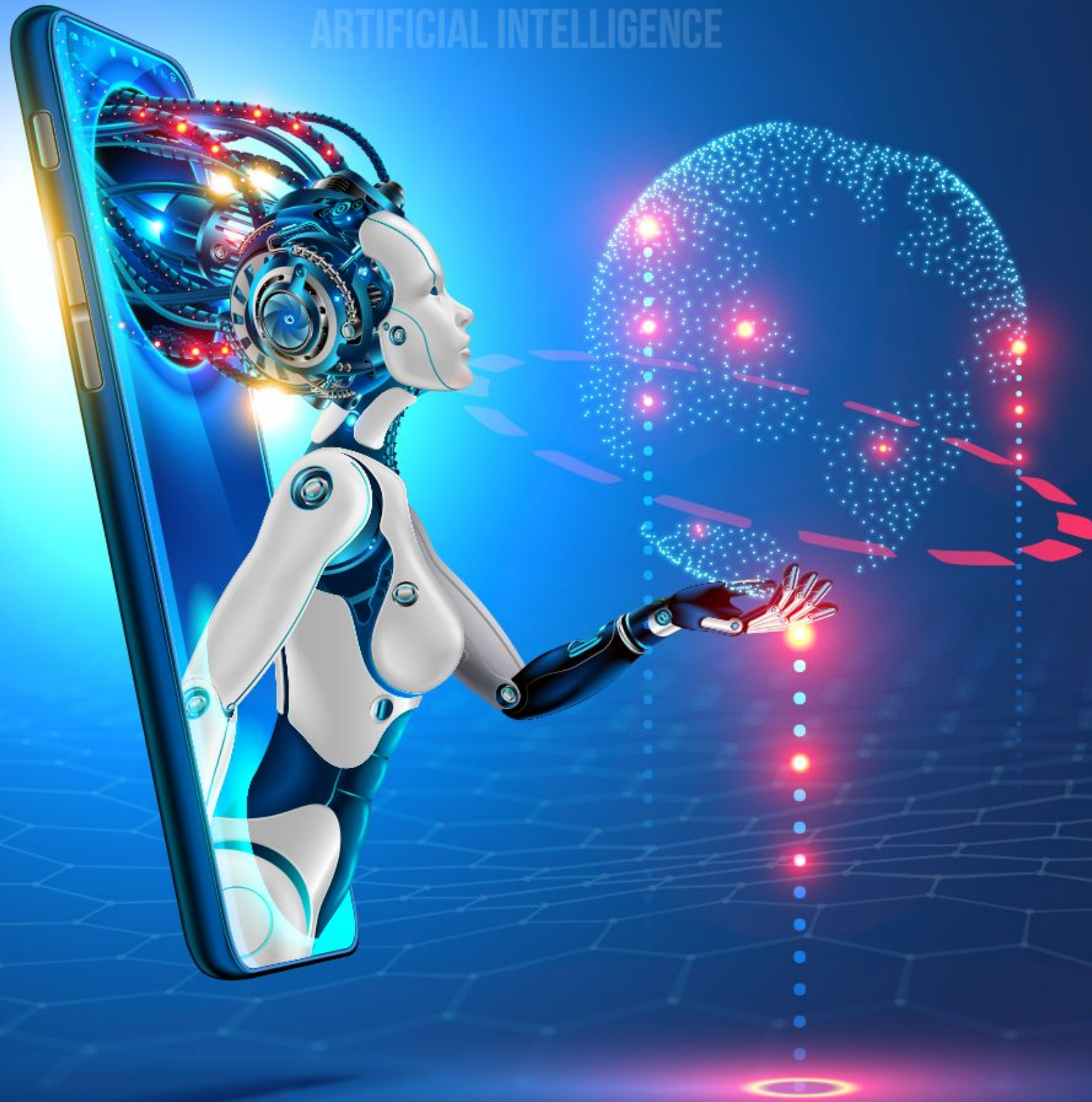


DATA AND
ARTIFICIAL INTELLIGENCE



simplilearn

PURDUE
UNIVERSITY

Natural Language Processing and Speech Recognition

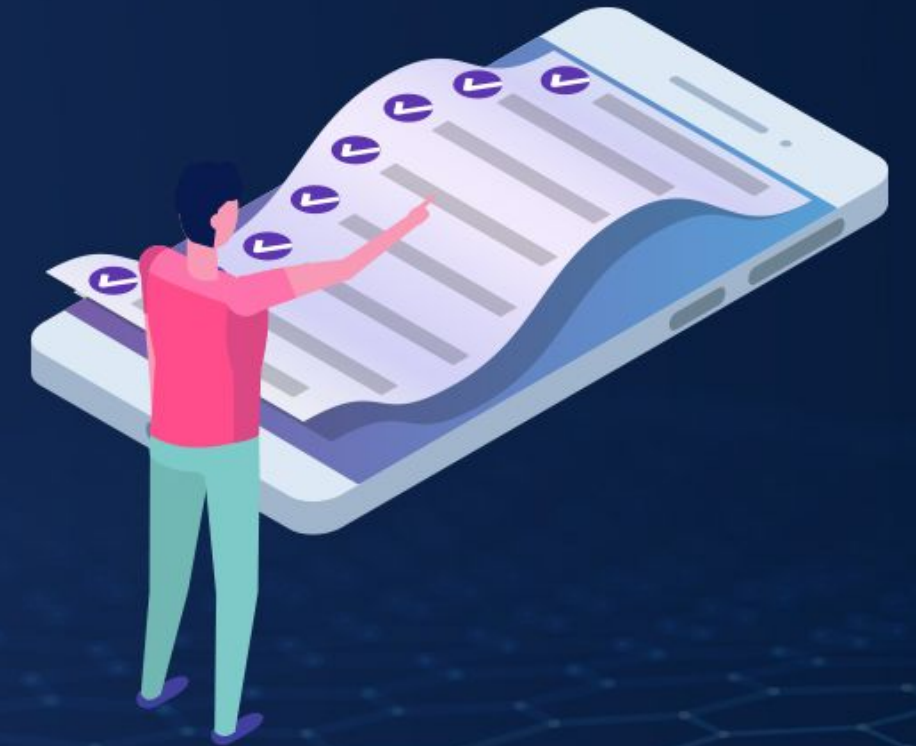


NLP with Machine Learning and Deep Learning

Learning Objectives

By the end of this lesson, you will be able to:

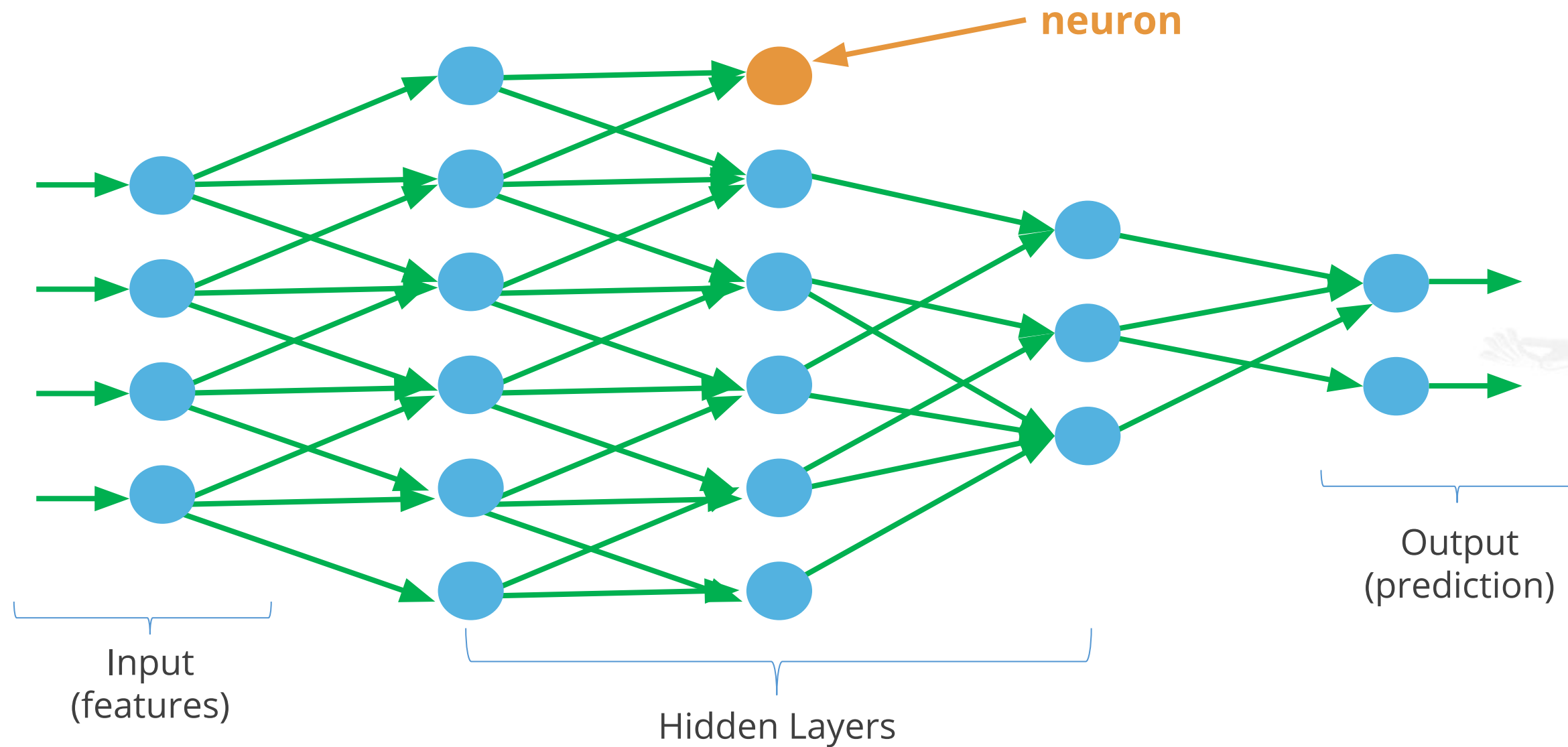
- 🕒 Explain Neural Network and Recurrent Neural Networks
- 🕒 Explain Neural Machine Translation
- 🕒 Define Text Classification and Text Summarization
- 🕒 Explain document clustering, attention mechanism, and question answering engine
- 🕒 Demonstrate spam-ham classification using ML
- 🕒 Create a script summarizing a news by finding only important information



Neural Networks and Recurrent Neural Networks (RNN)

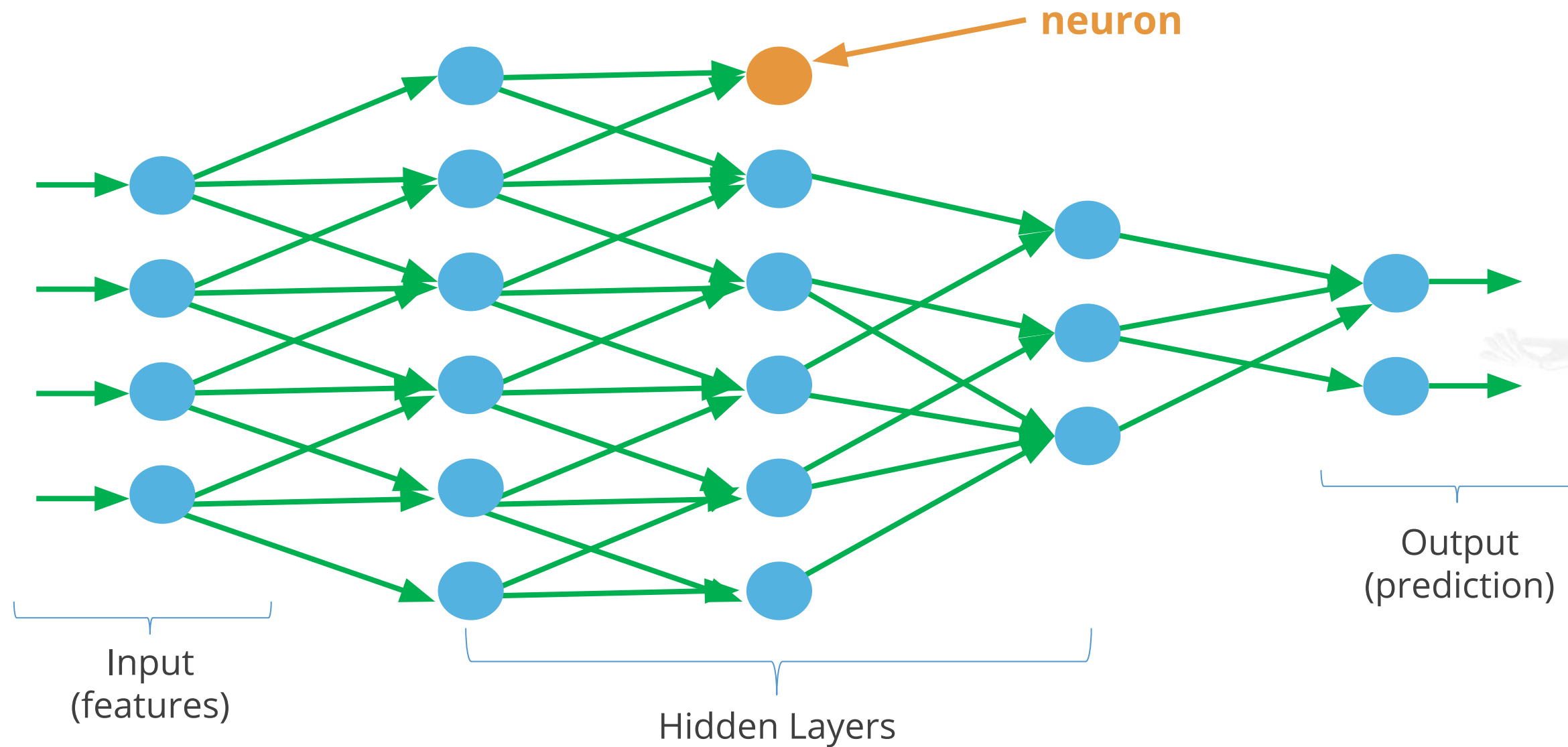
Neural Network

Neural Networks used in Deep Learning, consists of different layers connected to each other and work on the structure and functions of a human brain. It learns from huge volumes of data and uses complex algorithms to train a neural net.

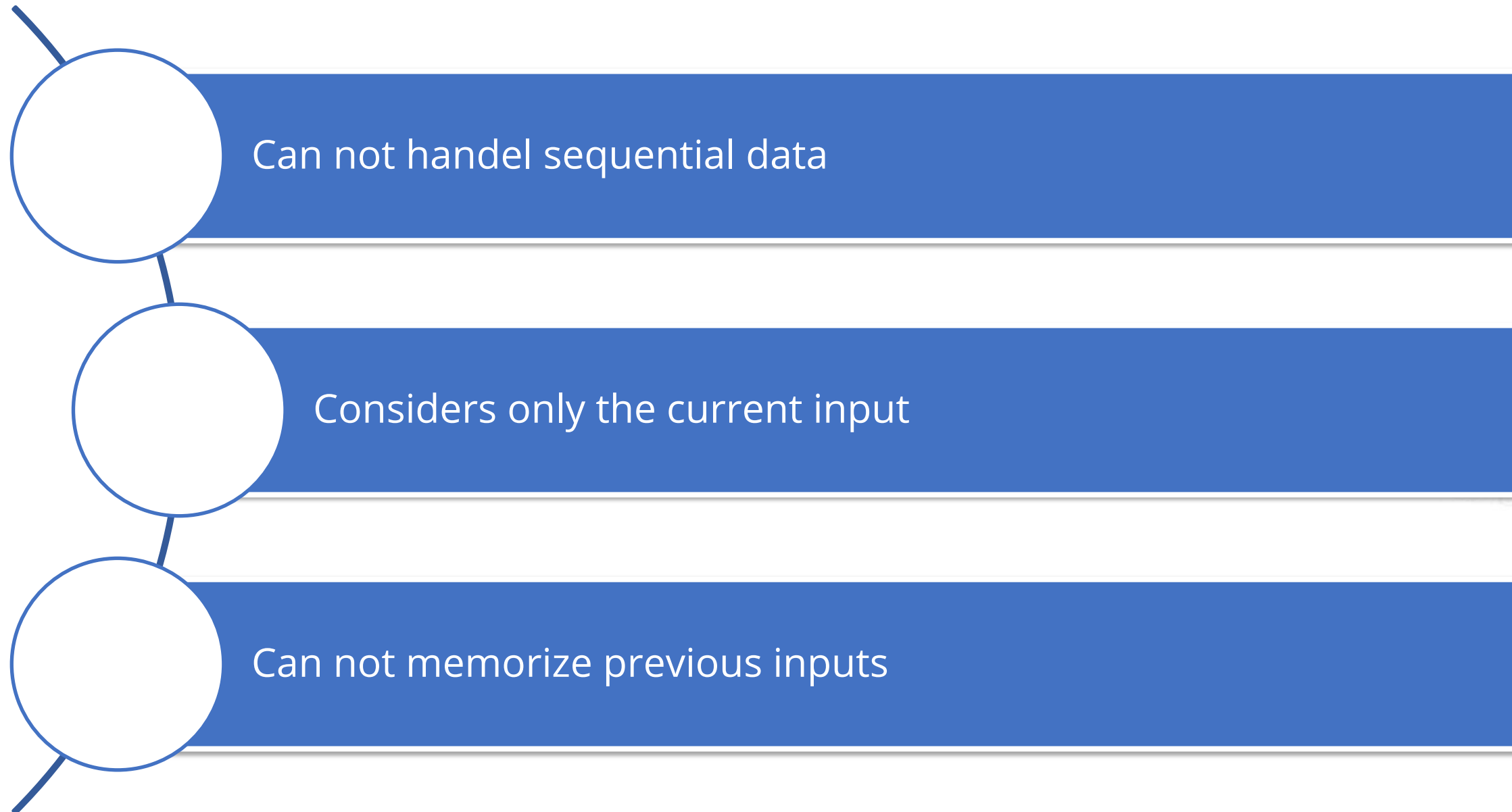


Feed Forward Neural Network

It is a type of neural network where every unit in a layer is connected with all the units in the previous layer

