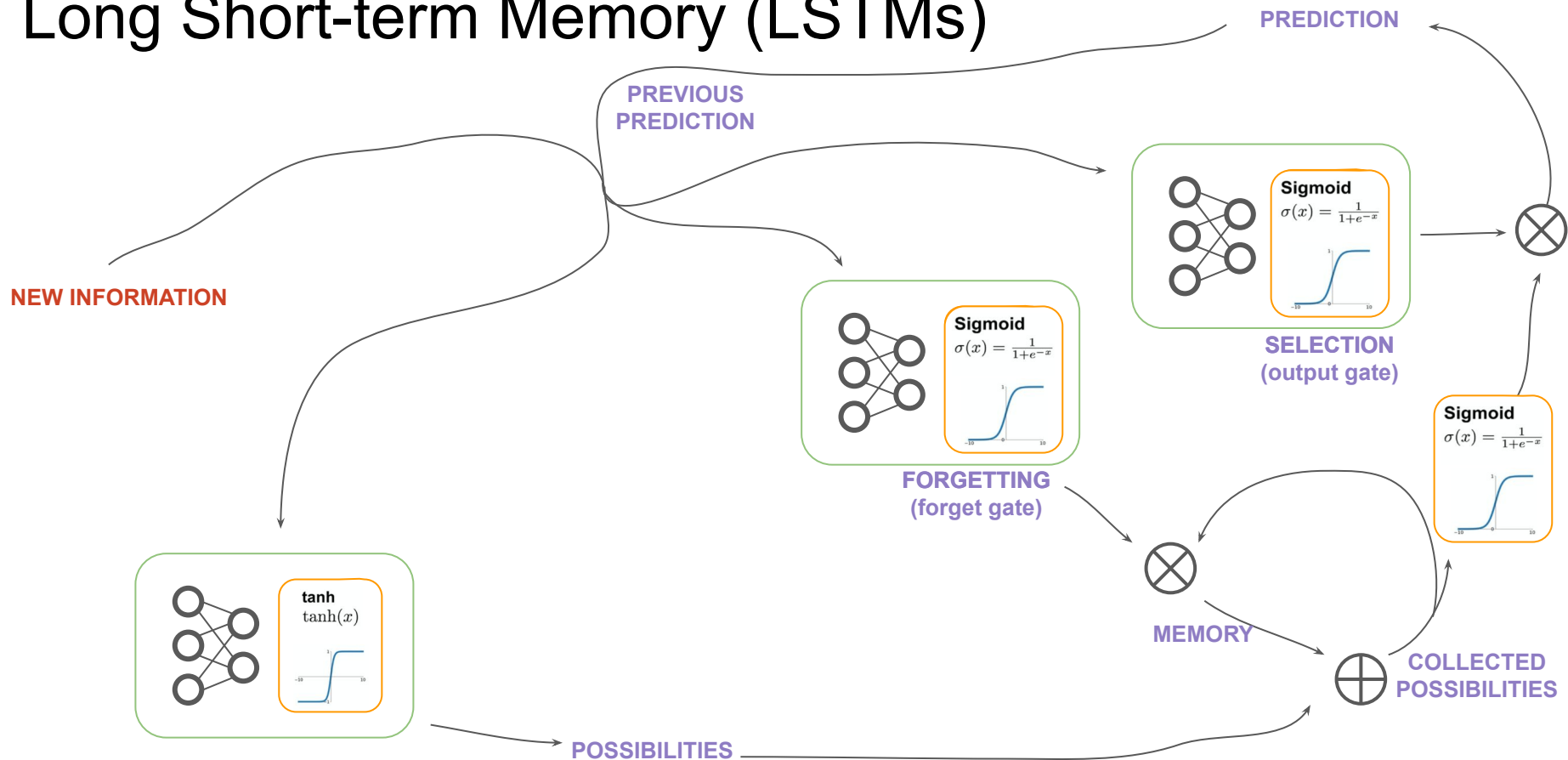
Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

PREDICTION

PREVIOUS PREDICTION

TO

SELECTION (output gate)

FORGETTING (forget gate)

IGNORING (input gate)