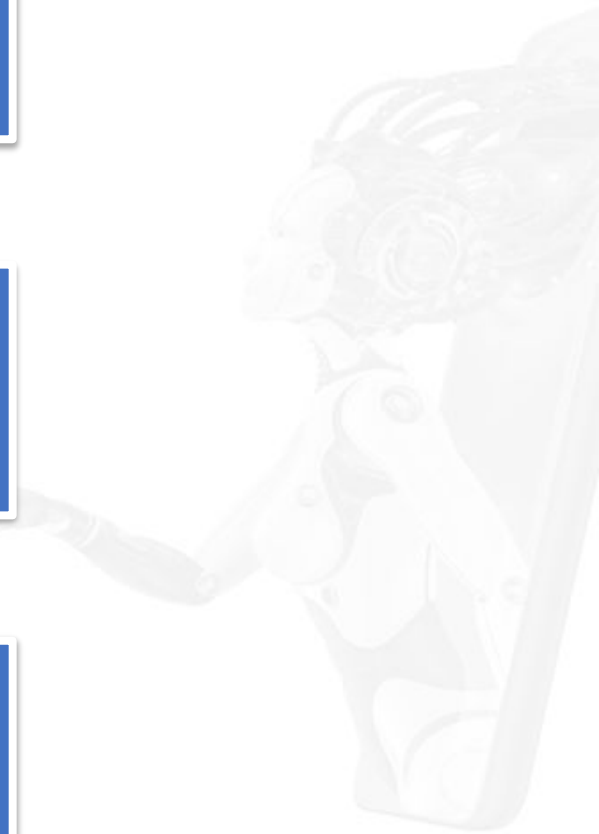
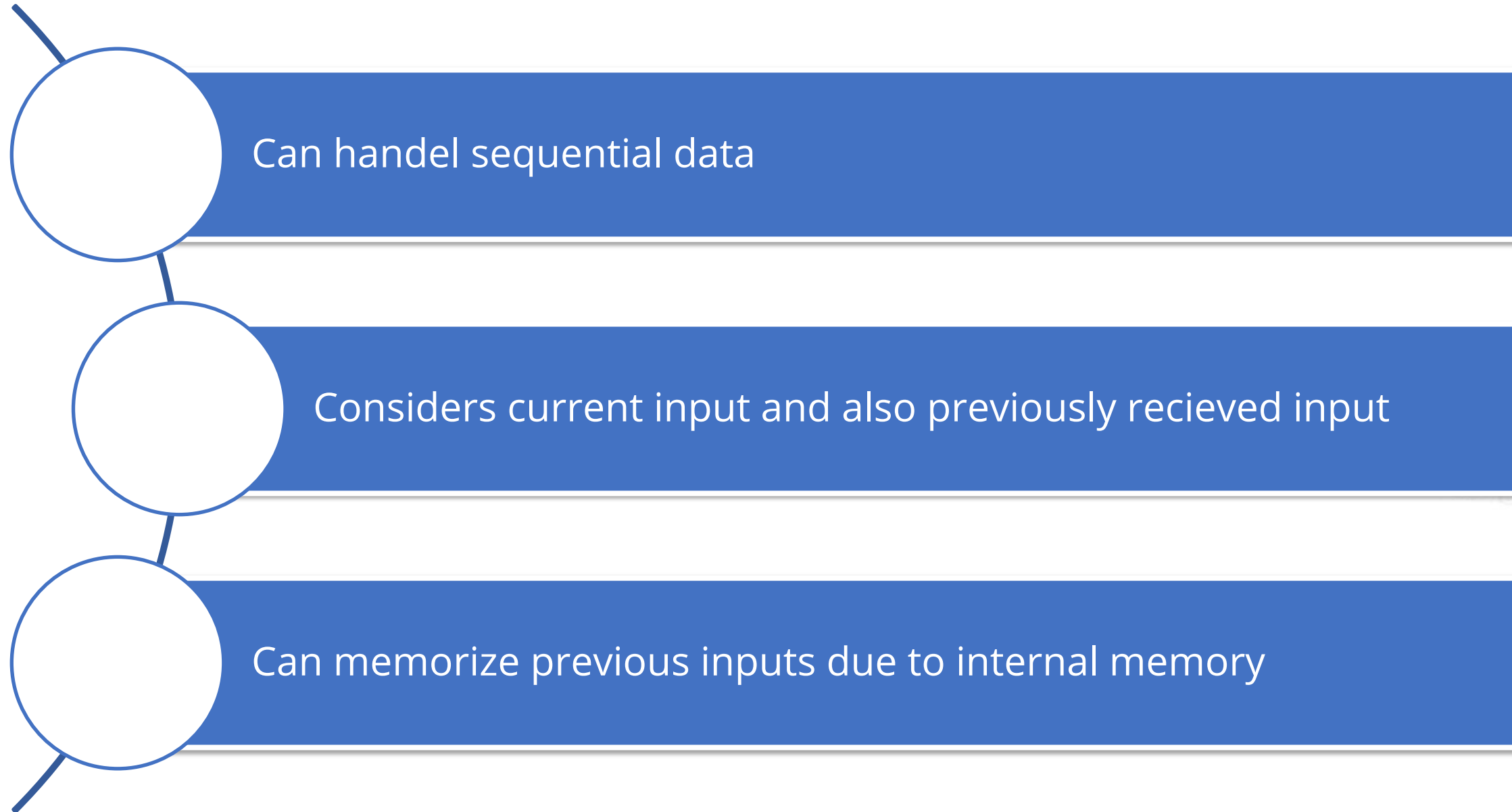


Disadvantages of Feed Forward Neural Network



Note: To over come these, RNN came in to the picture.

Recurrent Neural Networks (RNN)



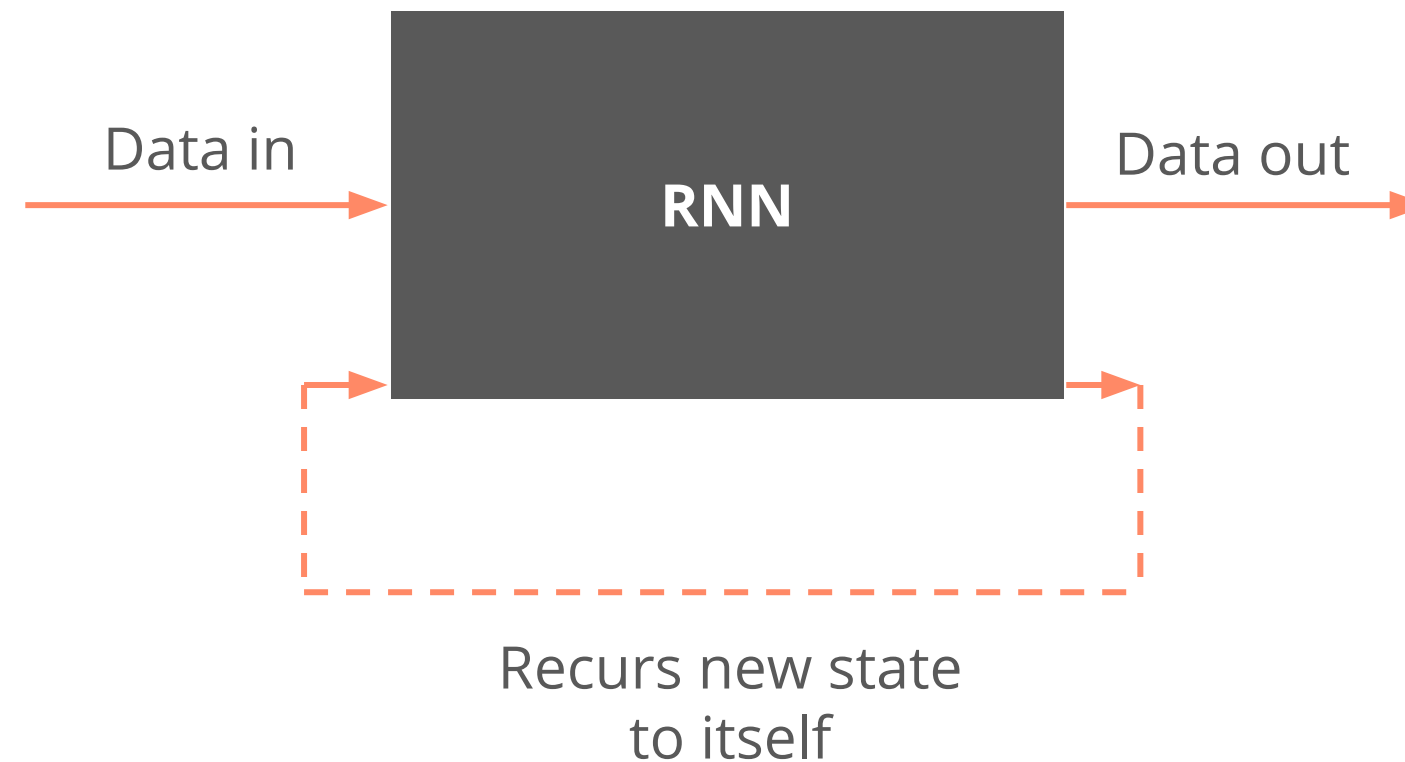
Recurrent Neural Networks (RNN)

Recurrent Neural Network(RNN) are a type of Neural Network where the output from previous step are fed as input to the current step.



The RNN Model

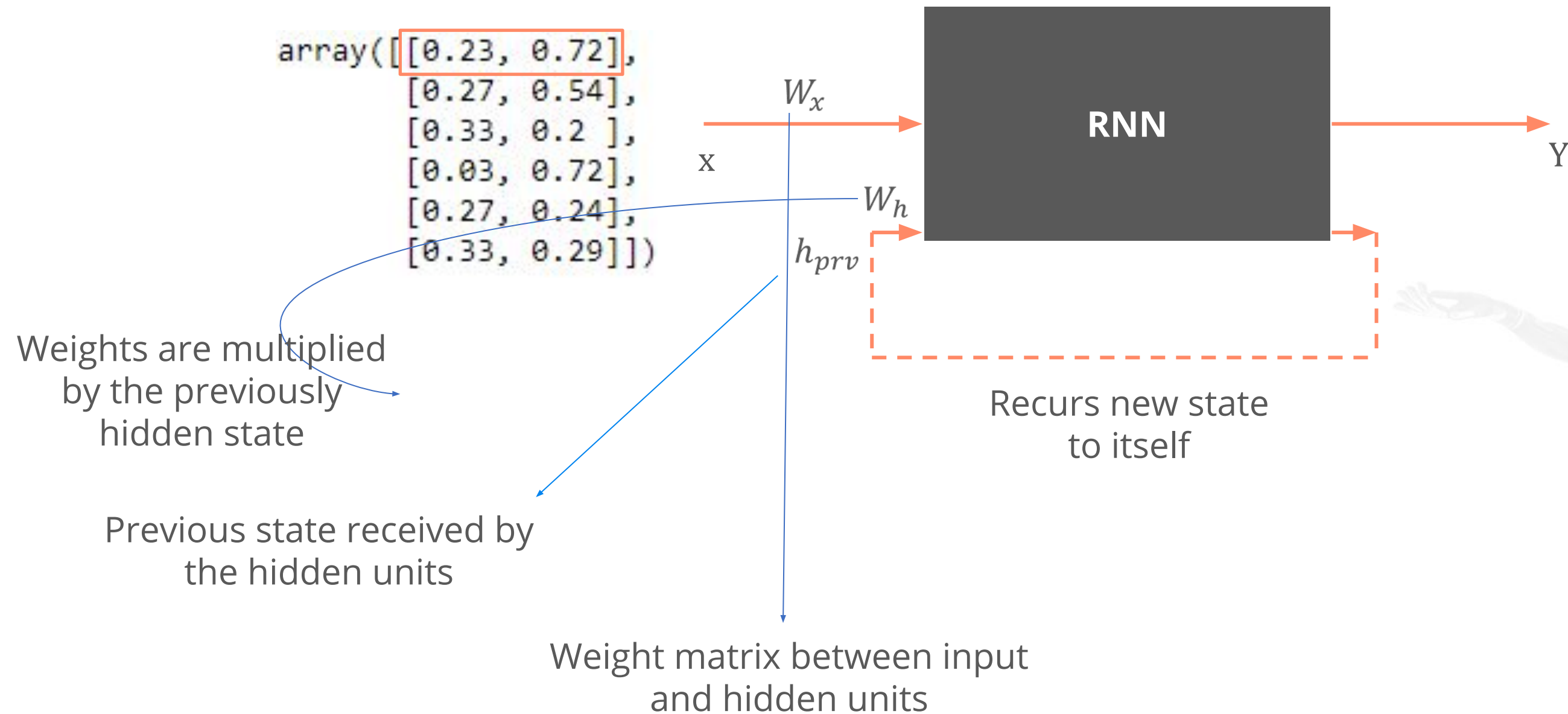
The RNN remembers the analysis done upto a given point by maintaining a **state**.



Note: You can think of the **state** as the **memory** of RNN which recurs into the net with each new input.

RNN: Working

The first data point flows into the network as input data, denoted as x .



Introduction to Attention Mechanism

What Is Attention Mechanism?

1

Attention mechanism is used in machine translation.

2

It has bidirectional RNN layer which takes word input from any language (Ex. French) and generate feature output for each word.

3

Another RNN is used after bidirectional RNN to generate the translations. This takes all features from previous step bidirectional RNN and predicts word to target language.

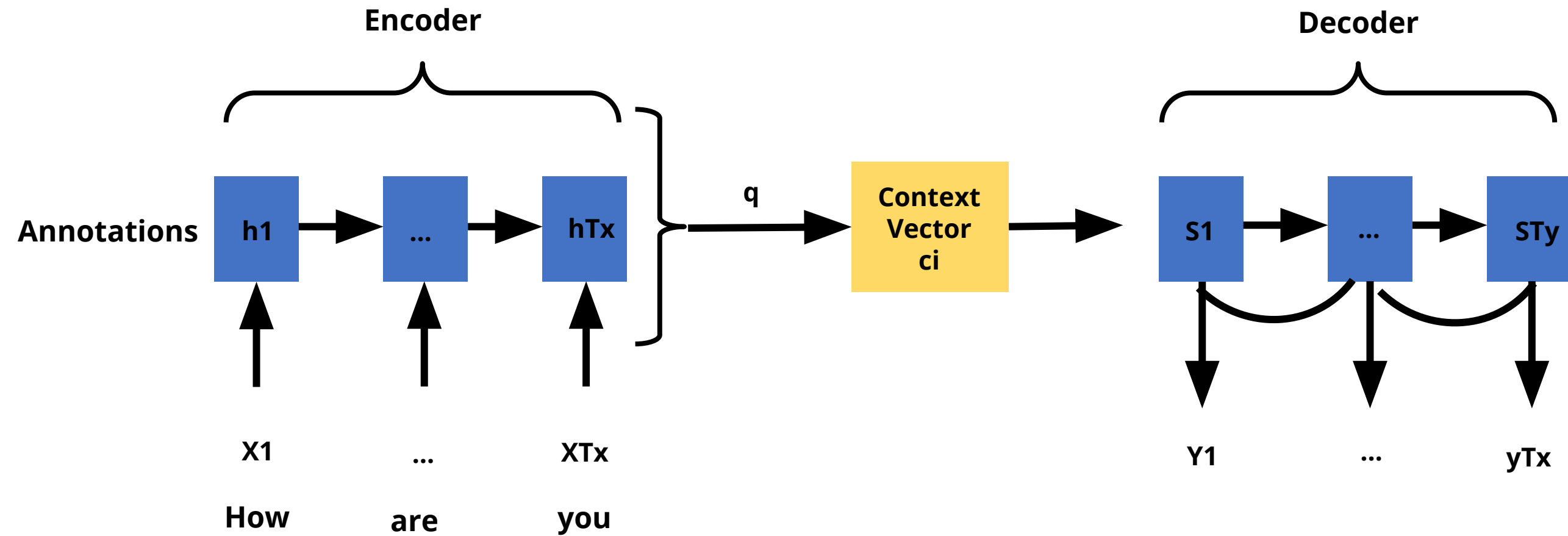
4

It uses matrix where columns are input sentences and rows are the translated sentences.