MEMORY

COLLECTED POSSIBILITIES

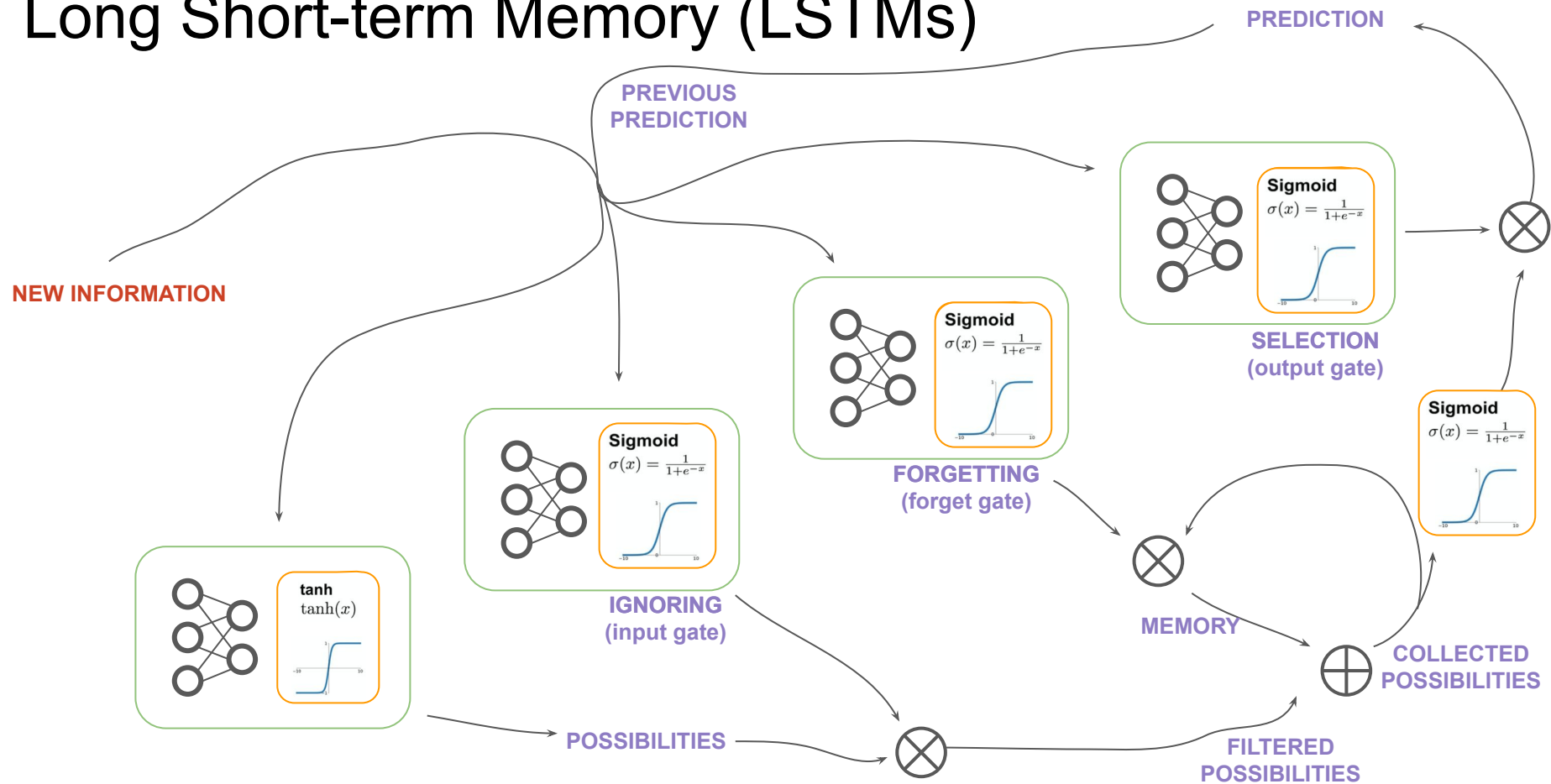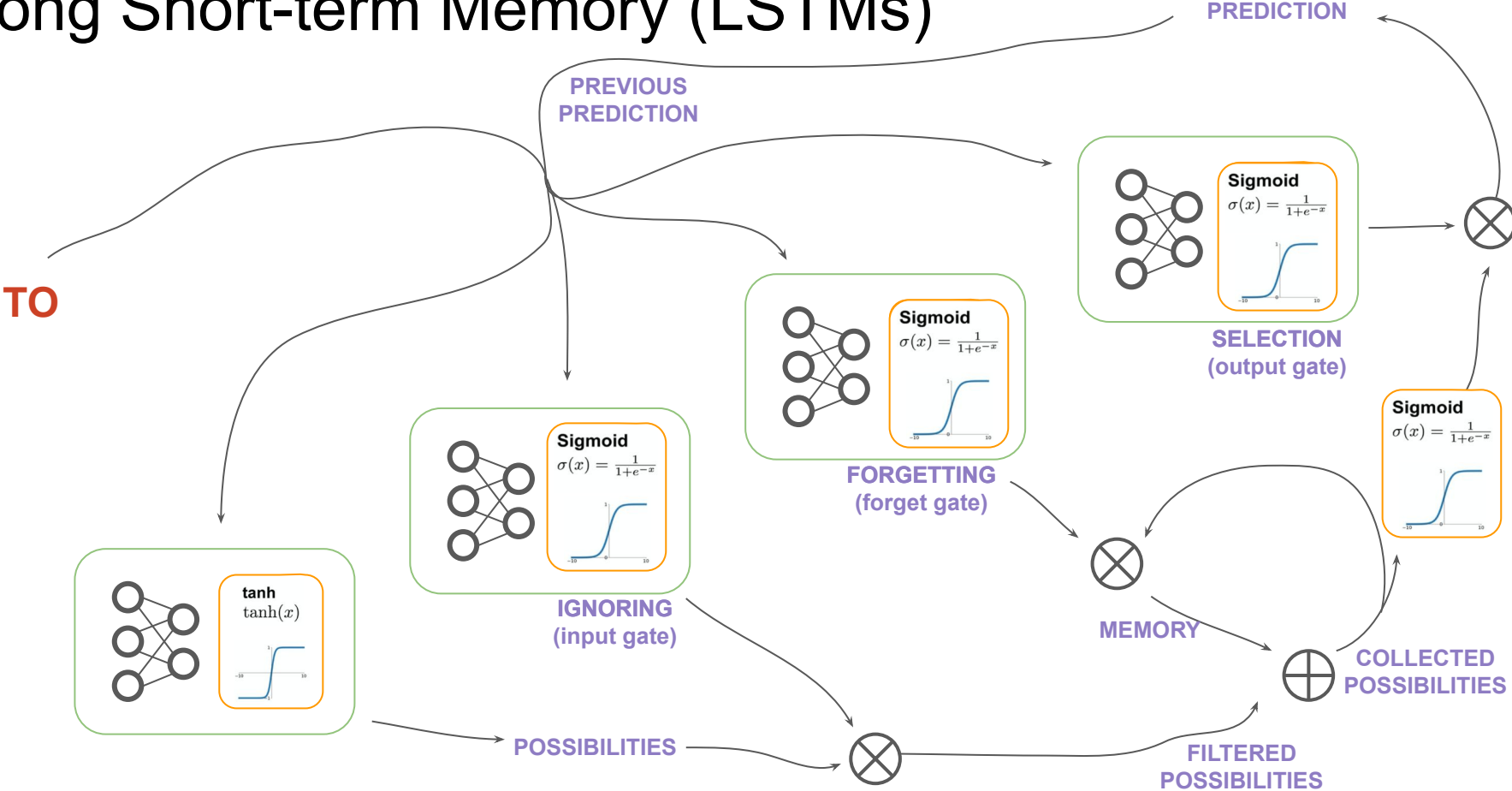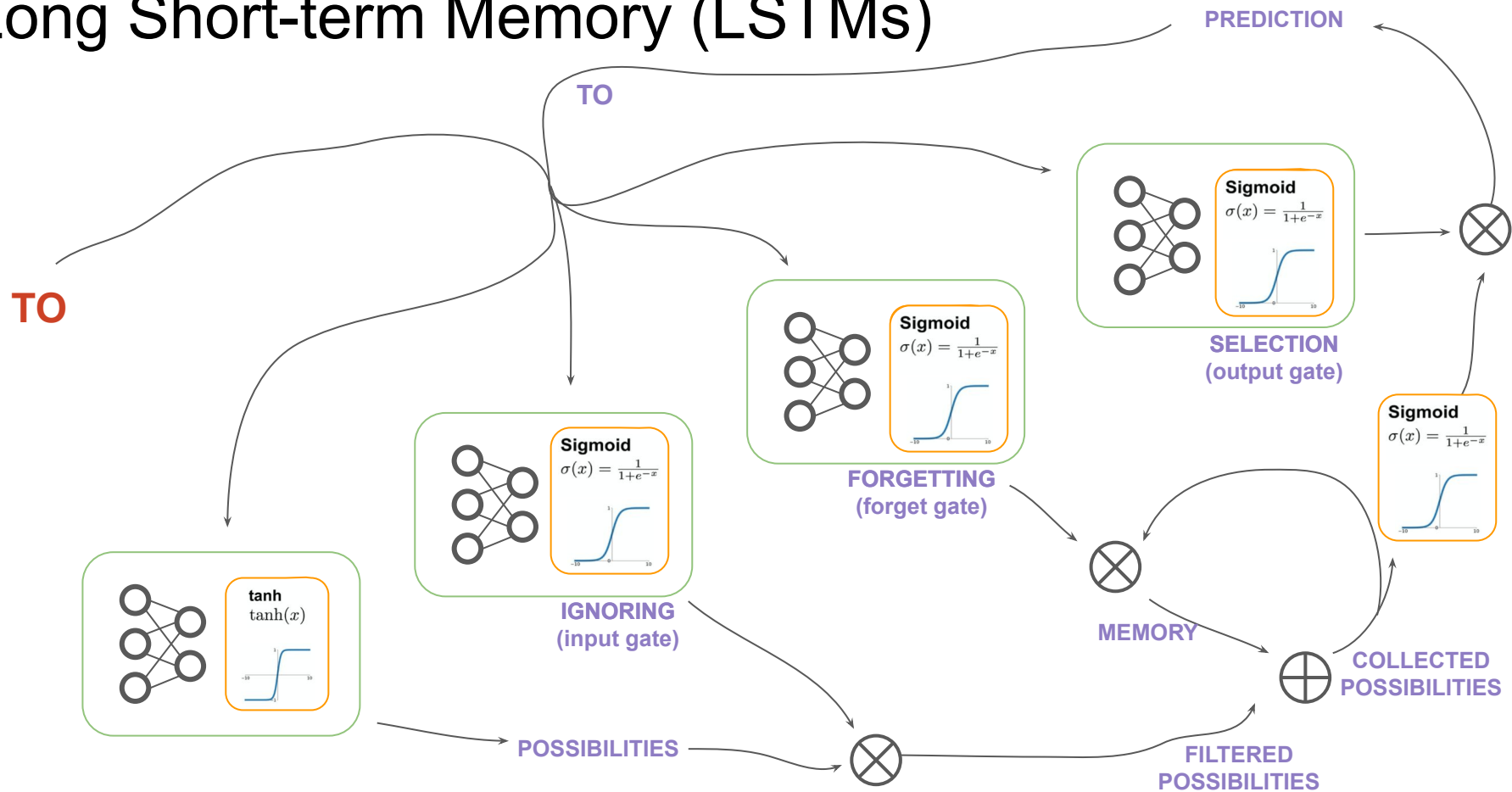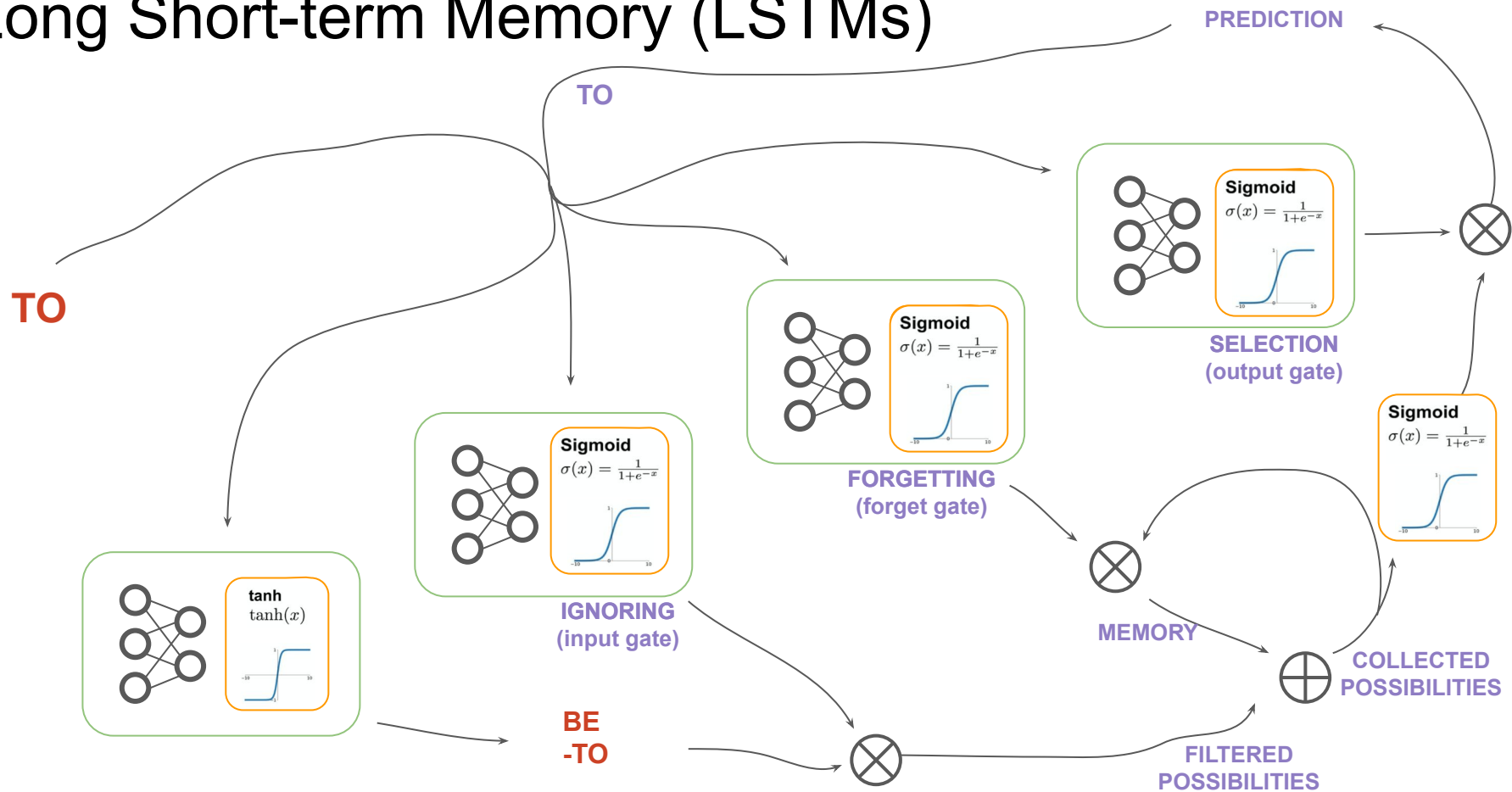POSSIBILITIES

FILTERED POSSIBILITIES

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

a

Long Short-term Memory (LSTMs)

PREDICTION

TO

TO

SELECTION
(output gate)

FORGETTING
(forget gate)

THAT
OR
APPLE ...

IGNORING
(input gate)

MEMORY

COLLECTED
POSSIBILITIES

BE
-TO

BE
-TO

Long Short-term Memory (LSTMs)

PREDICTION

TO

TO

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

SELECTION
(output gate)

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

FORGETTING
(forget gate)

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

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

THAT
OR
APPLE ...

BE
-TO

MEMORY

IGNORING
(input gate)

tanh
$\tanh(x)$

BE
-TO

BE
-TO

BE
-TO

BE
-TO

Long Short-term Memory (LSTMs)

PREDICTION

TO

TO

TO

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

BE > -TO

SELECTION
(output gate)

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

FORGETTING
(forget gate)

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

**BE
-TO**

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

**THAT
OR
APPLE ...**

**BE
-TO**

IGNORING
(input gate)

MEMORY

**BE
-TO**

tanh
$\tanh(x)$

**BE
-TO**
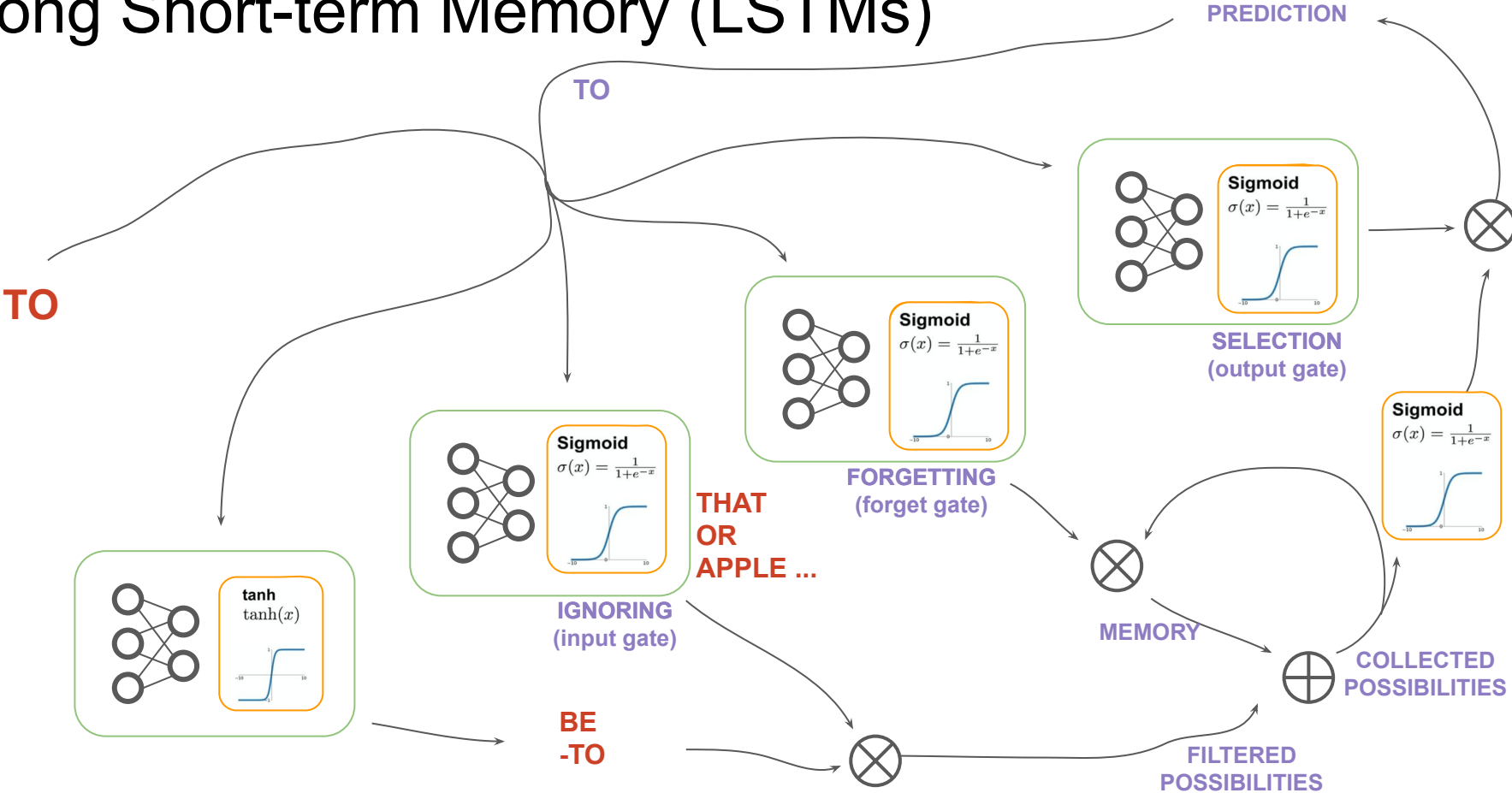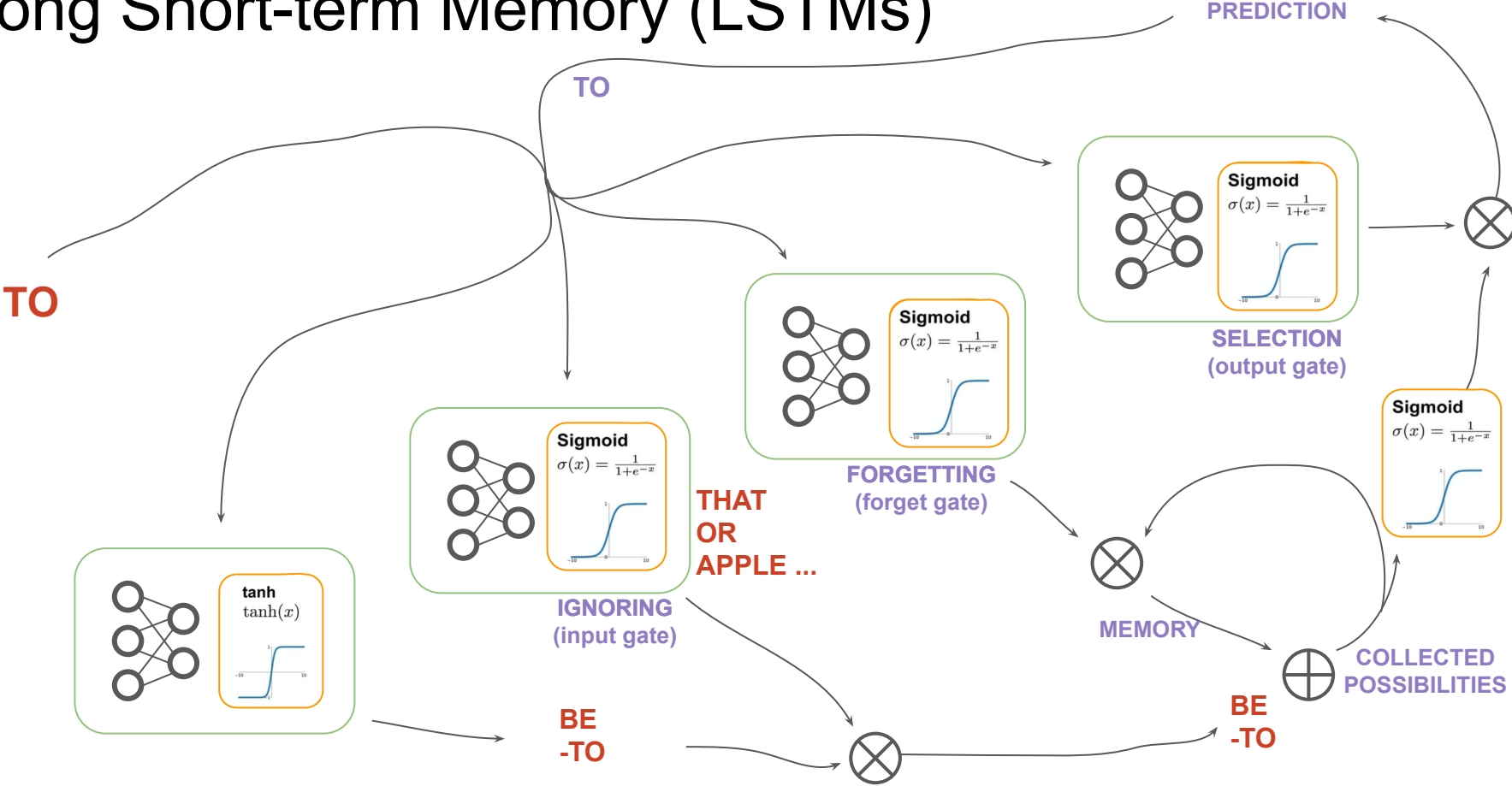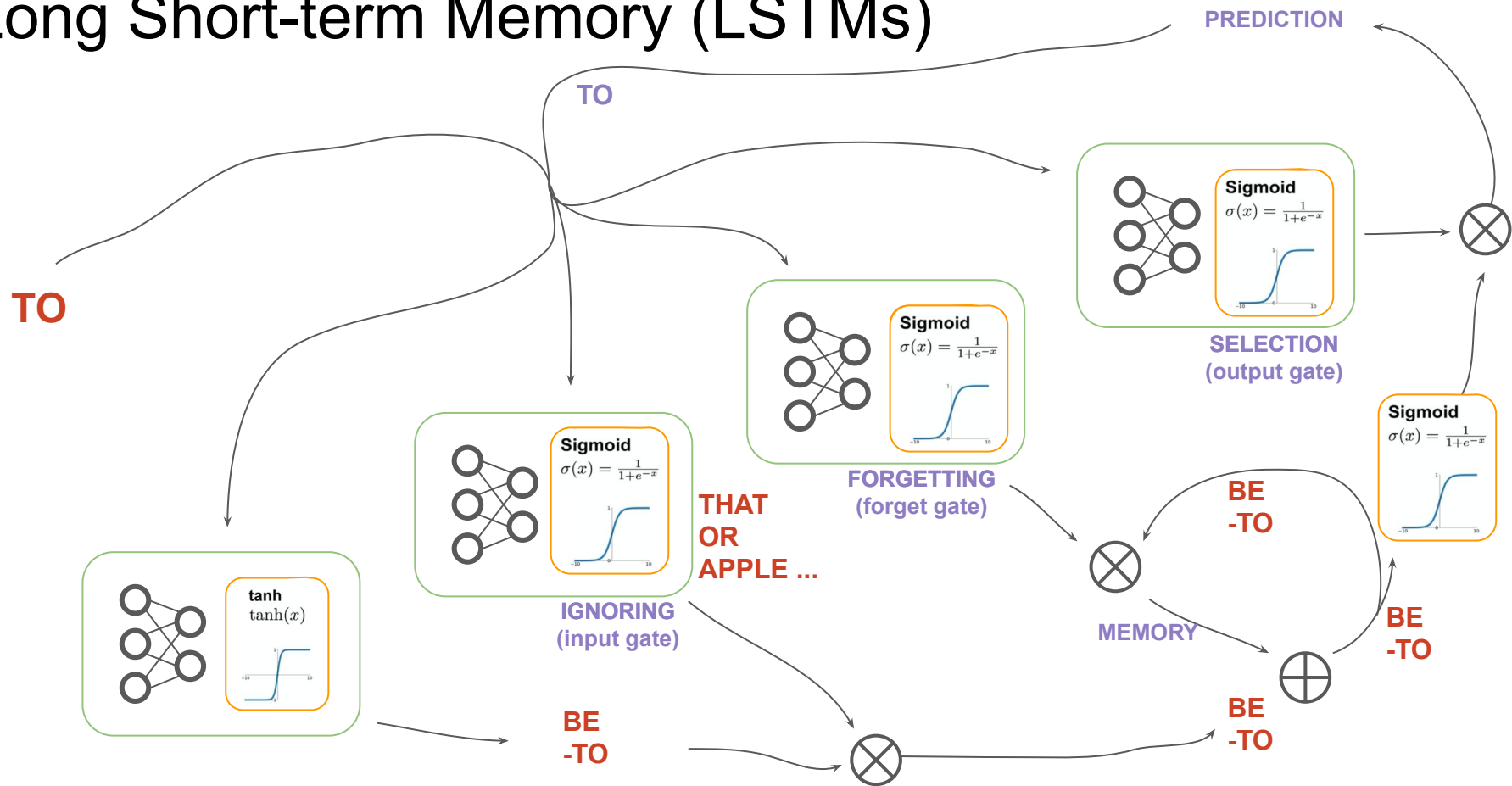
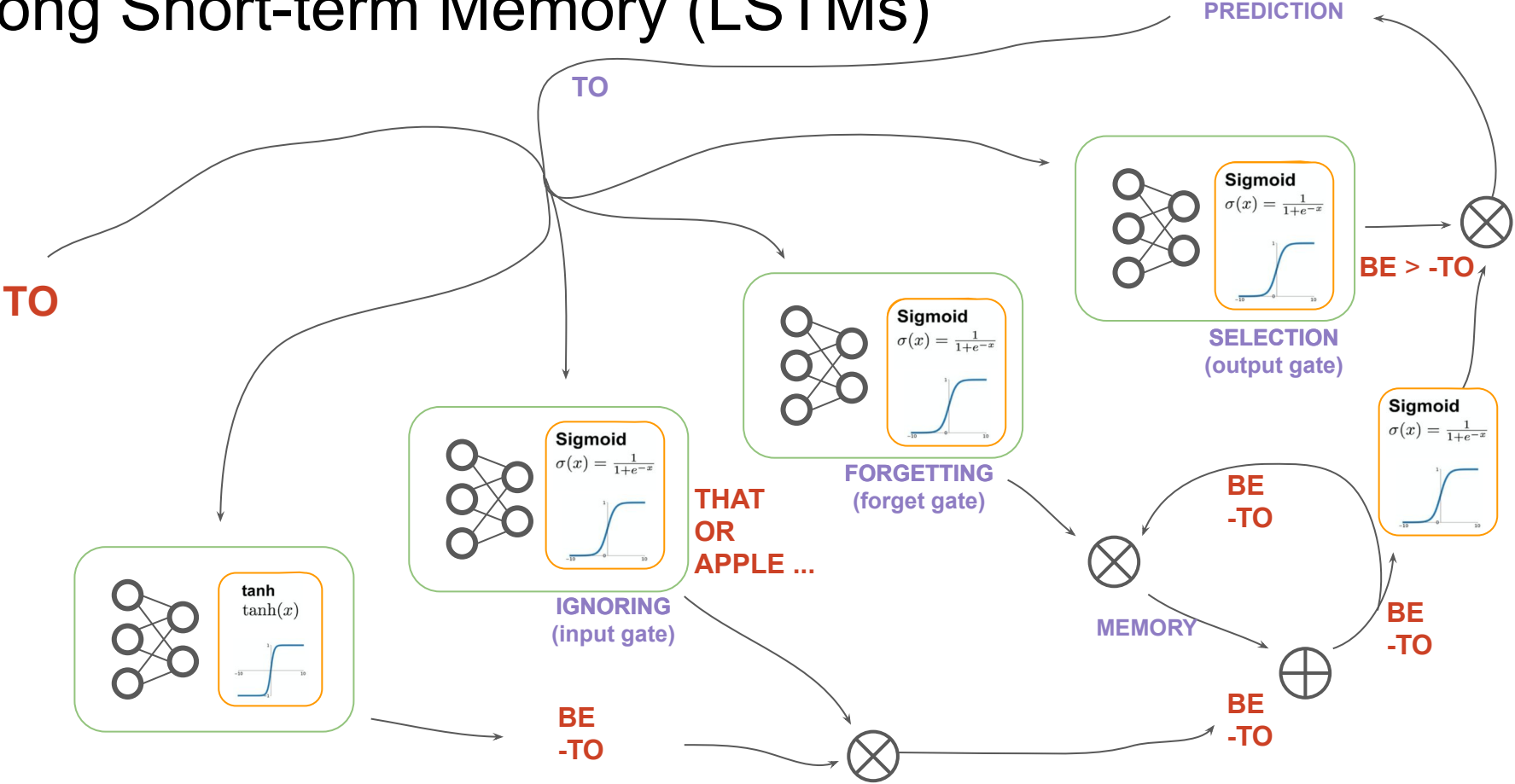# Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