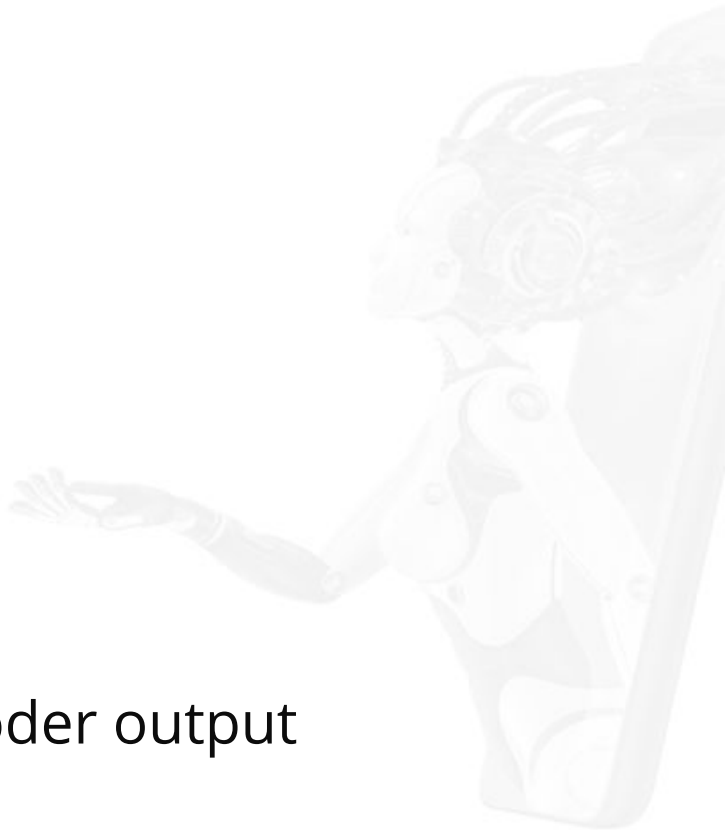
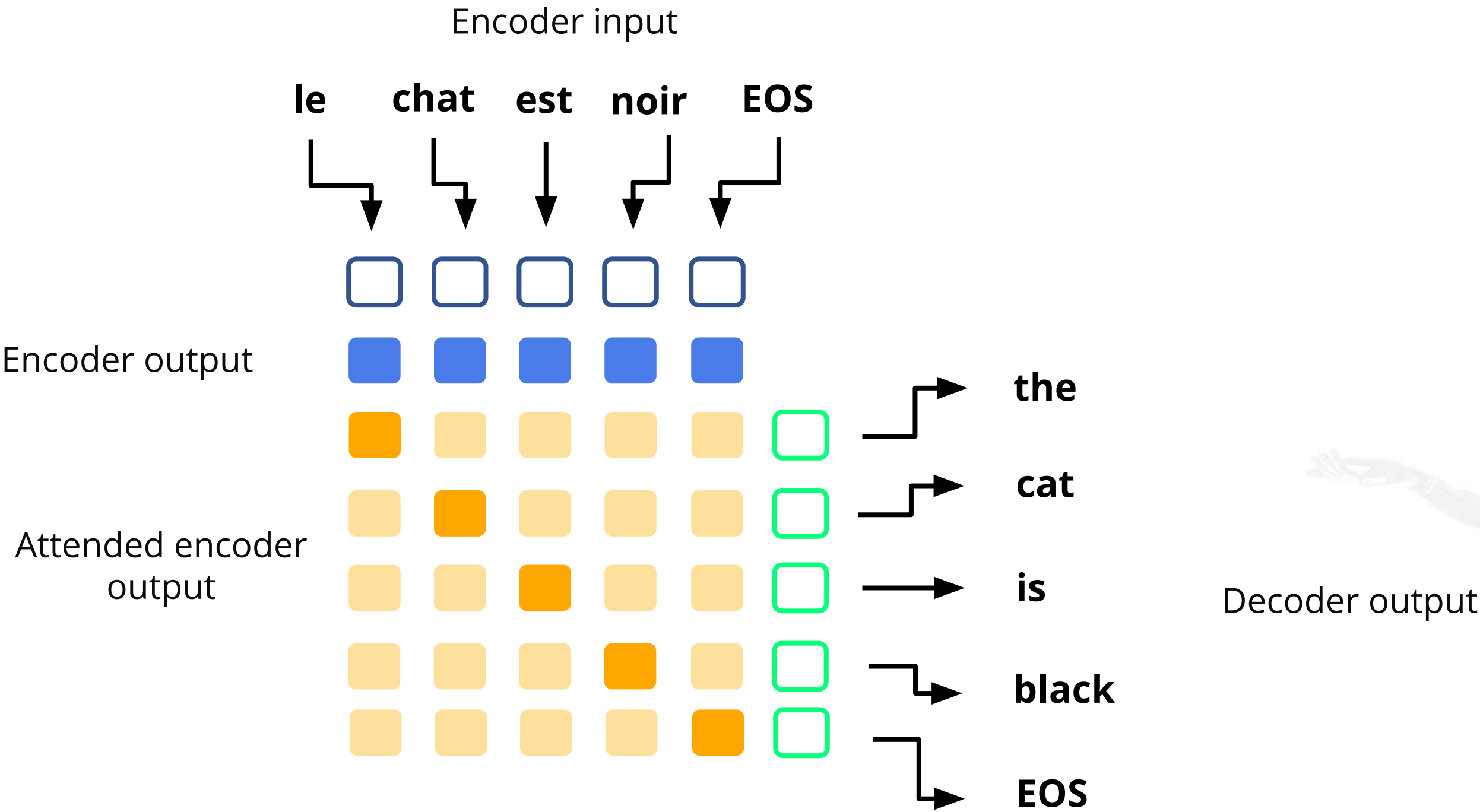
5

Word prediction is done using features extracted from bidirectional RNN and based on relation to previous word in translation RNN.

What Is Attention Mechanism?



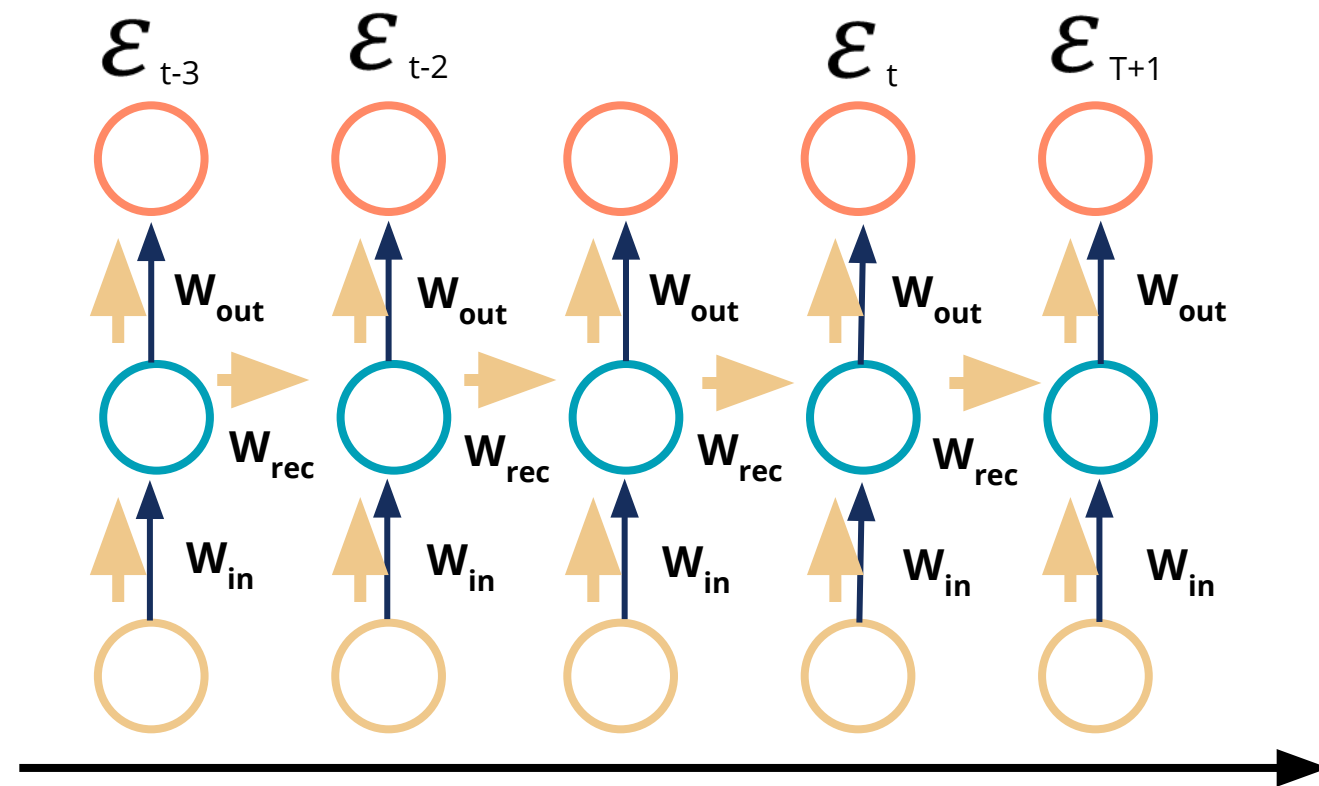
Attention Mechanism: Example



Long Short-Term Memory (LSTM)

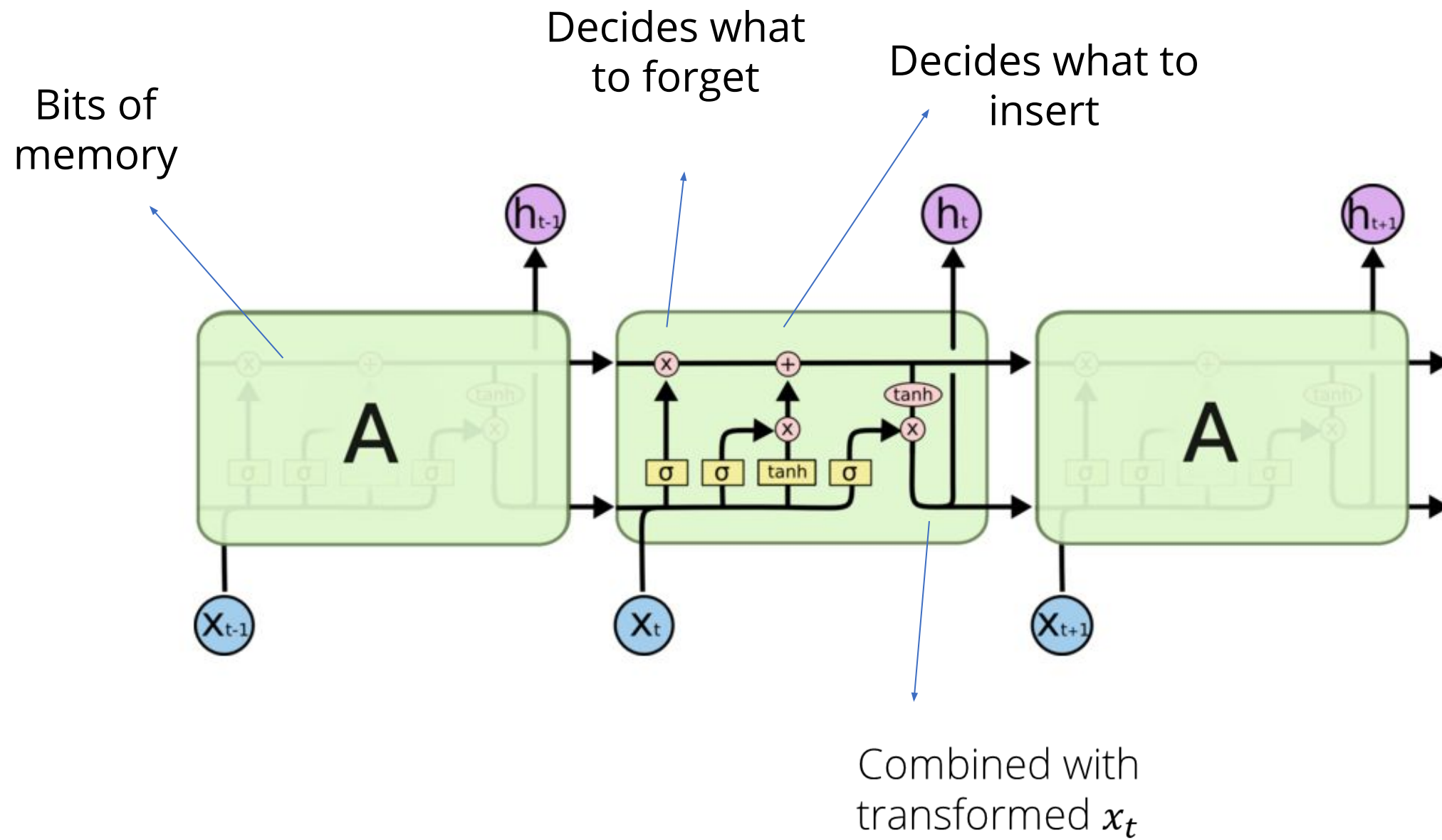
The Problem of Vanishing Gradient with RNNs

The problem arises while updating weights in RNN. These weights connect the hidden layers to themselves in the unrolled temporal loop.



Note: When any figure is multiplied by a small number, its value decreases very quickly.

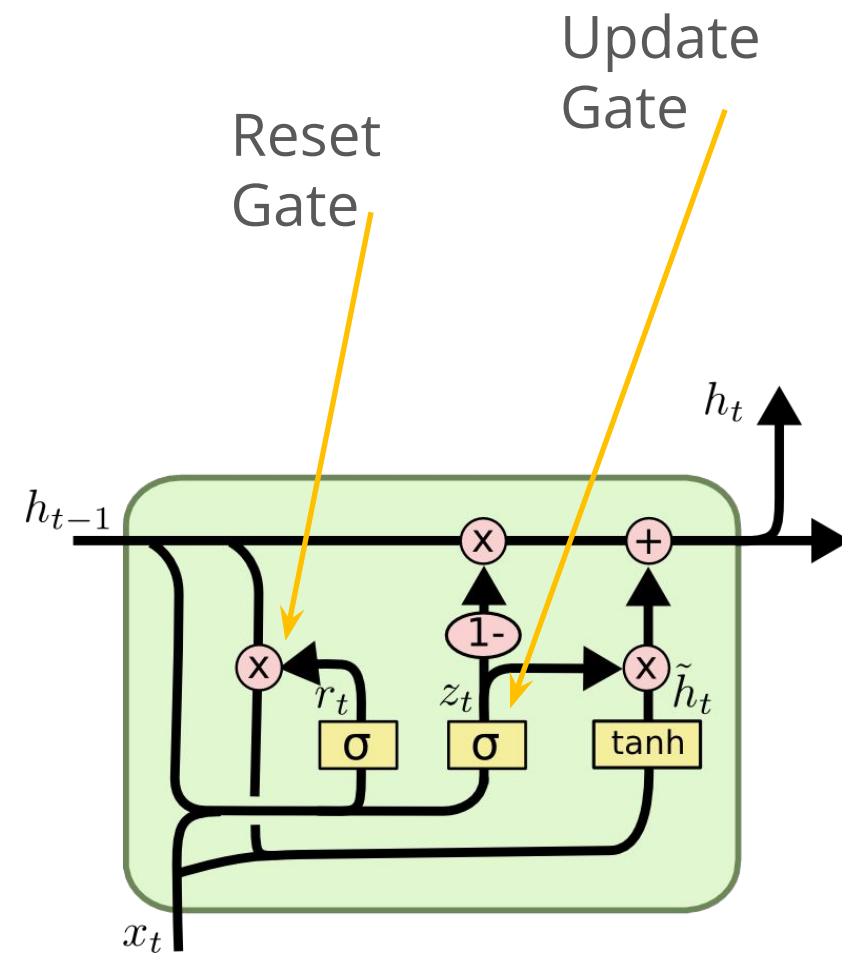
LSTM Architecture



Gated Recurrent Unit (GRU)

GRU Architecture

Performs label predictions against random data.



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

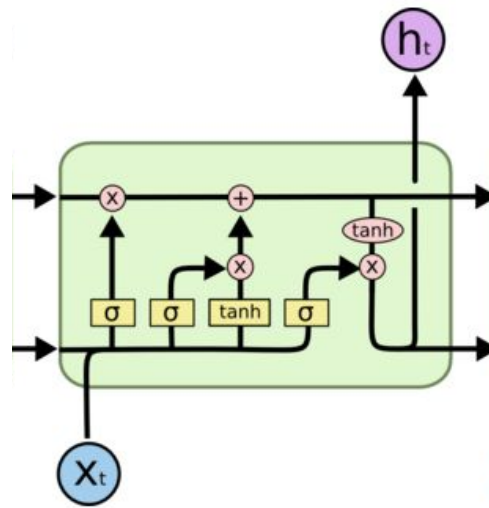
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

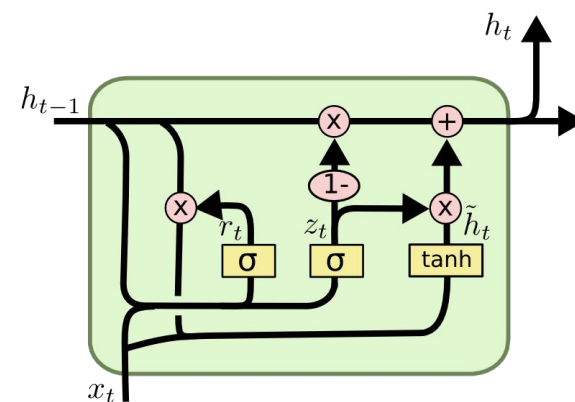
LSTM vs. GRU

LSTM



- Tracks long-term dependencies while mitigating the vanishing or exploding gradient problems. It does so via input, forget, and output gates
- Controls the exposure of memory content

GRU

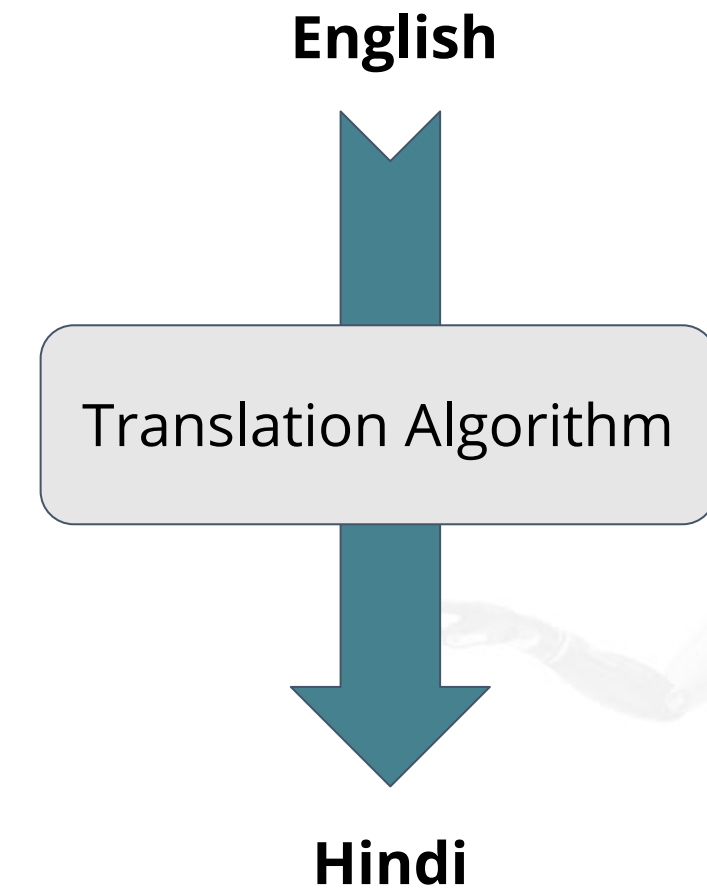


- Tracks long-term dependencies using a reset gate and an update gate
- Exposes the entire cell state to other units in the network

Introduction to Machine Translation

What Is Machine Translation?