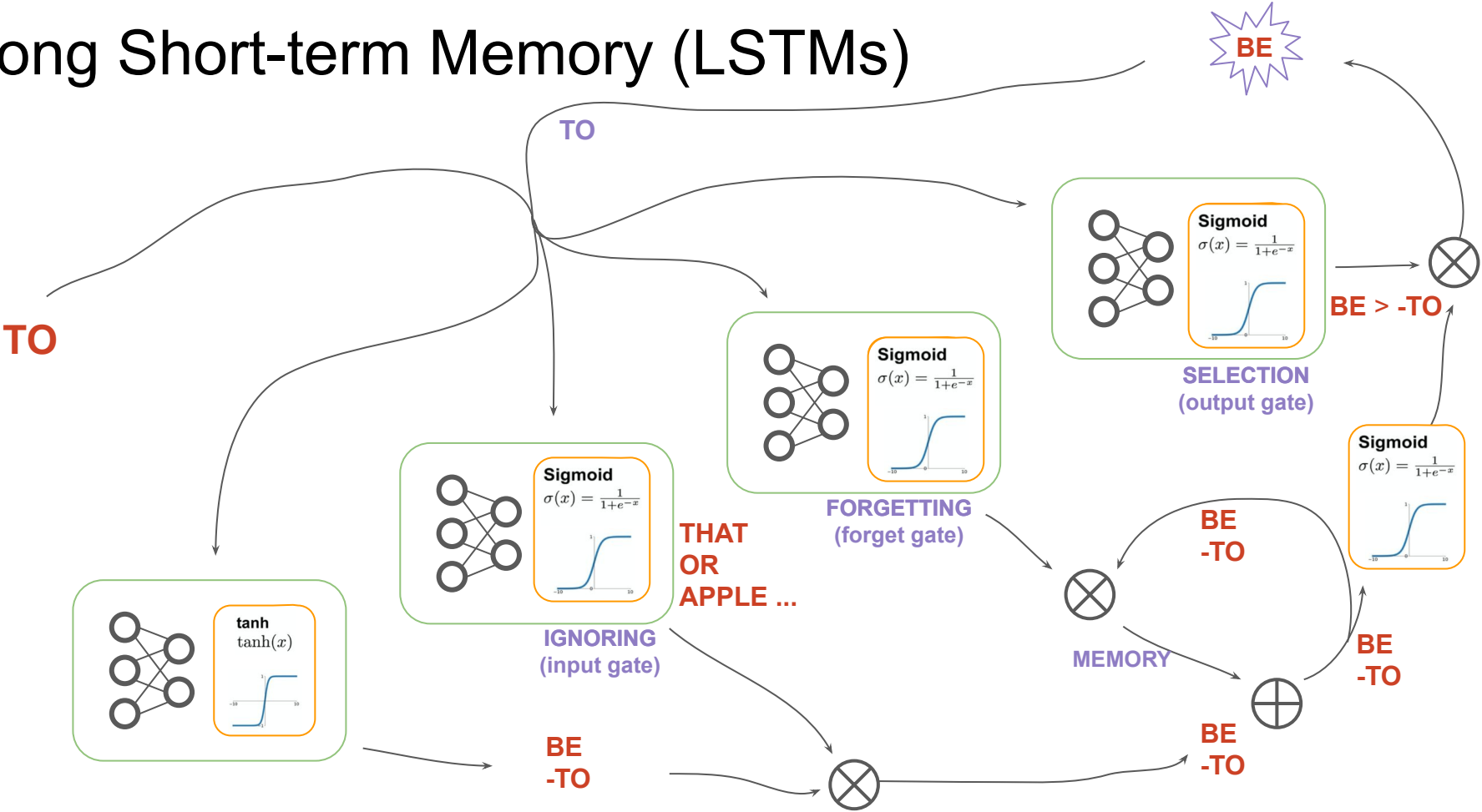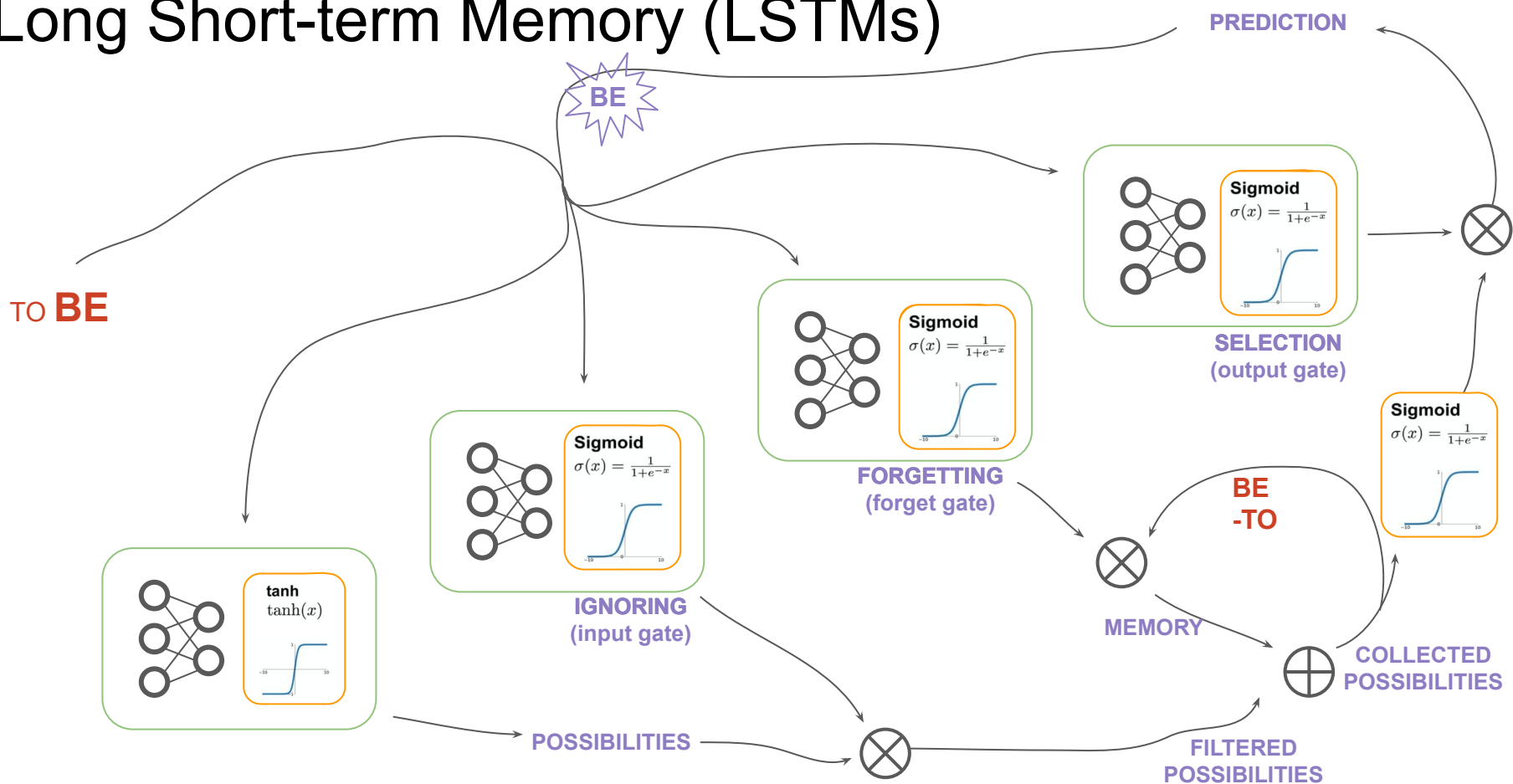PREDICTION

BE

TO **BE**

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

SELECTION
(output gate)

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

FORGETTING
(forget gate)

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

IGNORING
(input gate)

tanh
$\tanh(x)$

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

BE
-TO

MEMORY

COLLECTED
POSSIBILITIES

POSSIBILITIES

FILTERED
POSSIBILITIES

Long Short-term Memory (LSTMs)

PREDICTION

BE

TO **BE**

SELECTION
(output gate)

FORGETTING
(forget gate)

IGNORING
(input gate)

BE
-TO

MEMORY

COLLECTED
POSSIBILITIES

POSSIBILITIES

FILTERED
POSSIBILITIES

Long Short-term Memory (LSTMs)

PREDICTION

BE

TO **BE**

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

SELECTION
(output gate)

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

FORGETTING
(forget gate)

BE
-TO

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

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

IGNORING
(input gate)

MEMORY

tanh
$\tanh(x)$

COLLECTED
POSSIBILITIES

OR
:
-BE

FILTERED
POSSIBILITIES

Long Short-term Memory (LSTMs)

PREDICTION

BE

TO **BE**

APPLE
QUESTION
EXAMPLE ...