- 1 Machine translation is a process to translate the text or speech from one language to other language.
- 2 It is a subfield of computational linguistics.
- 3 Translation process preserves the meaning of the input text.



Types of Machine Translation

Rule-Based Machine Translation

It uses a combination of language and grammar rules with common words.

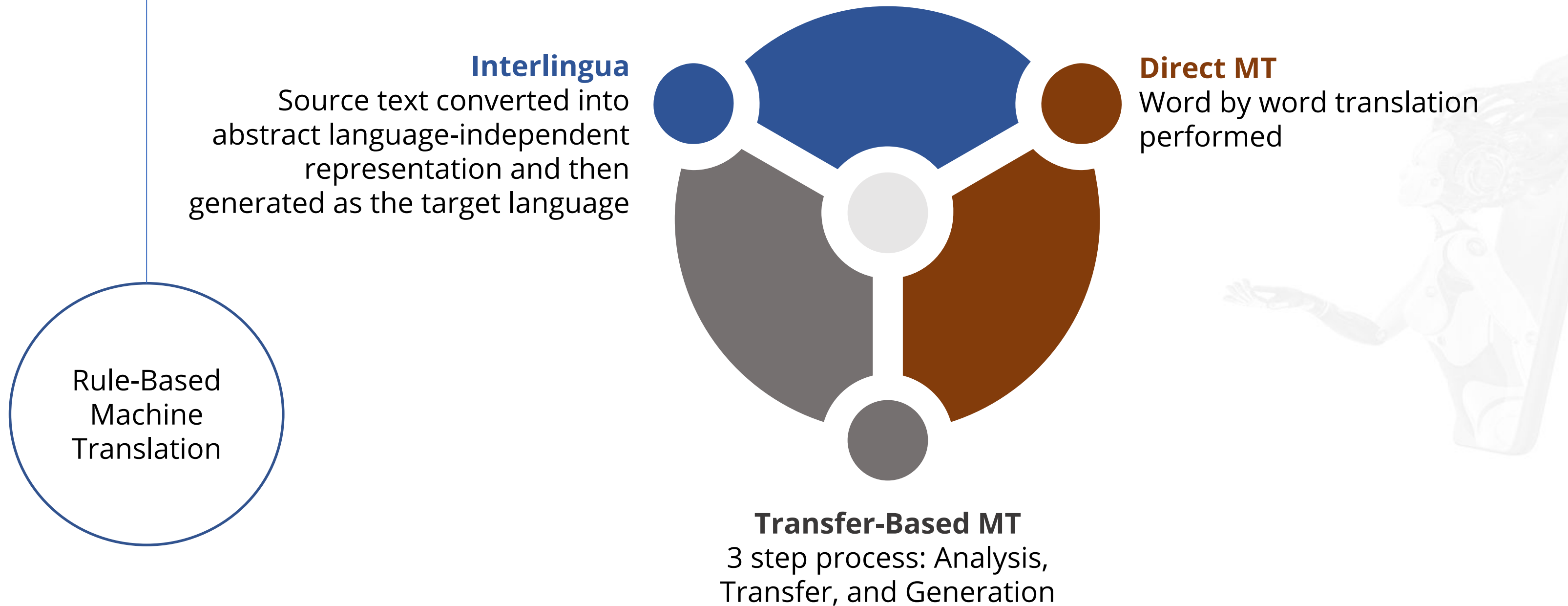


Statistical Machine Translation

It uses statistical translation models based on bilingual text corpora.



Types of Machine Translation



Types of Machine Translation

- Uses statistical approach where data is the main part used for translation
- Requires huge amount of data to train
- Neural machine translation approach is used in machine translation that uses a large artificial neural network

Statistical
Machine
Translation

Training

Testing

Model Building

Introduction to Neural Machine Translation

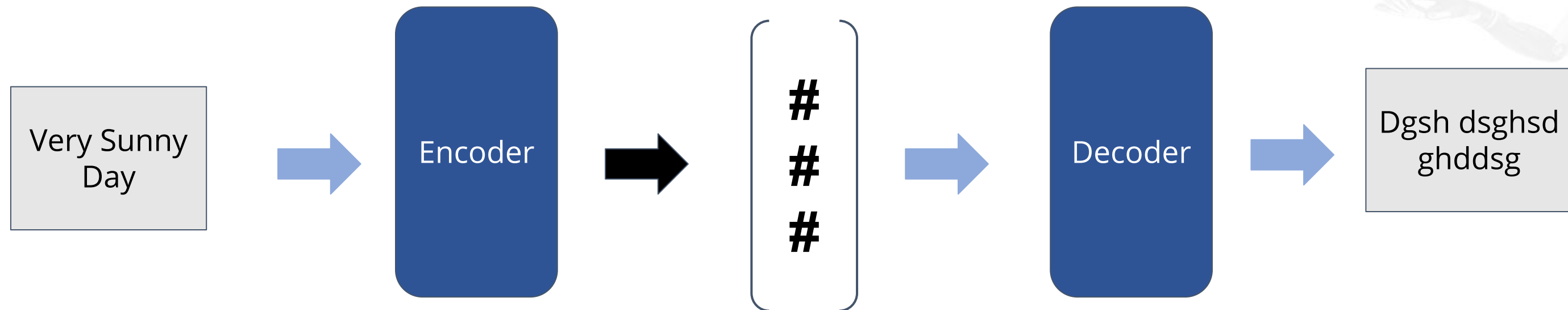
Neural Machine Translation



Neural Machine Translation

NMT uses the Encoder Decoder structure for **seq2seq** model.

Encoder converts the source text into intermediate state, and this is converted into target text by the decoder.



Neural Machine Translation

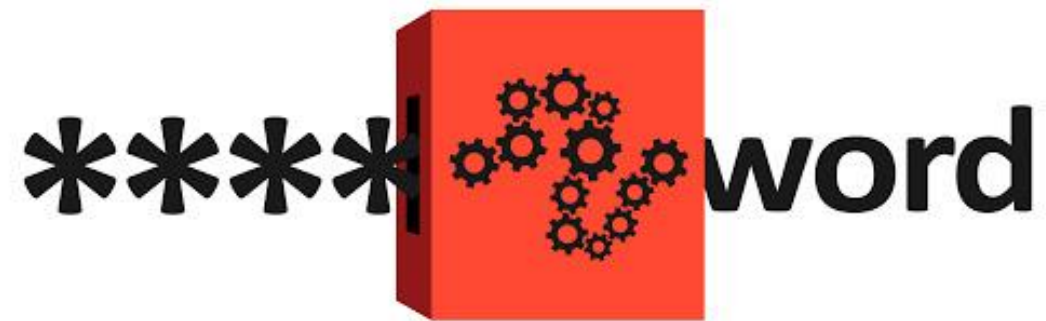
Encoder



Encoder embeds in English and passes the embedded text through multiple LSTM, encodes it, and gives it to the decoder LSTM layer.

Neural Machine Translation

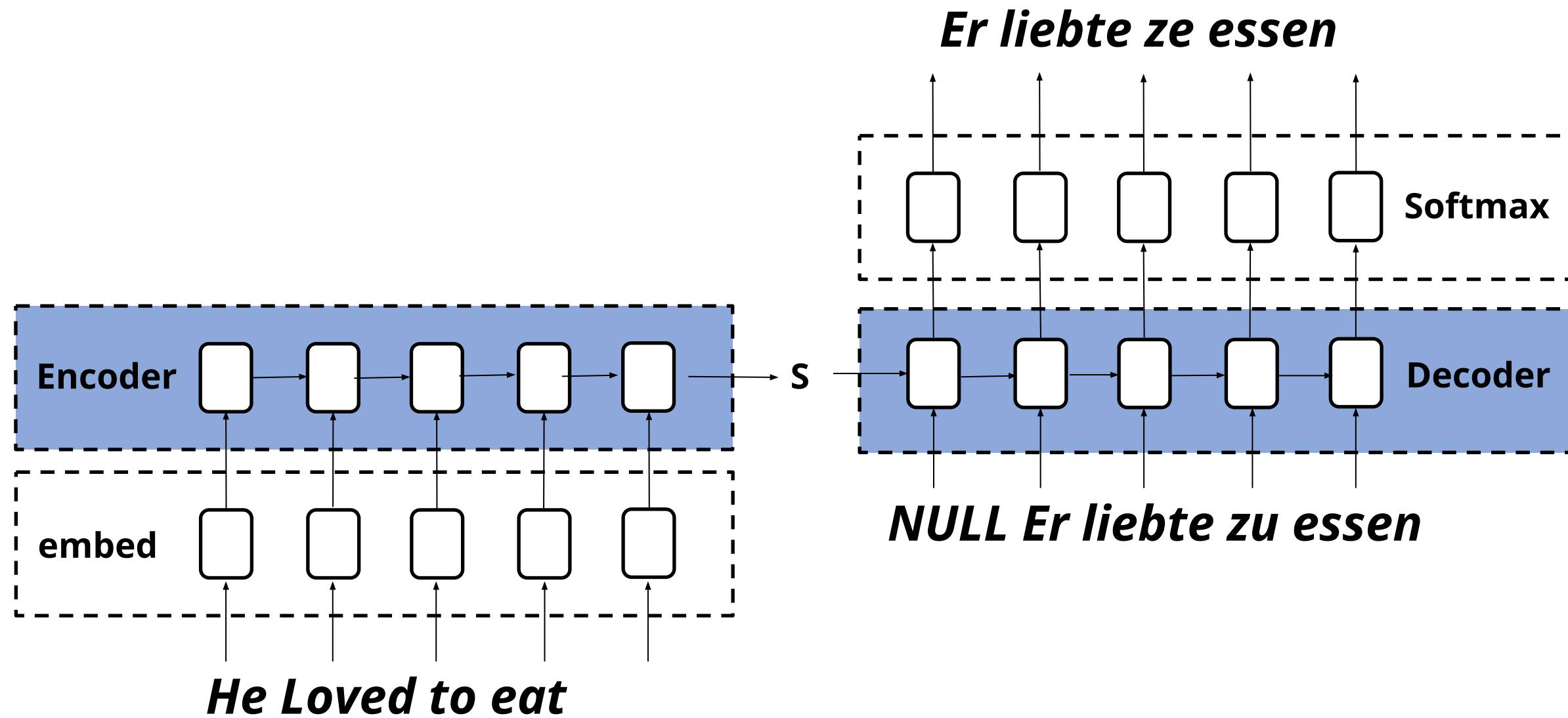
Decoder



- The Decoder takes text from the language which is to be translated into, embeds, and matches the text with the encoded English text in LSTM layer.
- The softmax function in the last layer in decoder decides the output of the decoder in translation.

Neural Machine Translation

Encoder and Decoder together as a sequence model:



seq2seq example

Music Generator

Random seed = 5

Sequence Model



Text Captioning

Random seed =10

Sequence Model

*Neuron is a cell that can
Transmit electrical signals.
They can also refer to units
In ANN model.....*

Image Generator

Random seed =15

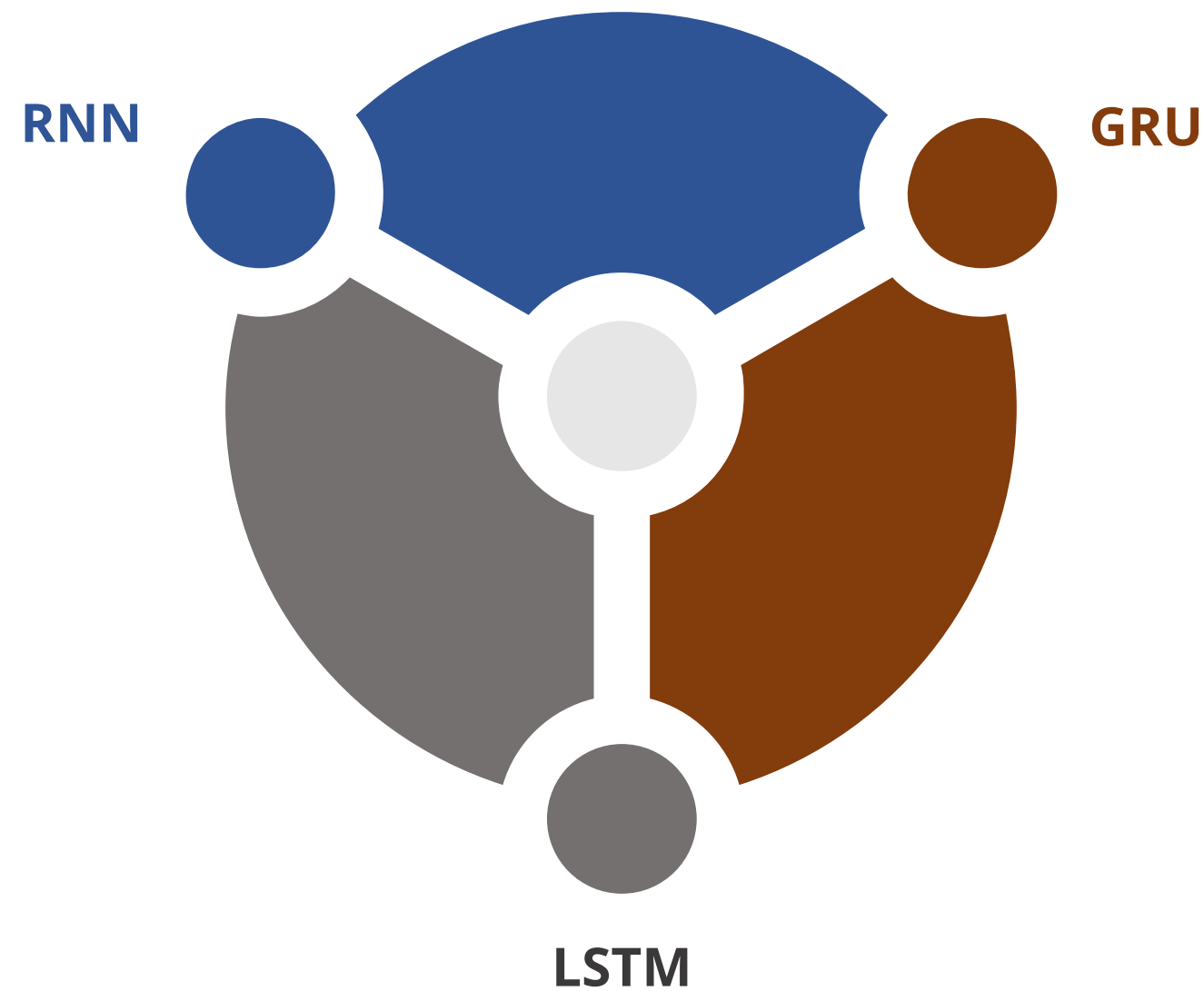
Sequence Model



Components of Encoder Decoder Architecture

Components of Encoder-Decoder Architecture

Following are the components of Encoder-Decoder architecture:



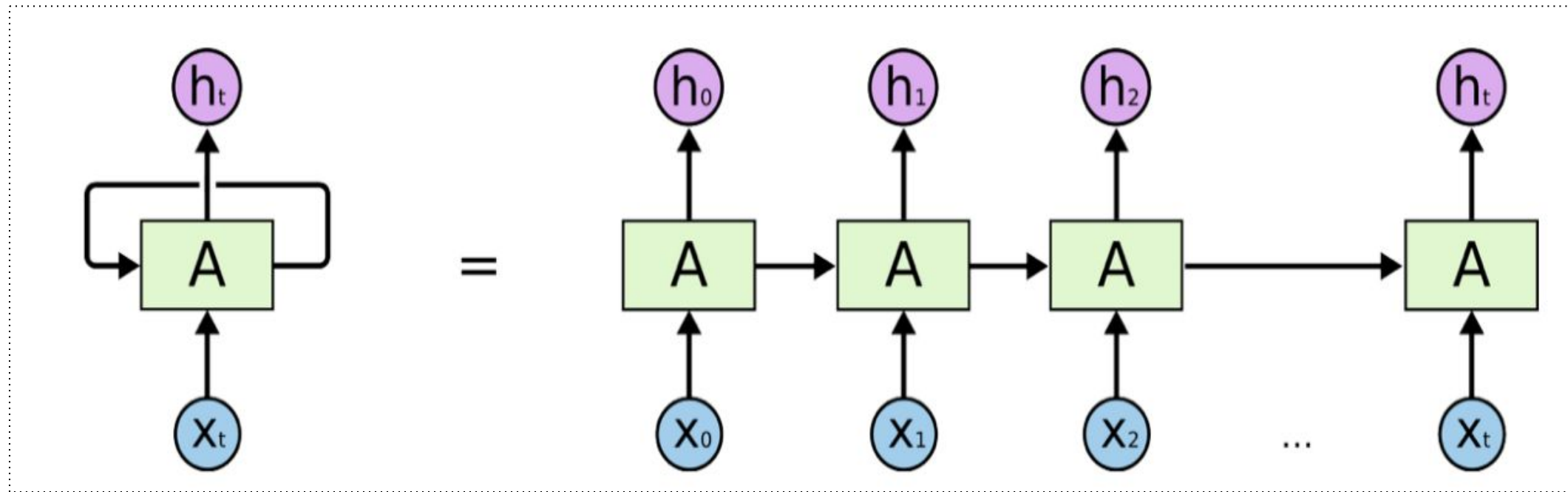
Components of Encoder-Decoder Architecture

Recurrent
Neural
Network
(RNN)

- Output from previous step are fed as input to the current step
- Hidden state of RNN remembers some information about a sequence
- Weights and losses are calculated in each layer

Components of Encoder-Decoder Architecture

Here one hidden layer shares information with the next hidden layer to understand the relation and information to produce the output.