IGNORING
(input gate)

FORGETTING
(forget gate)

SELECTION
(output gate)

BE
-TO
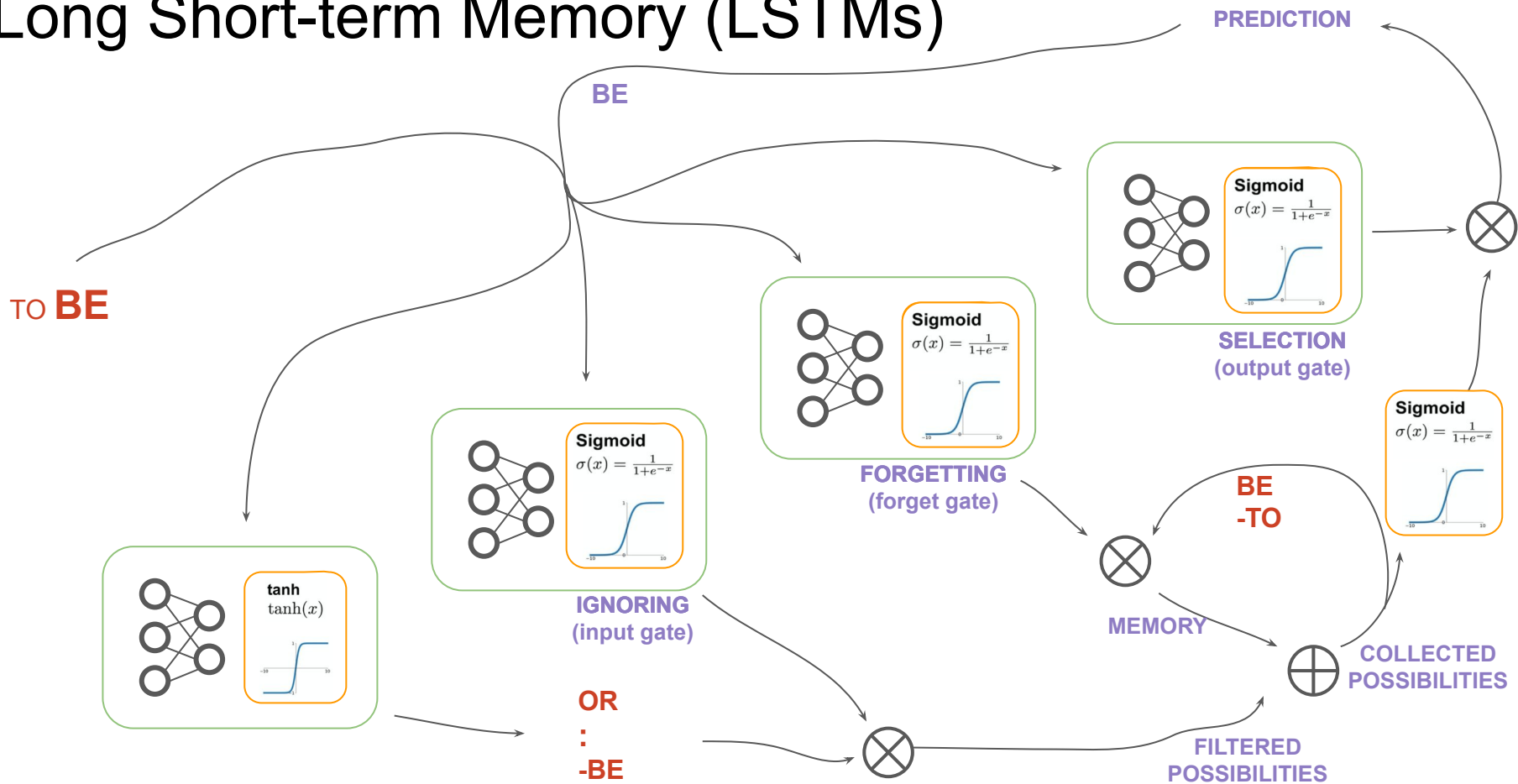
MEMORY

COLLECTED
POSSIBILITIES

OR
:
-BE

FILTERED
POSSIBILITIES

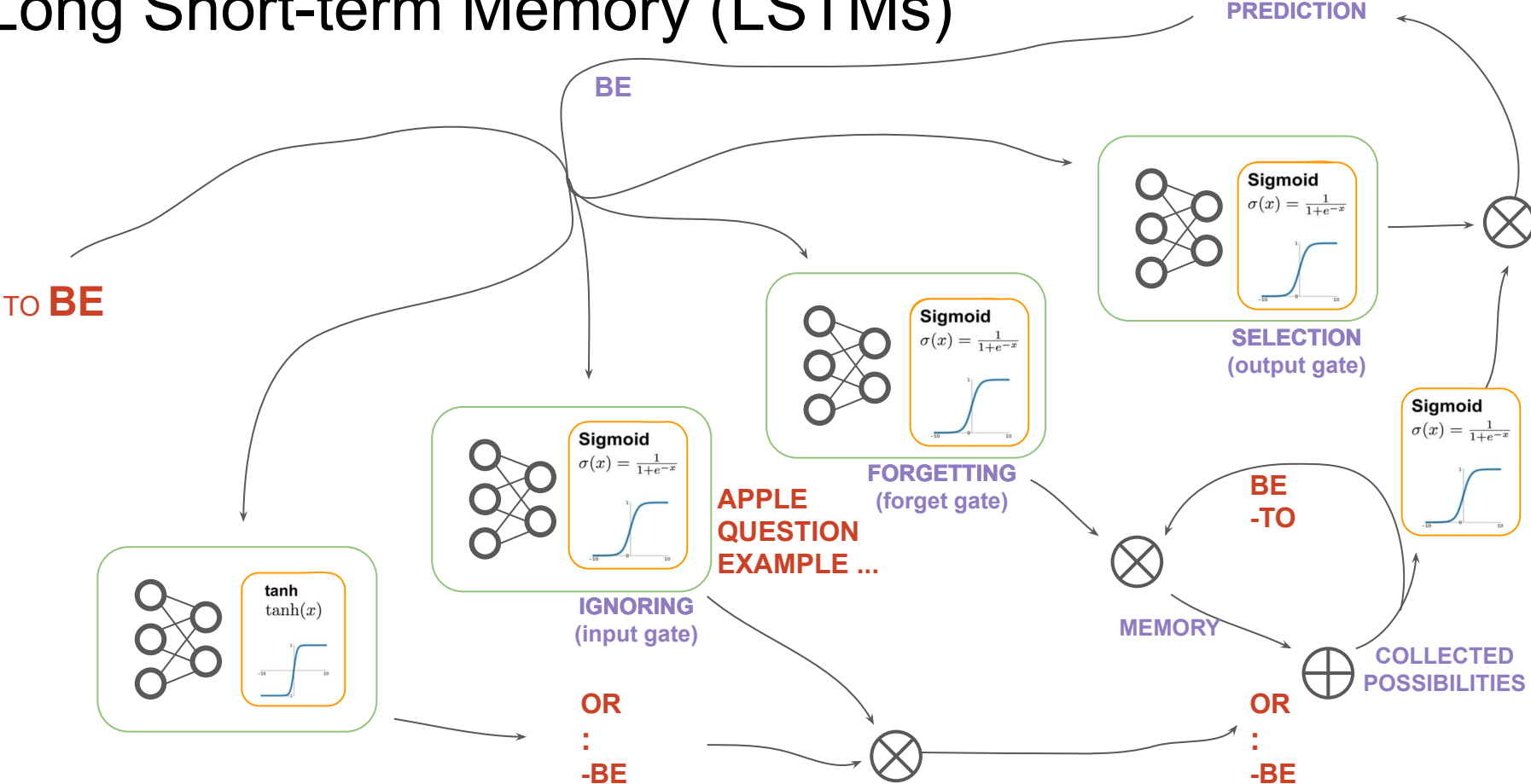# Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

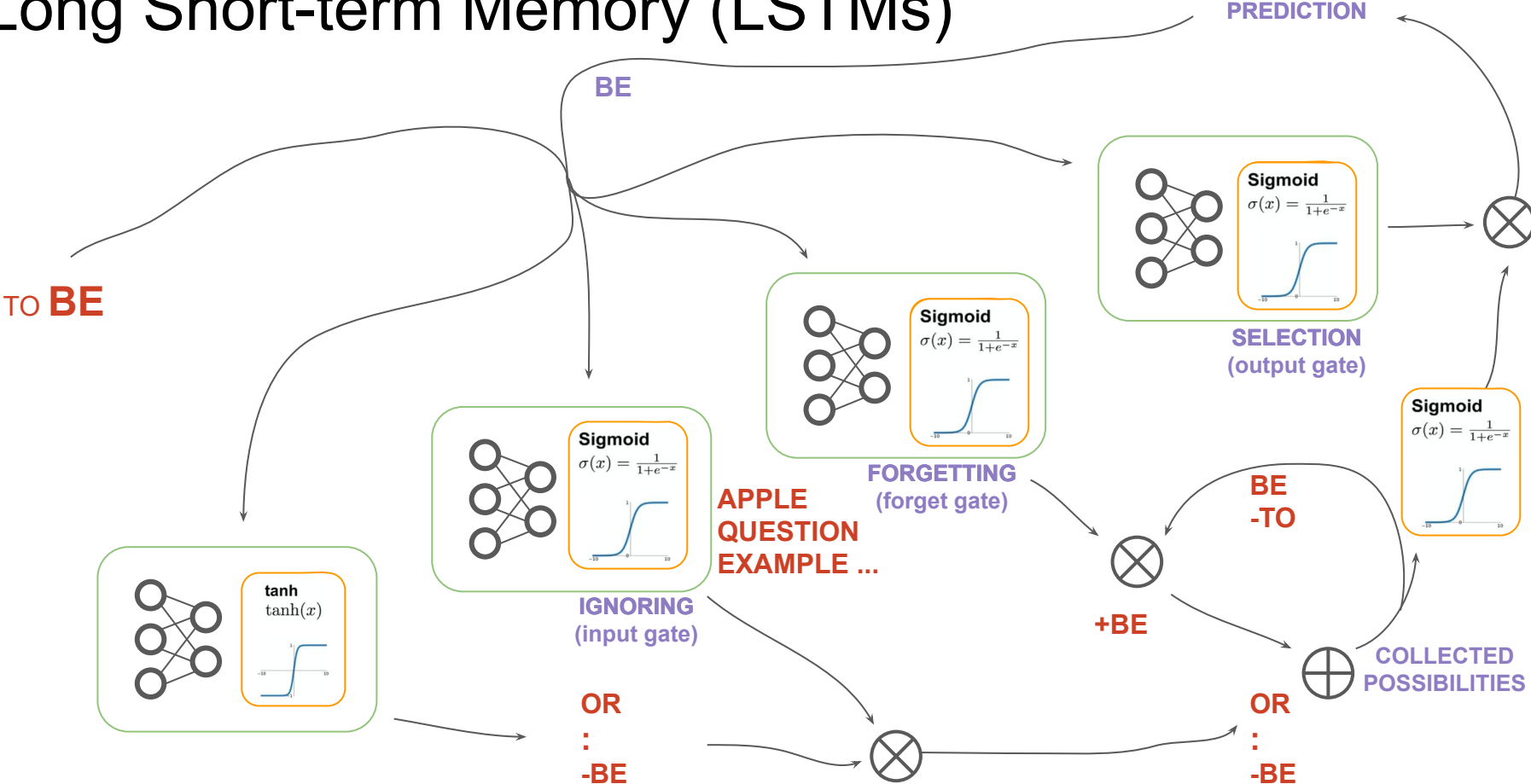PREDICTION

BE

TO **BE**

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

SELECTION
(output gate)

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

FORGETTING
(forget gate)

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

tanh
$\tanh(x)$

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

APPLE
QUESTION
EXAMPLE ...

IGNORING
(input gate)

BE
-TO

+BE

OR
:
-BE

COLLECTED
POSSIBILITIES

OR
:
-BE

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

PREDICTION

BE

TO **BE**

SELECTION
(output gate)

OR > :

FORGETTING
(forget gate)

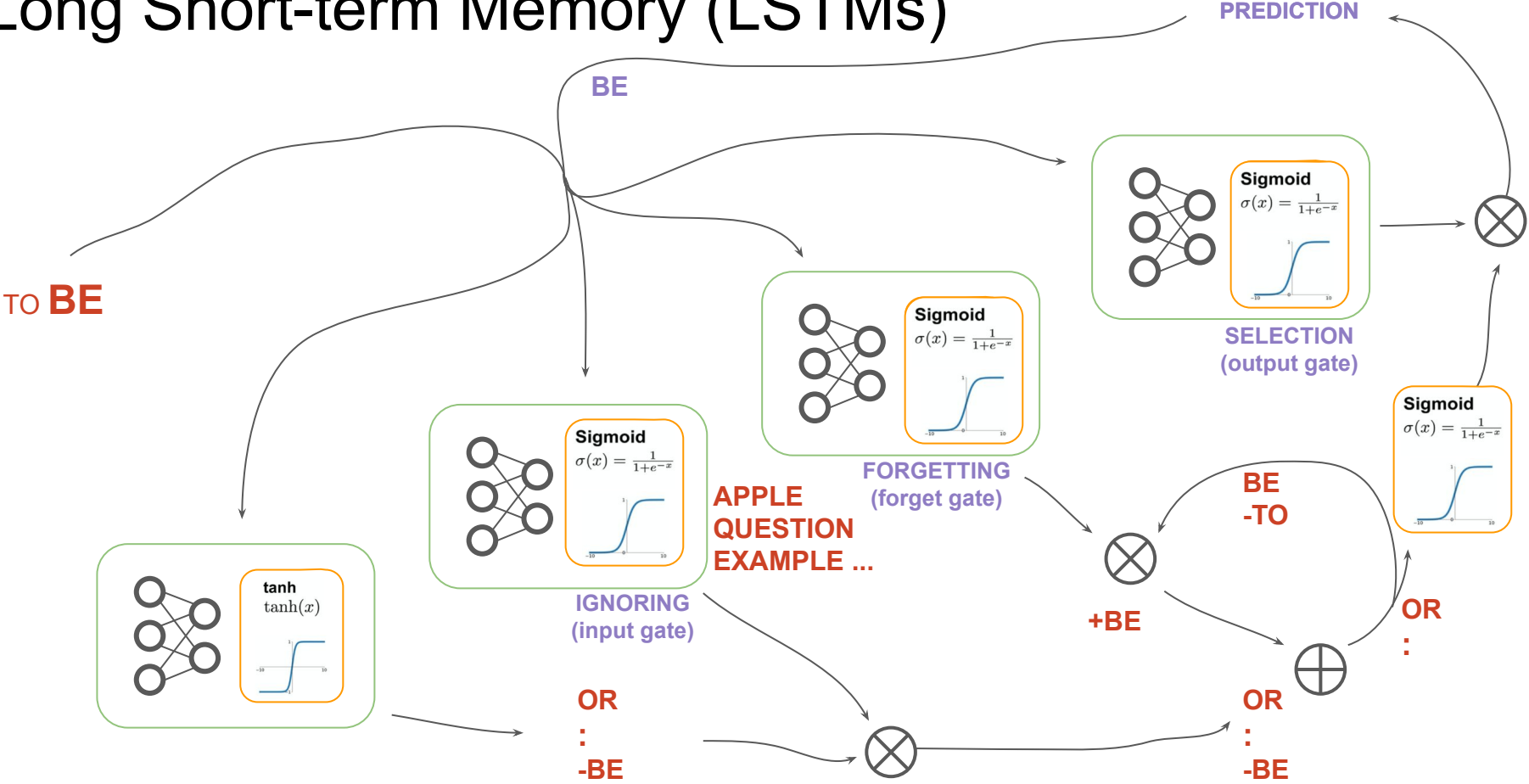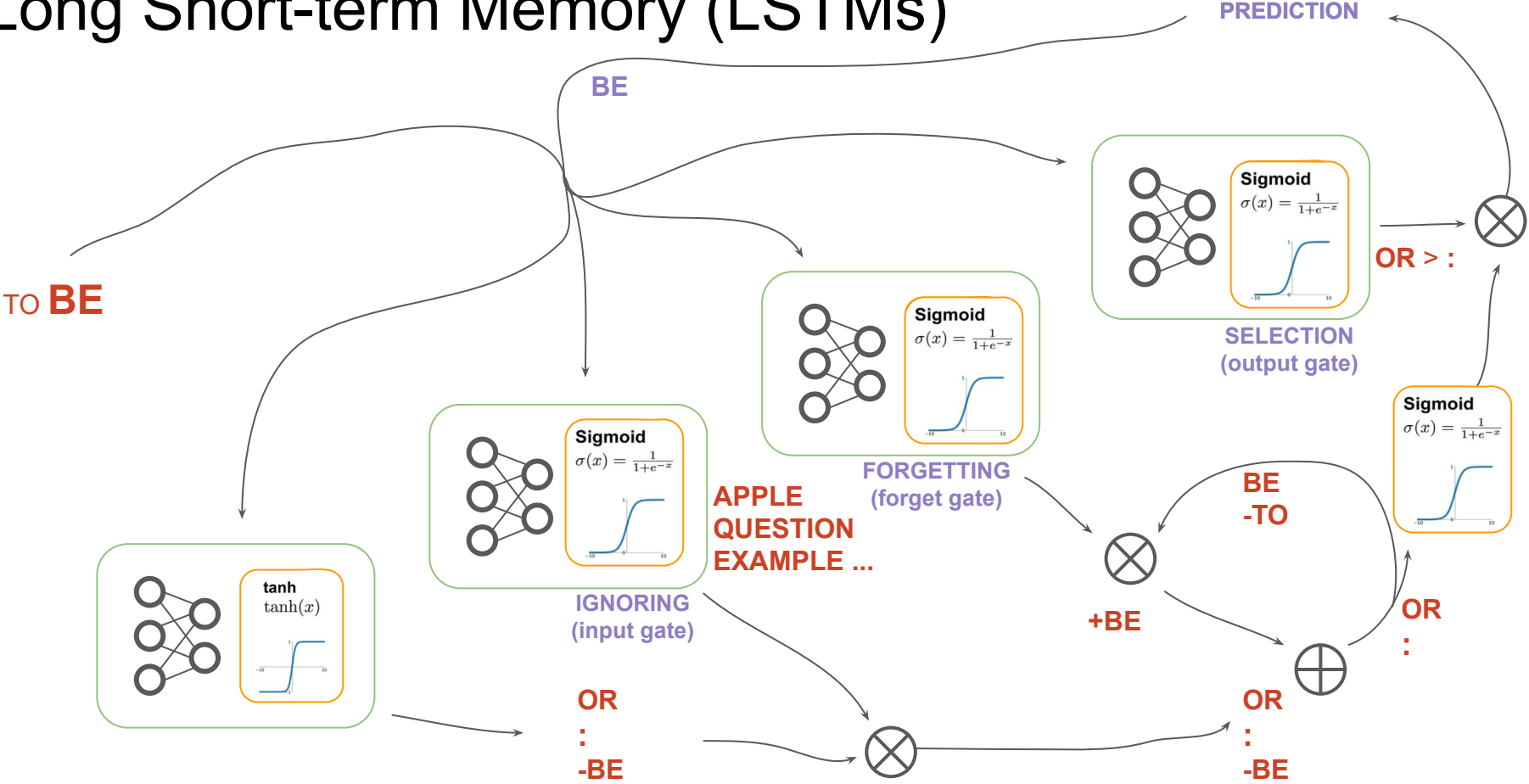APPLE
QUESTION
EXAMPLE ...