Calculating current state:

$$h_t = f(h_{t-1}, x_t)$$

h_t is current state

h_{t-1} is previous state

x_t is input state

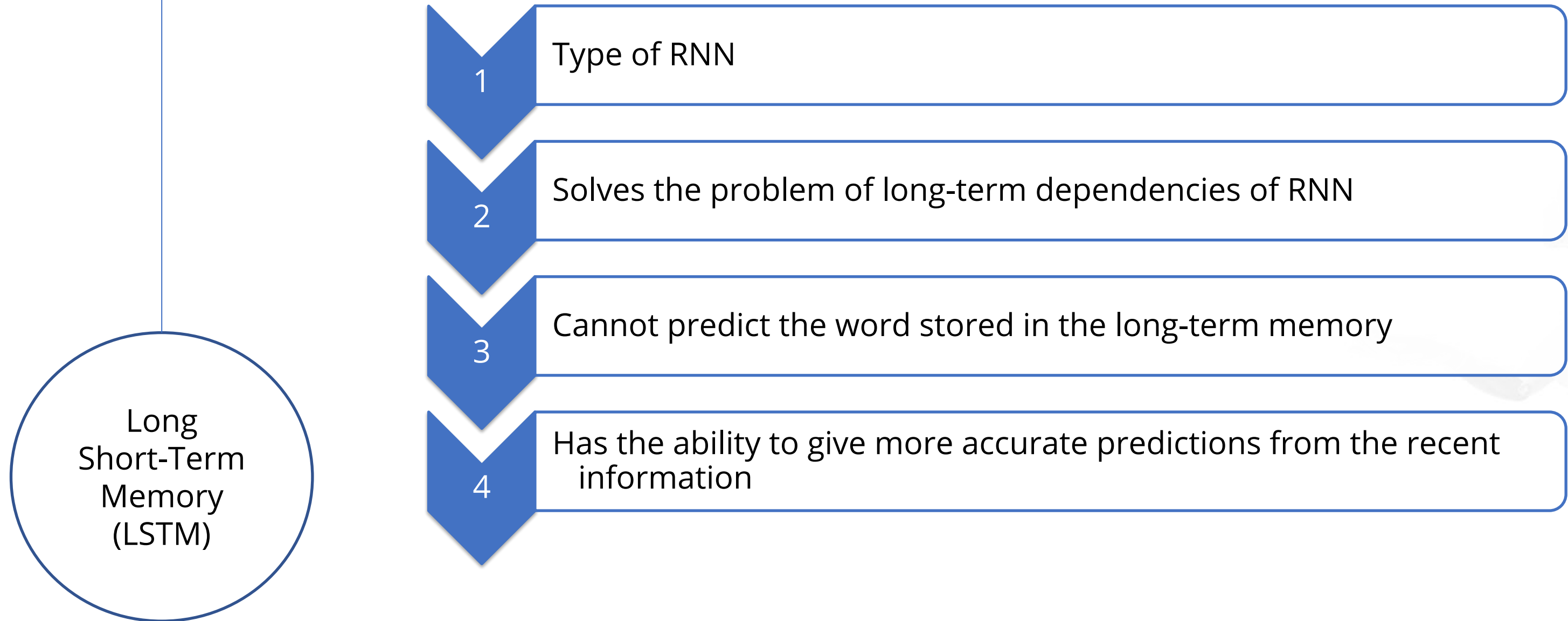
Calculating output:

$$y_t = W h_t$$

y_t is output

W is weight at output layer

Components of Encoder-Decoder Architecture

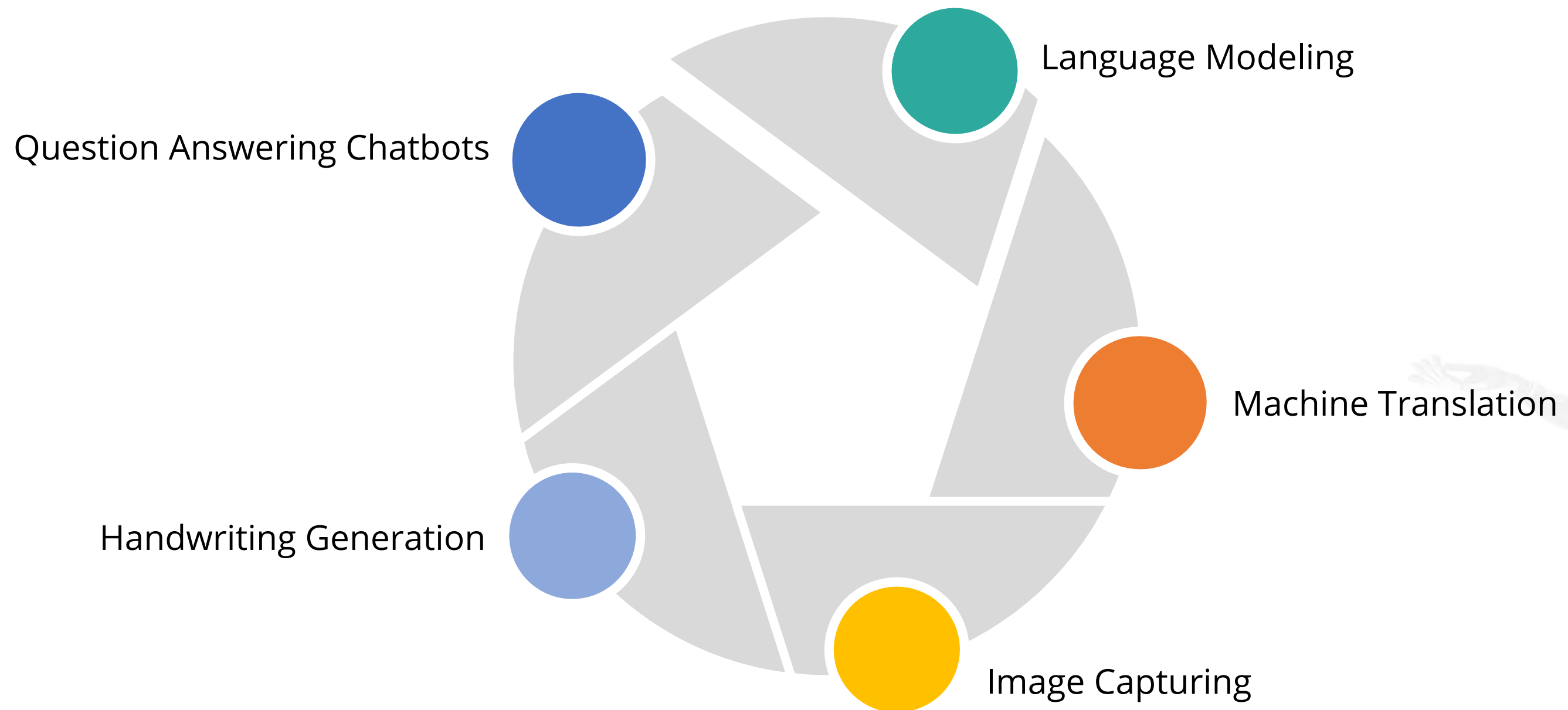


Components of Encoder-Decoder Architecture

- LSTM contains four neural networks which are arranged in a chain structure and have different memory blocks called cells.
- Information is retained by the cells and the memory manipulations are done by the gates.

Components of Encoder-Decoder Architecture

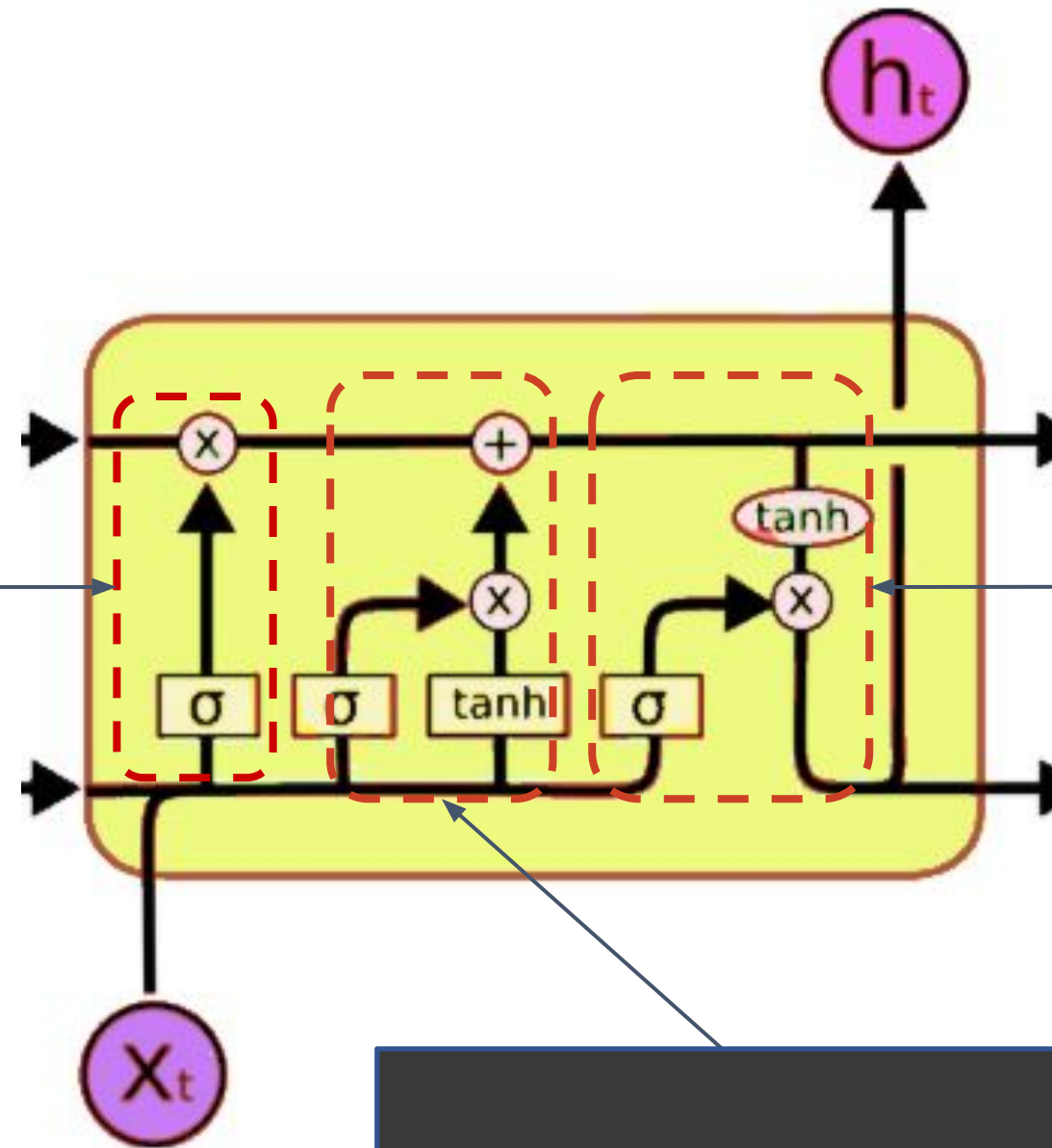
The following are the applications of LSTM:



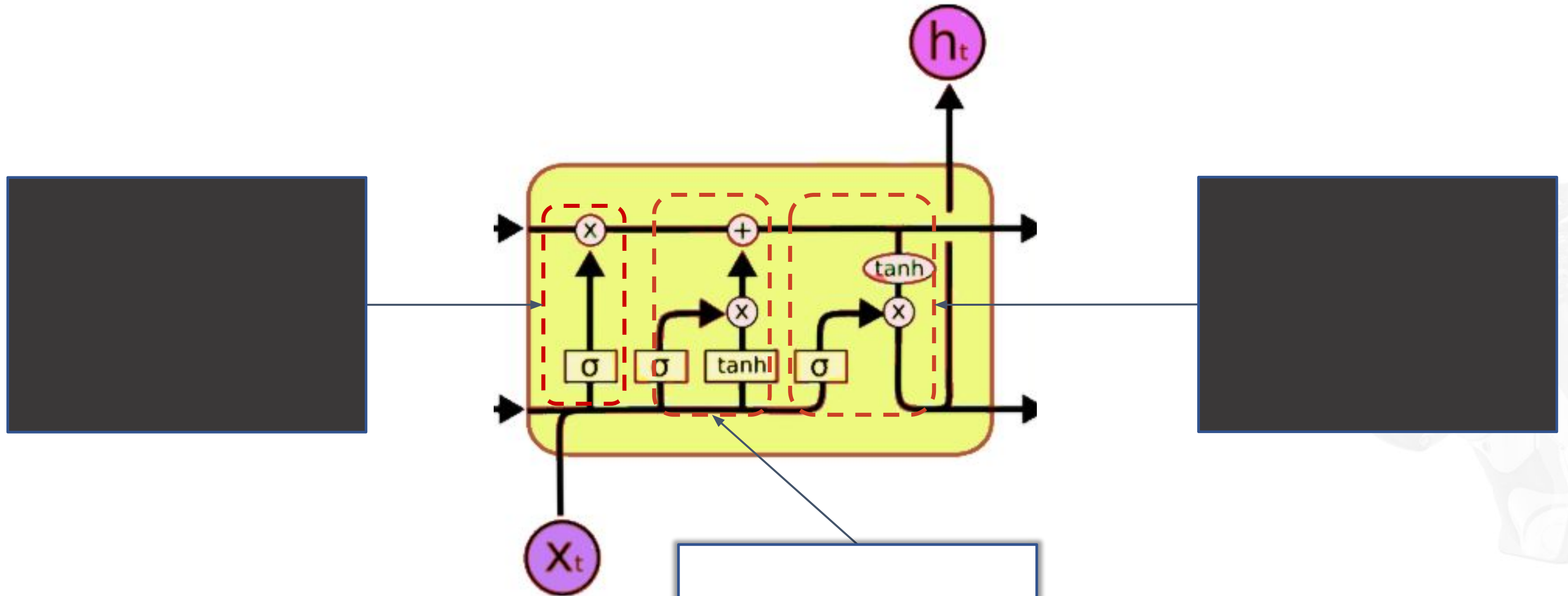
Components of Encoder-Decoder Architecture

Forget Gate:

Removes information that is no longer useful in the cell state



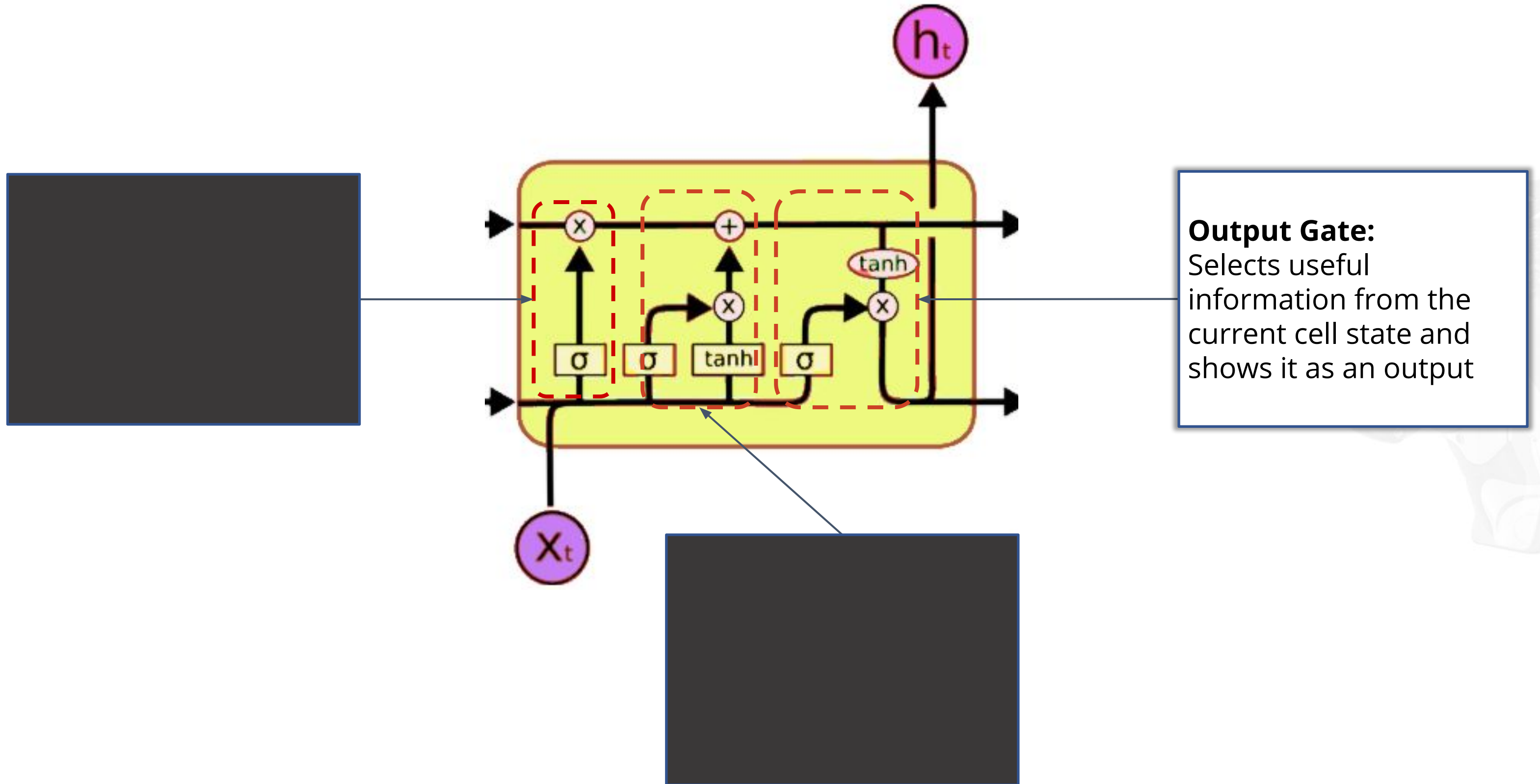
Components of Encoder-Decoder Architecture



Input Gate:

Adds some new information in the cell state

Components of Encoder-Decoder Architecture



Components of Encoder-Decoder Architecture

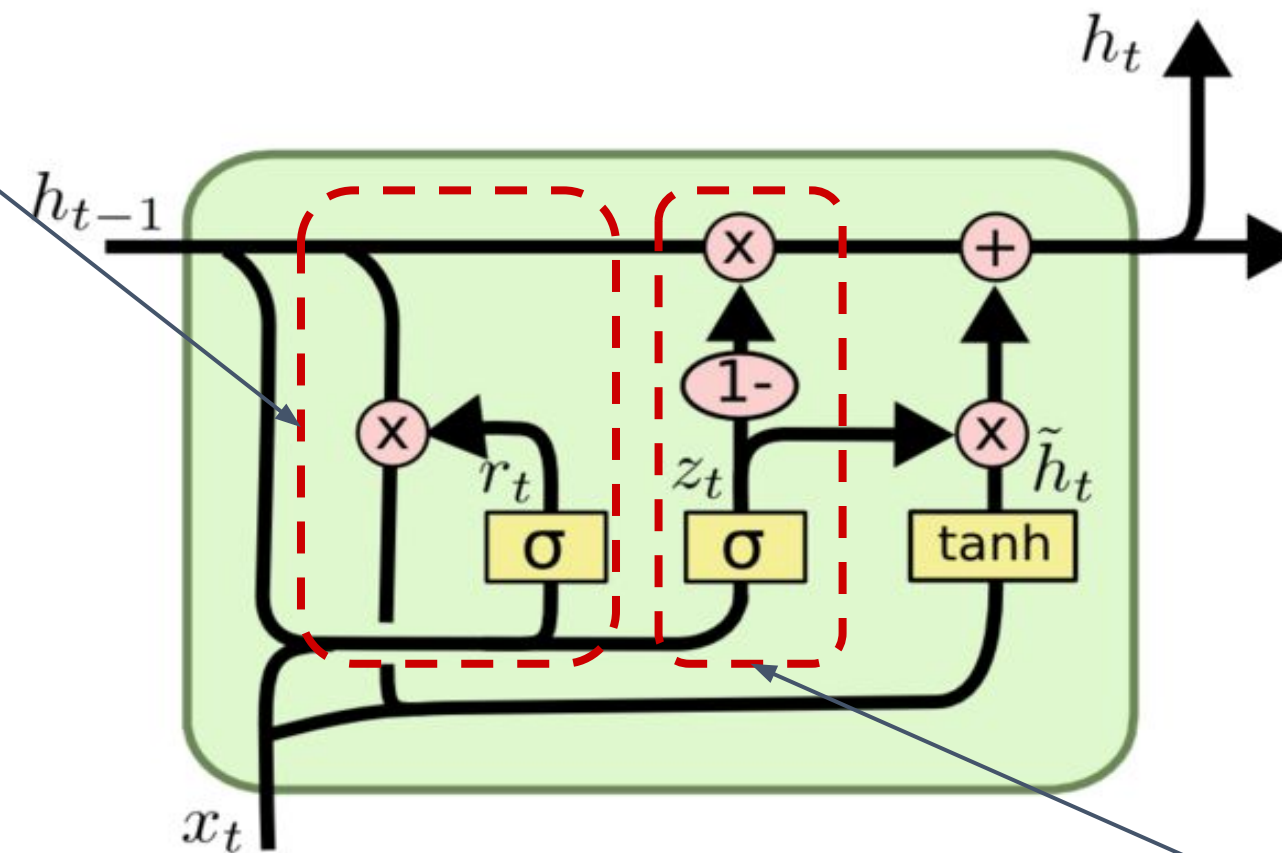
Gated
Recurrent
Unit (GRU)

- GRU can be considered as a variation of the LSTM because it is designed similarly
- It solves the vanishing gradient problem of a standard RNN
- It uses two gates

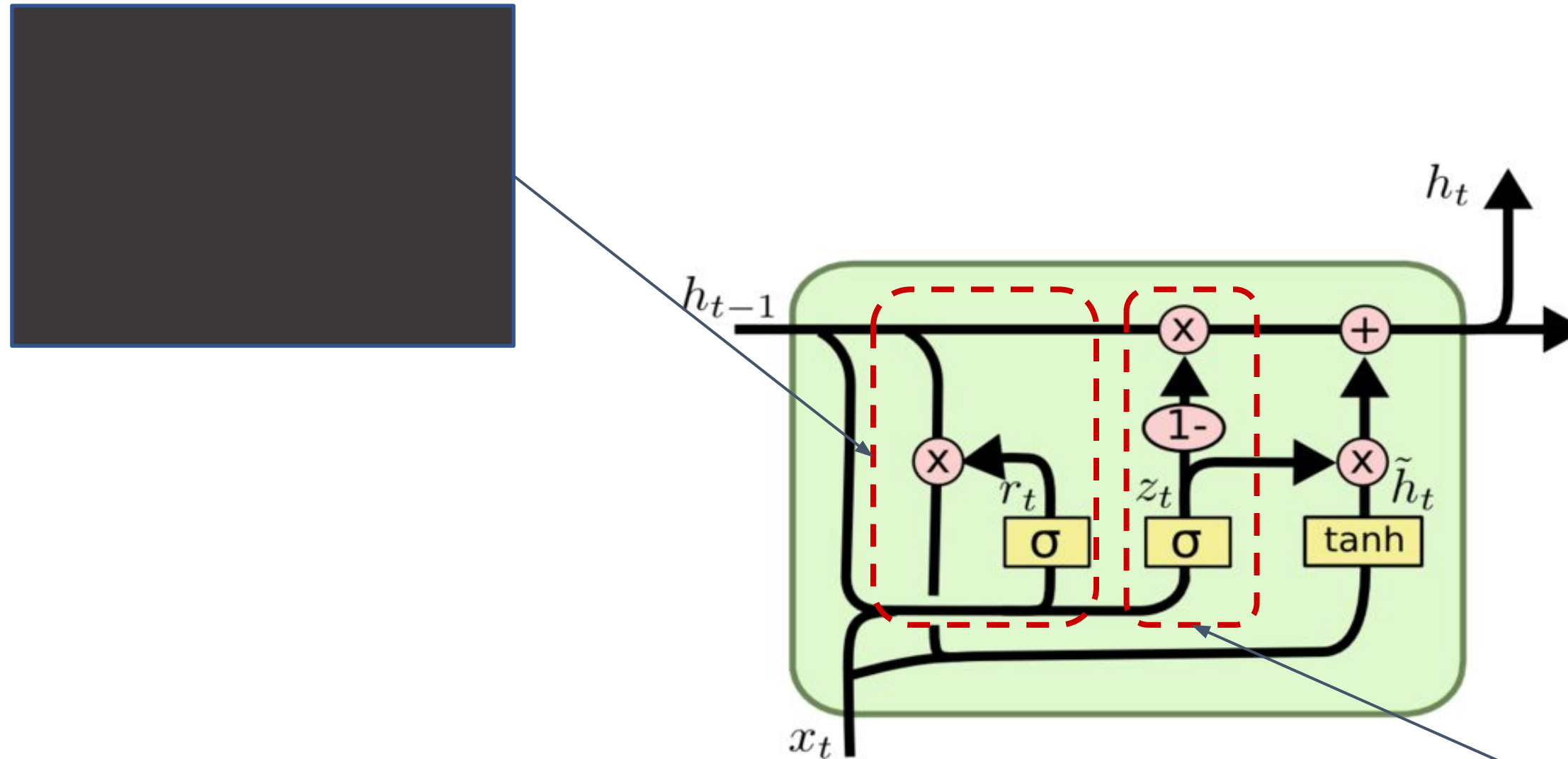
Components of Encoder-Decoder Architecture

Reset Gate:

Used by the model to determine how much of the past information to forget



Components of Encoder-Decoder Architecture

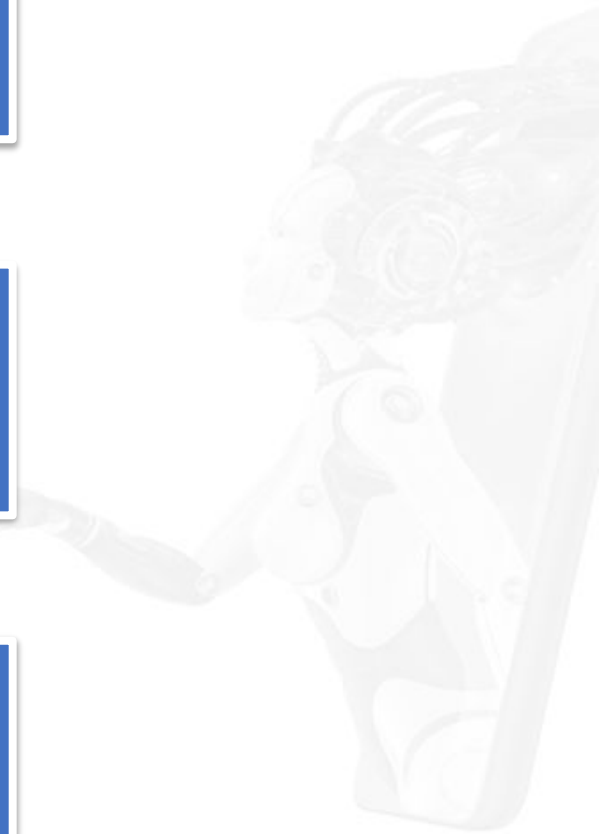
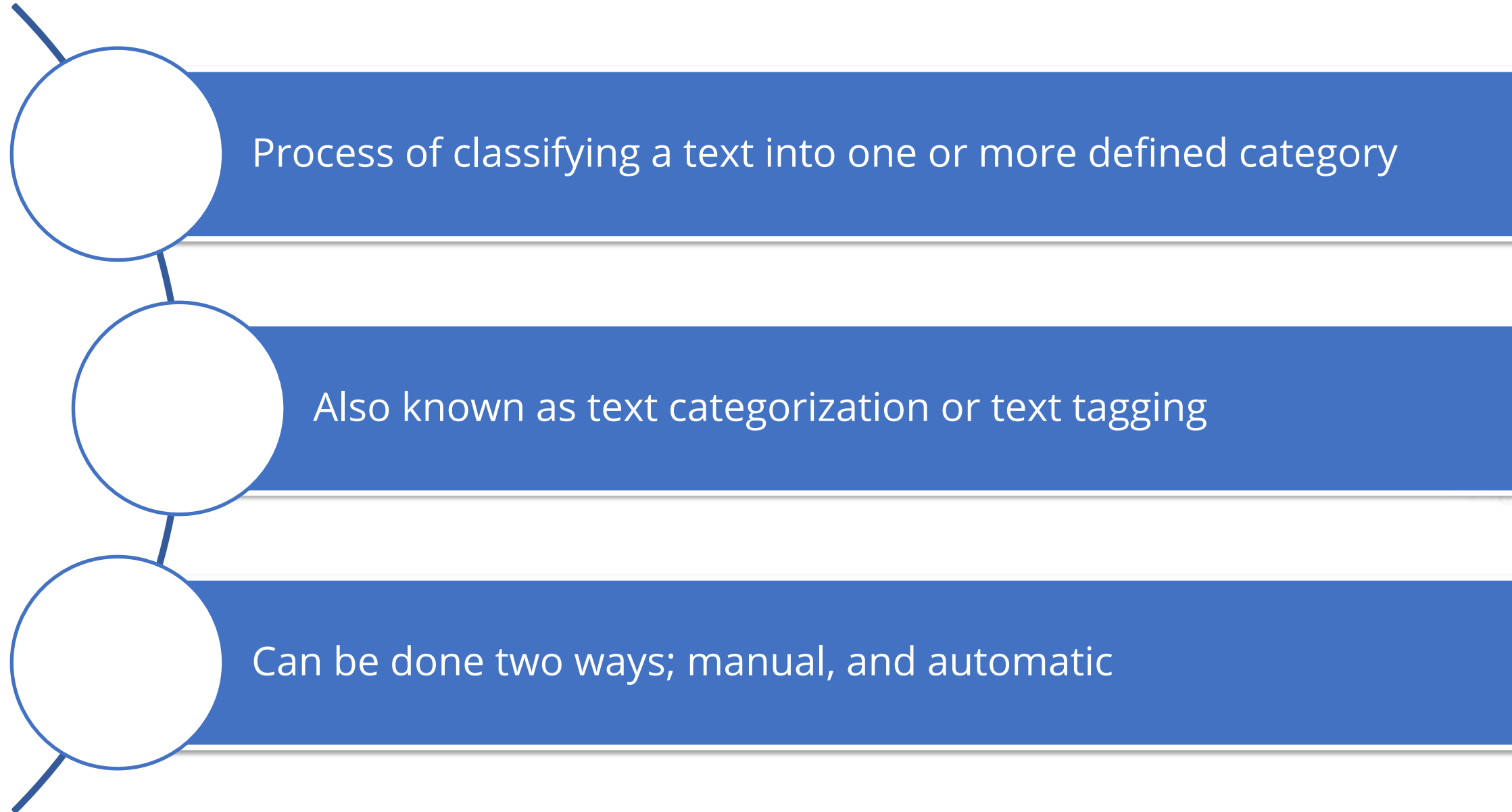


Updated Gate:

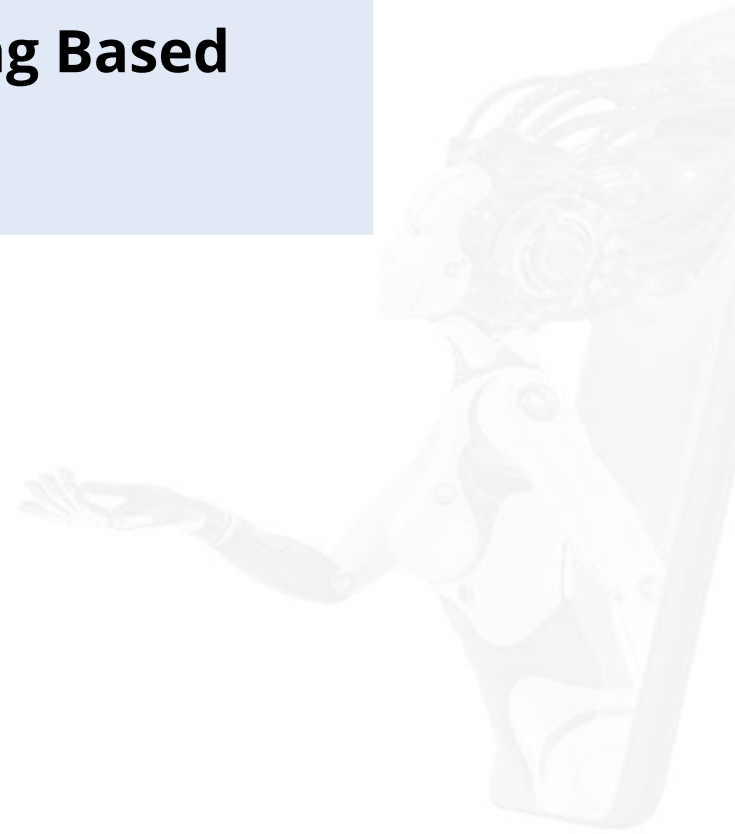
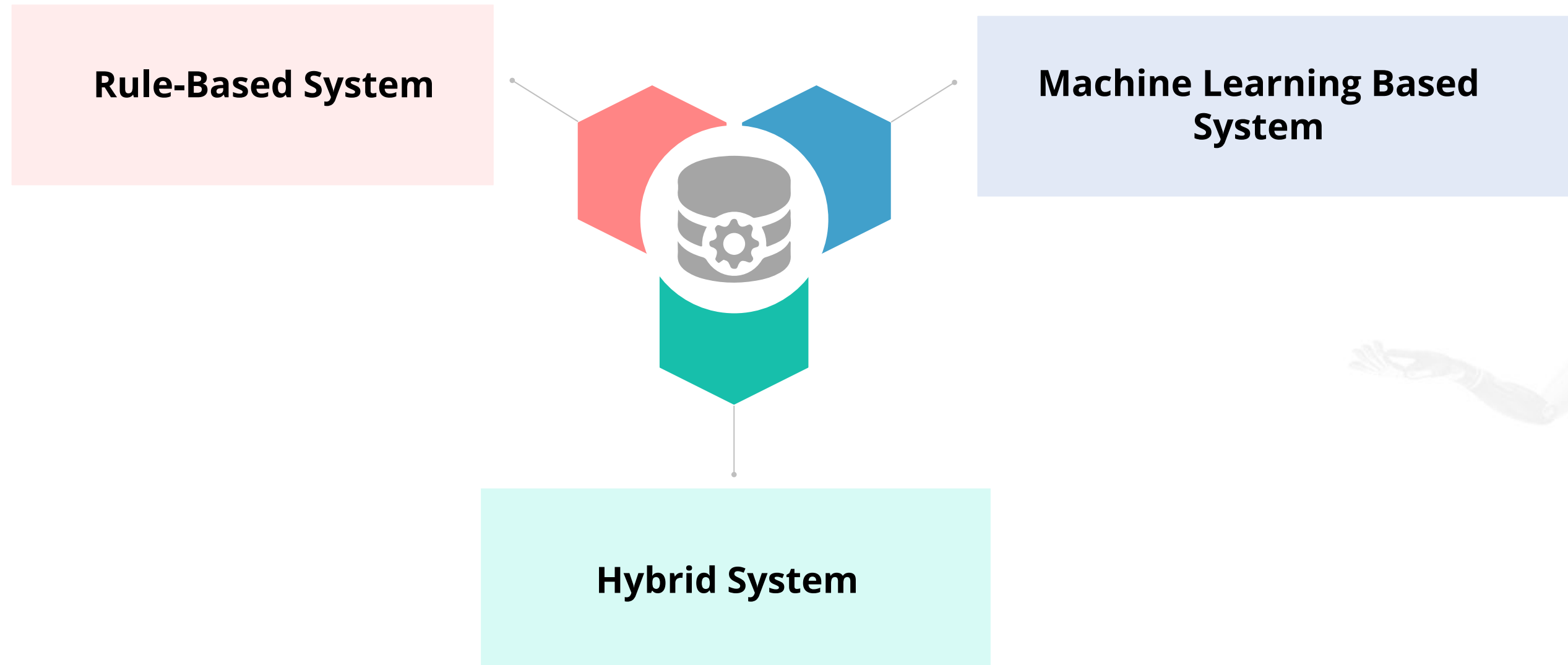
Helps the model to determine how much of the past information needs to be passed along to the future