IGNORING
(input gate)

BE
-TO

+BE

OR
:
-BE

OR
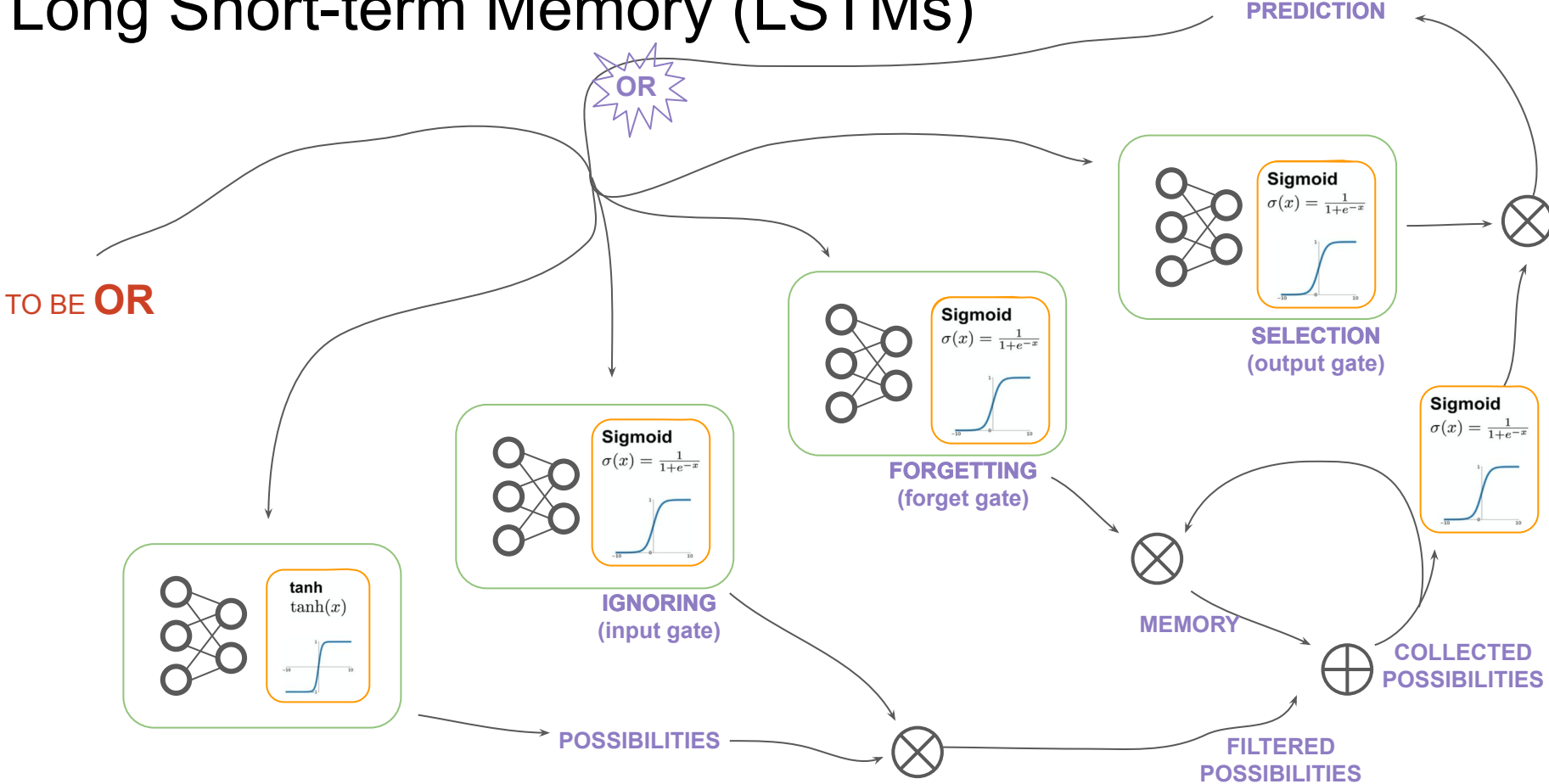:
-BE

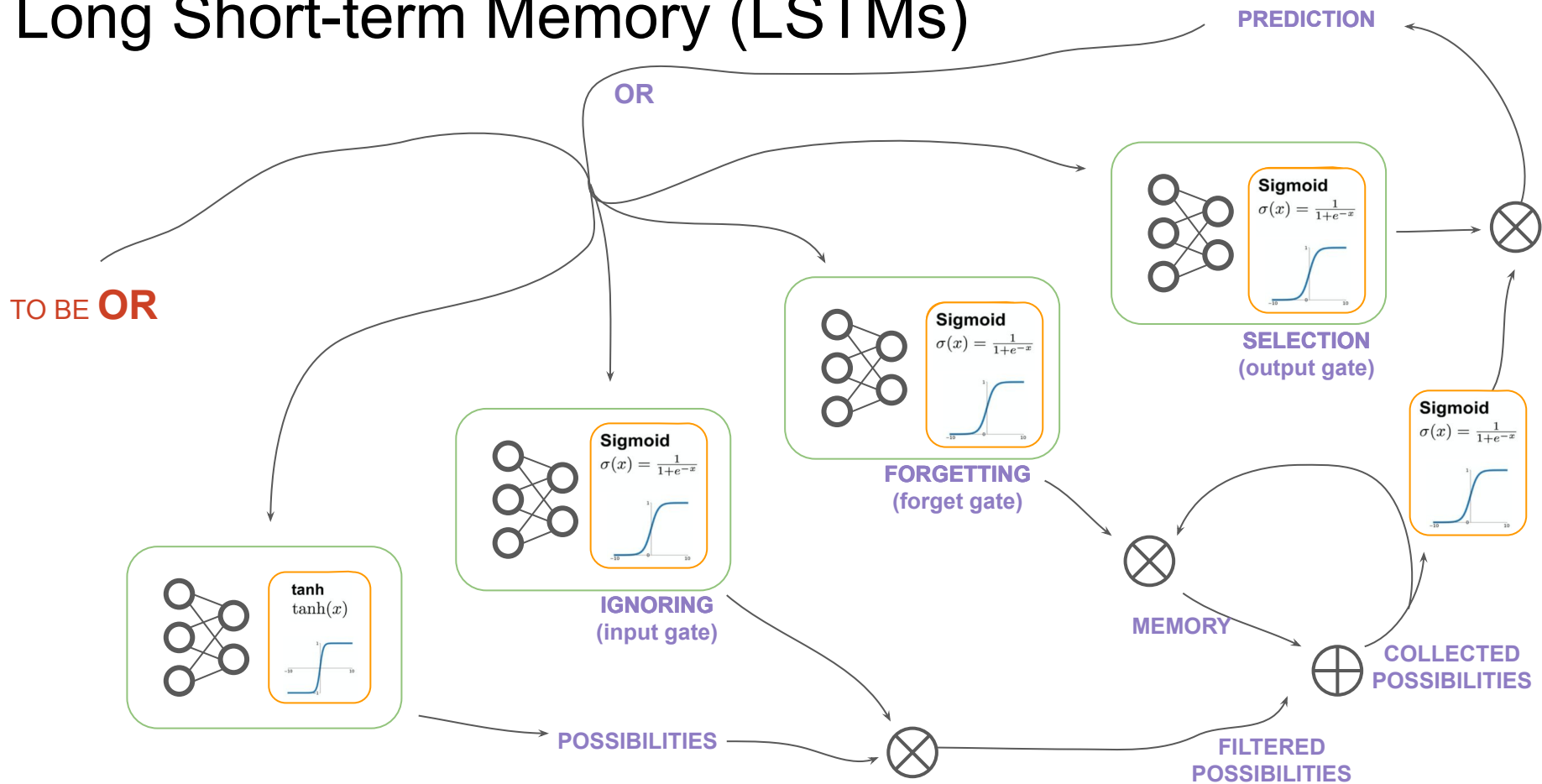OR
:

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

Long Short-term Memory (LSTMs)

Gated Recurrent Unit (GNUs)

- Similar to LSTMs
    - Uses a hidden state instead of the gates of the cell state
    - Has an "update" gate that is like "forget" in LSTMs
    - Has "reset" gate to decide how much past information to forget

Might be helpful…

- Tricks about regularizing and optimizing LSTM Language Models:
  - https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0

Might be helpful…

- Tricks about regularizing and optimizing LSTM Language Models:
  - https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0