Text Classification

Text Classification: Introduction



Text Classification: Approach



Text Classification: Approaches

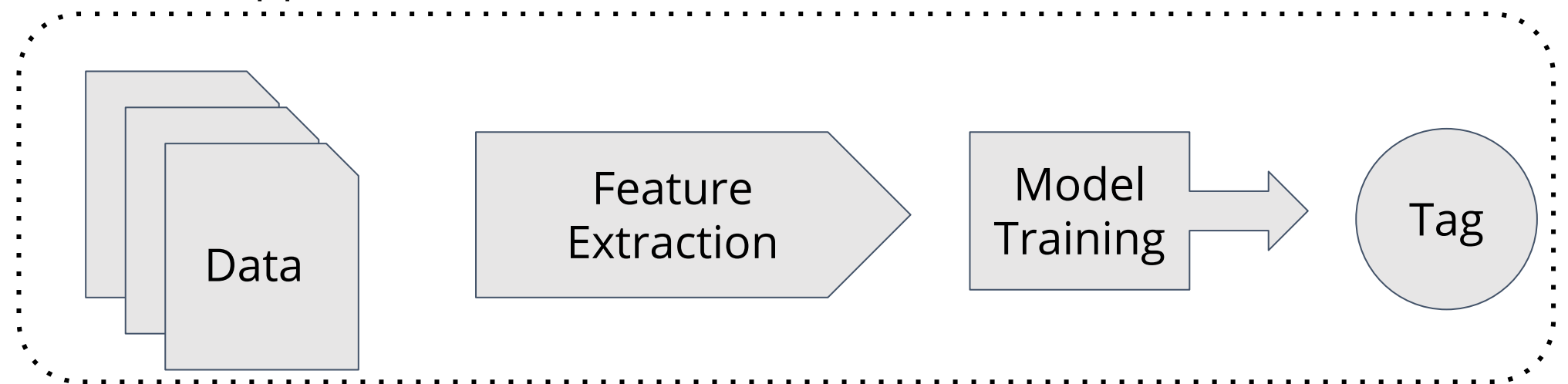
Rule-Based System

- Classifies texts into groups by using predefined handcrafted rules
- Can be improved over time, but it is time consuming
- Requires domain knowledge to build the rules
- Maintainability is tough

Text Classification: Approaches

Machine Learning Based
System

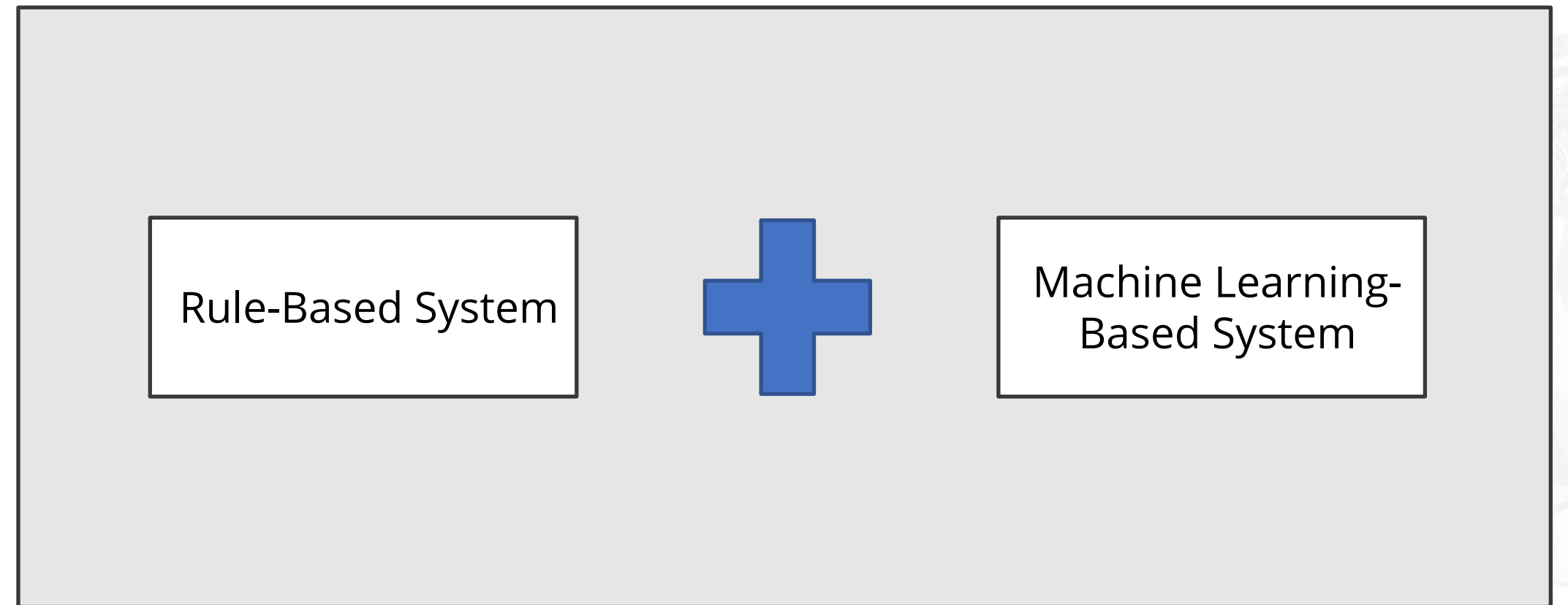
- Classifies the text based on past observations.
- Text classification algorithms:
 - Naive Bayes
 - Deep Learning
 - Support Vector Machine



Text Classification: Approaches



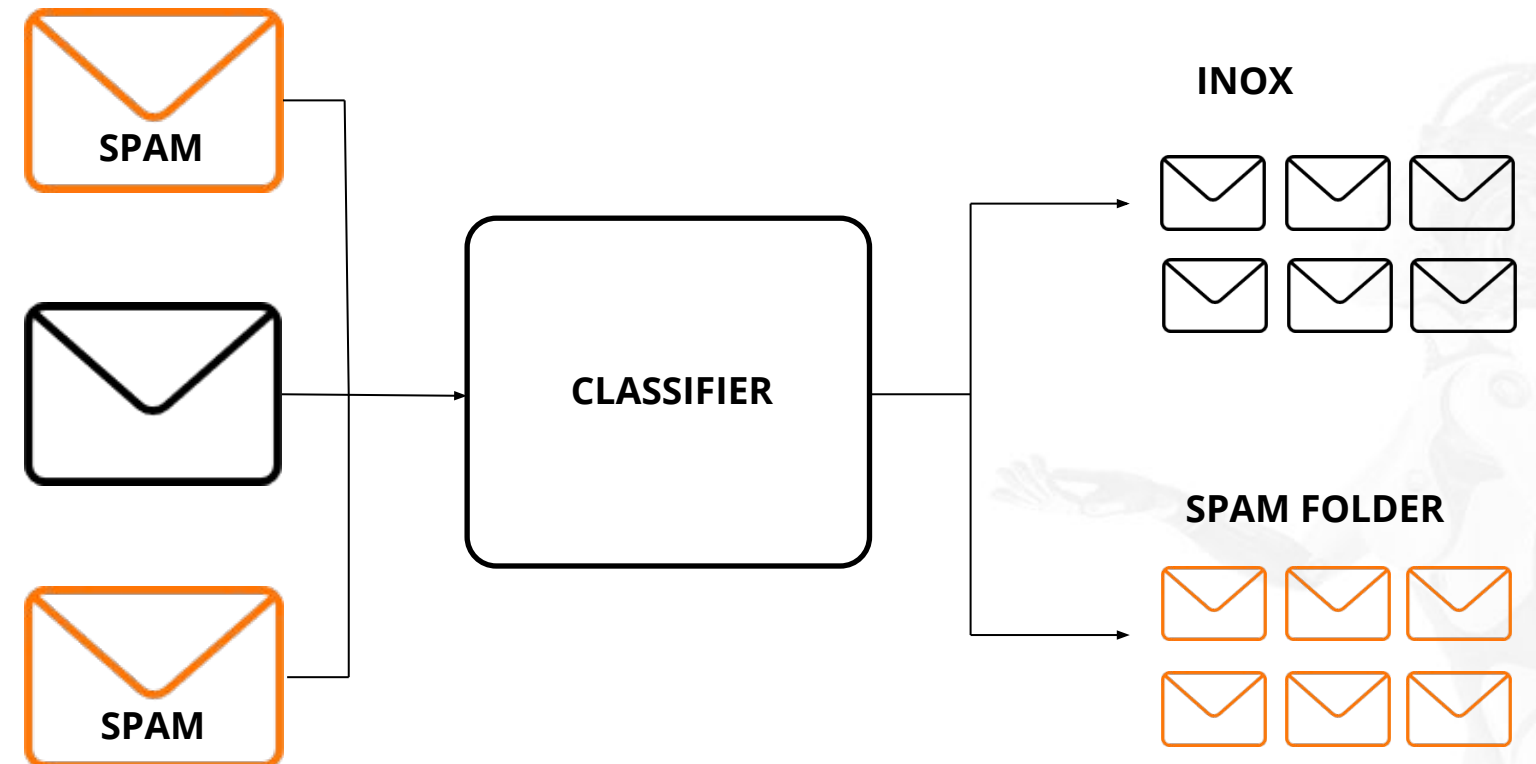
Hybrid System



Text Classification: Example

Spam-Ham Detection

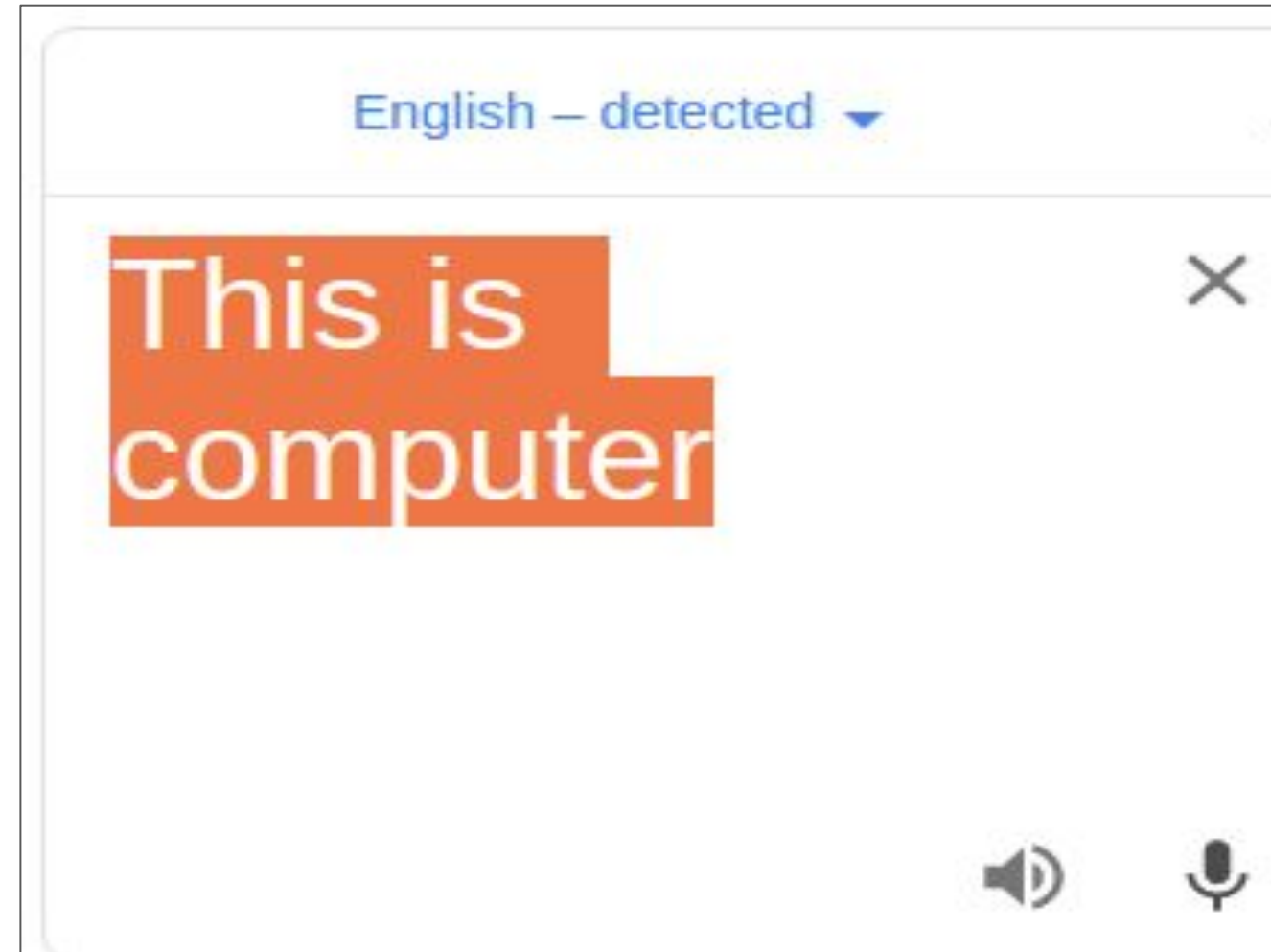
1. Email is the input data.
2. Once user receives the email, it is classified as spam or ham.
3. Spam is moved into junk folder. Ham sits in the inbox.



Text Classification: Example

Language Detector

Example: Google Language Detector



Text Classification: Example

Query Classification for Information Retrieval

I want to purchase Laptop.

Electronics

Track my parcel.

Logistics

Text Summarization

Text Summarization: Introduction

Text summarization is a technique to make the long piece of information, small and concise. The subset of data identified represents the crux of all contents.

Following are the two approaches for automatic text summarization:

Extraction

Abstraction

Text Summarization: Introduction

Extraction

It selects some portion from the original text like words and phrase.

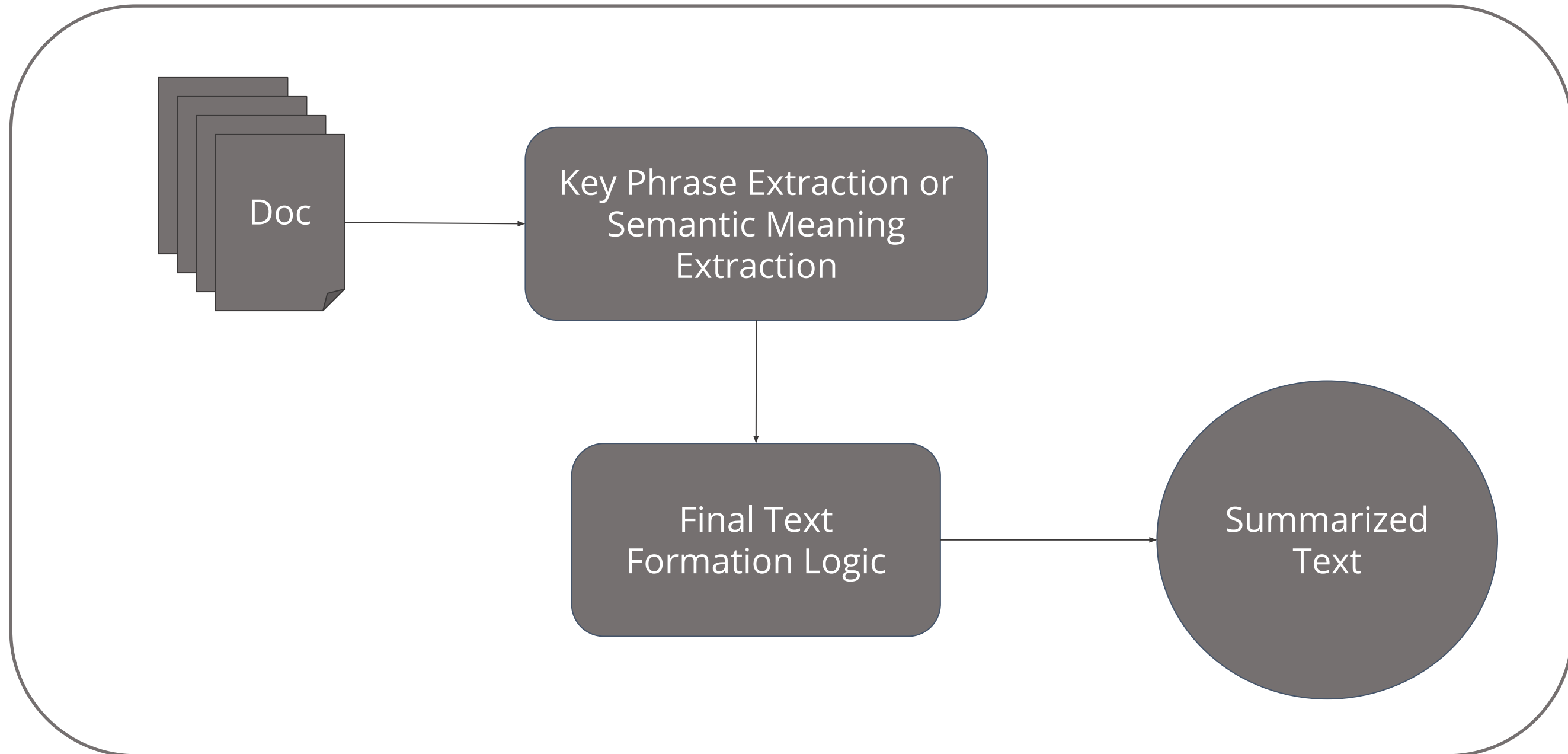
Abstraction

It gives you the semantic representation of content and uses Natural Language Generation technique to produce the final text.



Text Summarization: Working

Text summarization is considered as a supervised machine learning problem.



Text Summarization: Example



Document Clustering

Document Clustering: Introduction

1

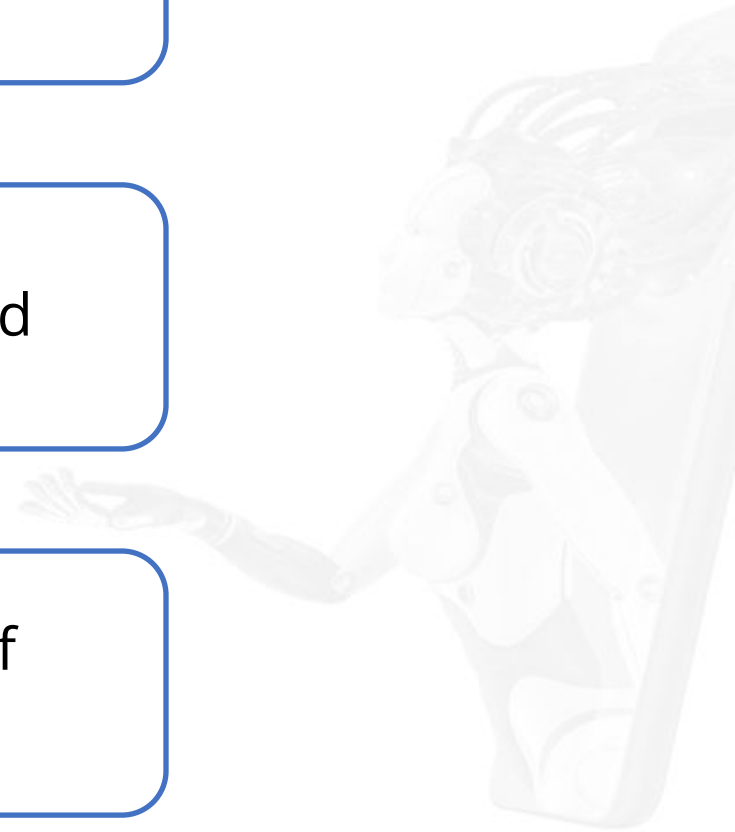
Document clustering is the grouping of same kind of documents in one or more cluster

2

The number of clusters can be defined, or it is automatically picked

3

It is an unsupervised process used to find structure in collection of unlabeled data



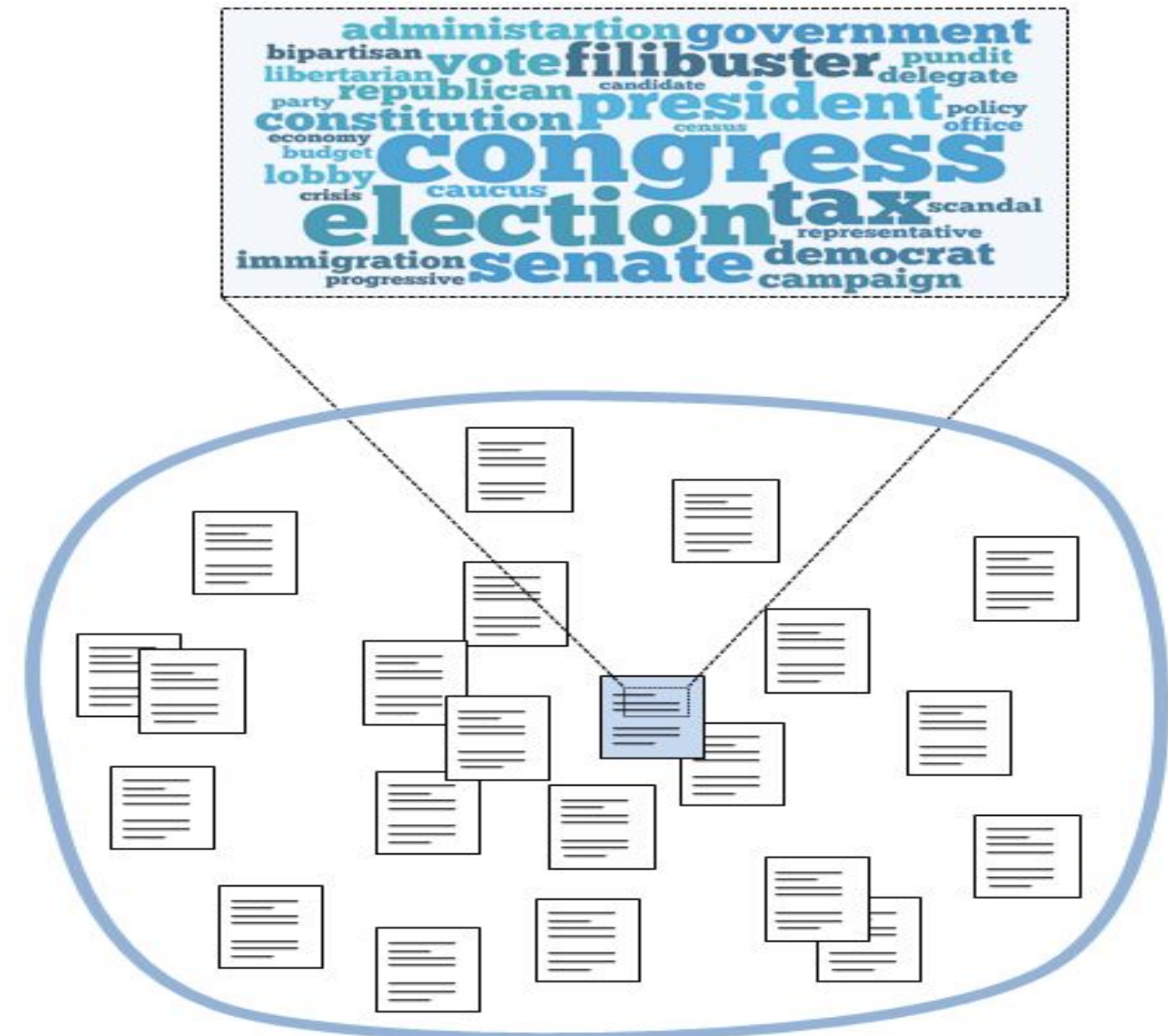
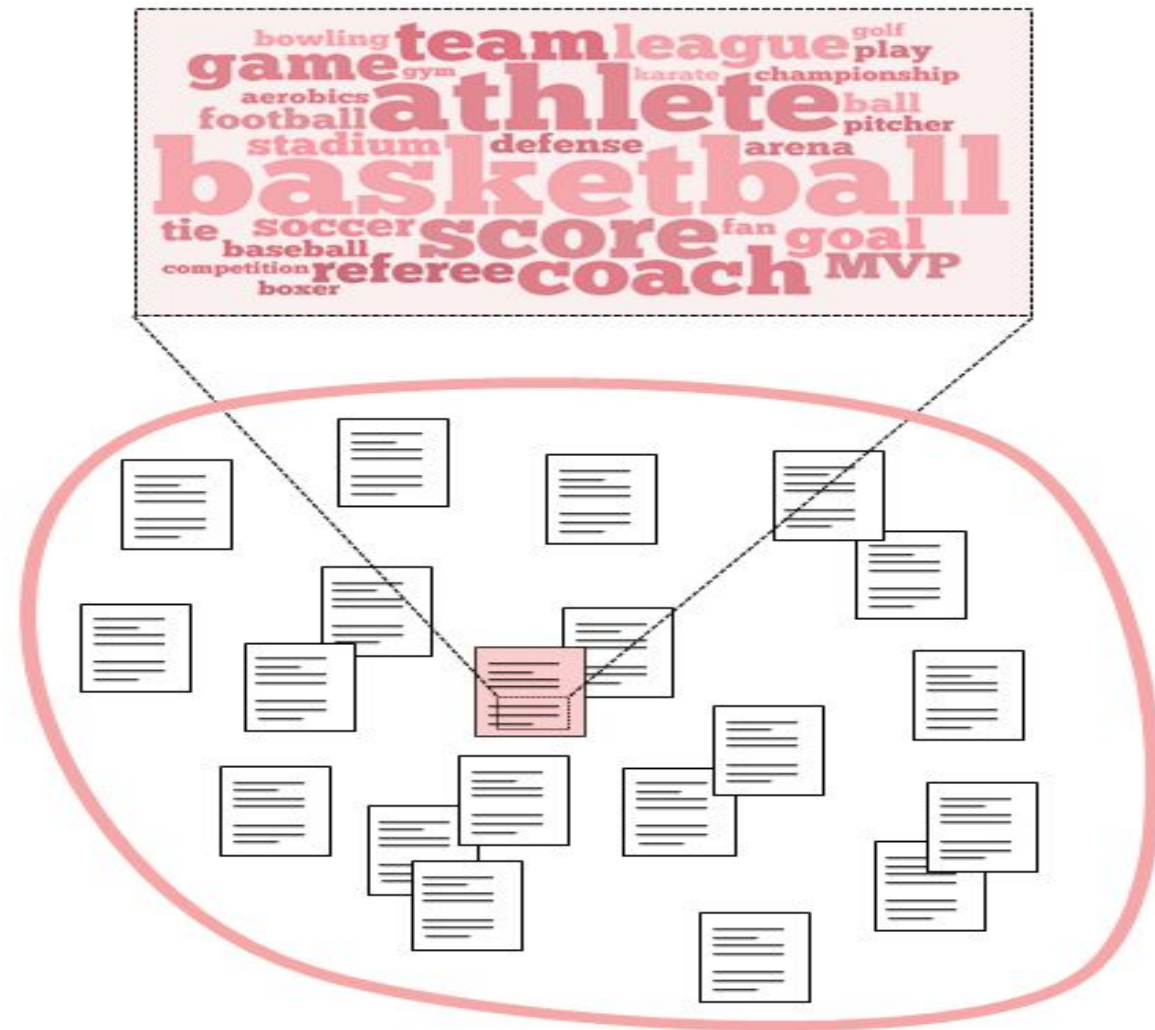
Document Clustering: Working

The following are the steps for clustering:

- 1 Remove the stop words, perform stemming, and tokenization.
- 2 Apply TF-IDF on text preprocessed in first step. TF-IDF will give importance to words occurring frequently in document but not that frequently in corpus.
- 3 Use clustering algorithm (k-means) on output of TF-IDF. Algorithm will have centroids in each cluster, which are far apart from each other.
- 4 Assign centroid as a topic for the documents.

Document Clustering: Example

Documents are clustered into two and here it is seen as document from each cluster having important keywords.

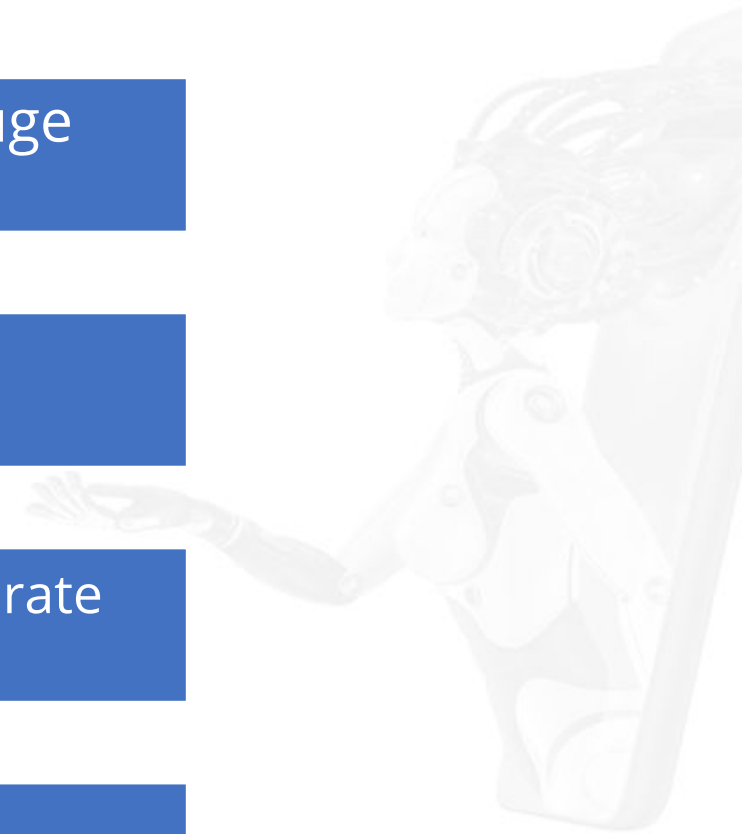


Source

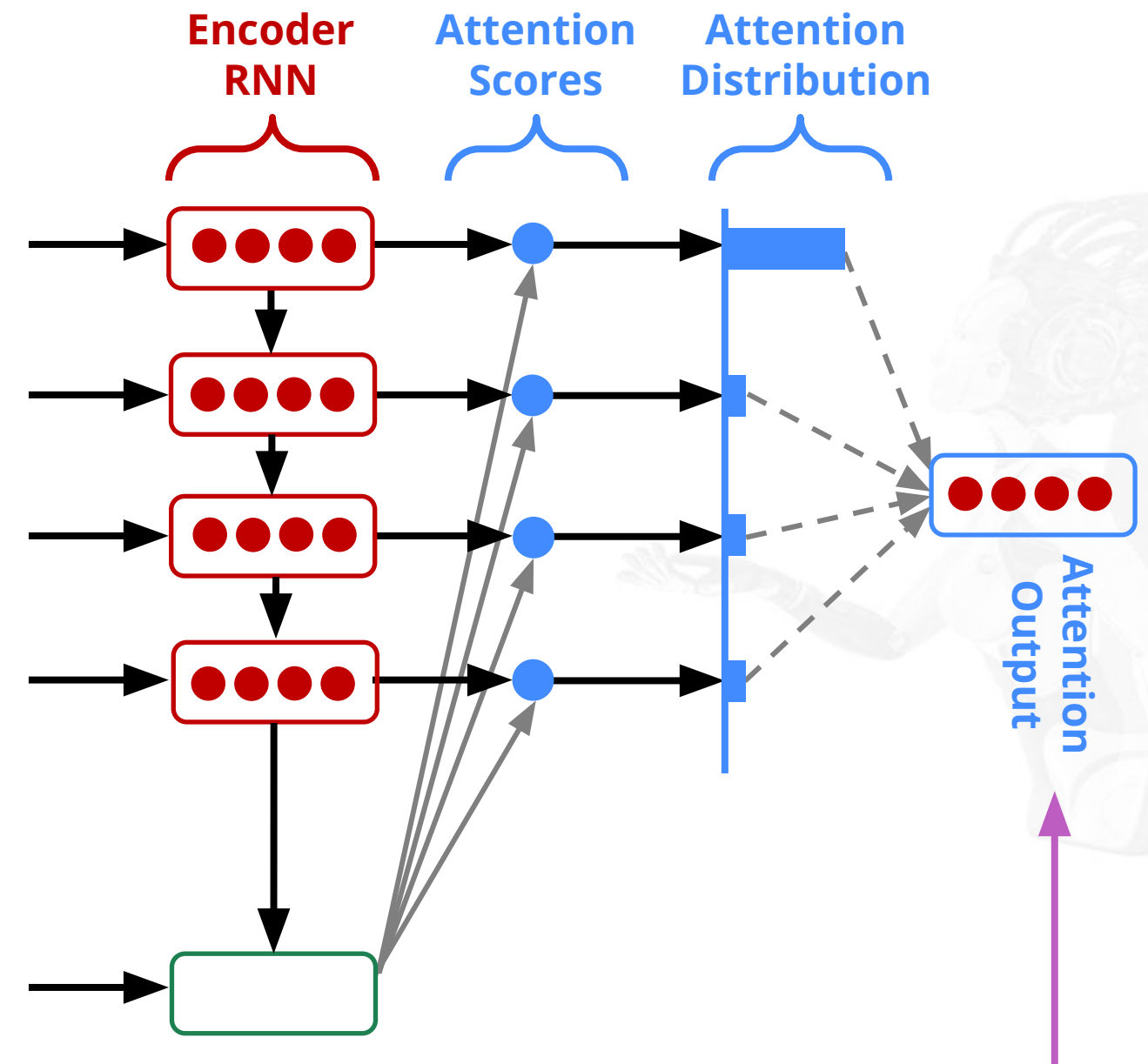
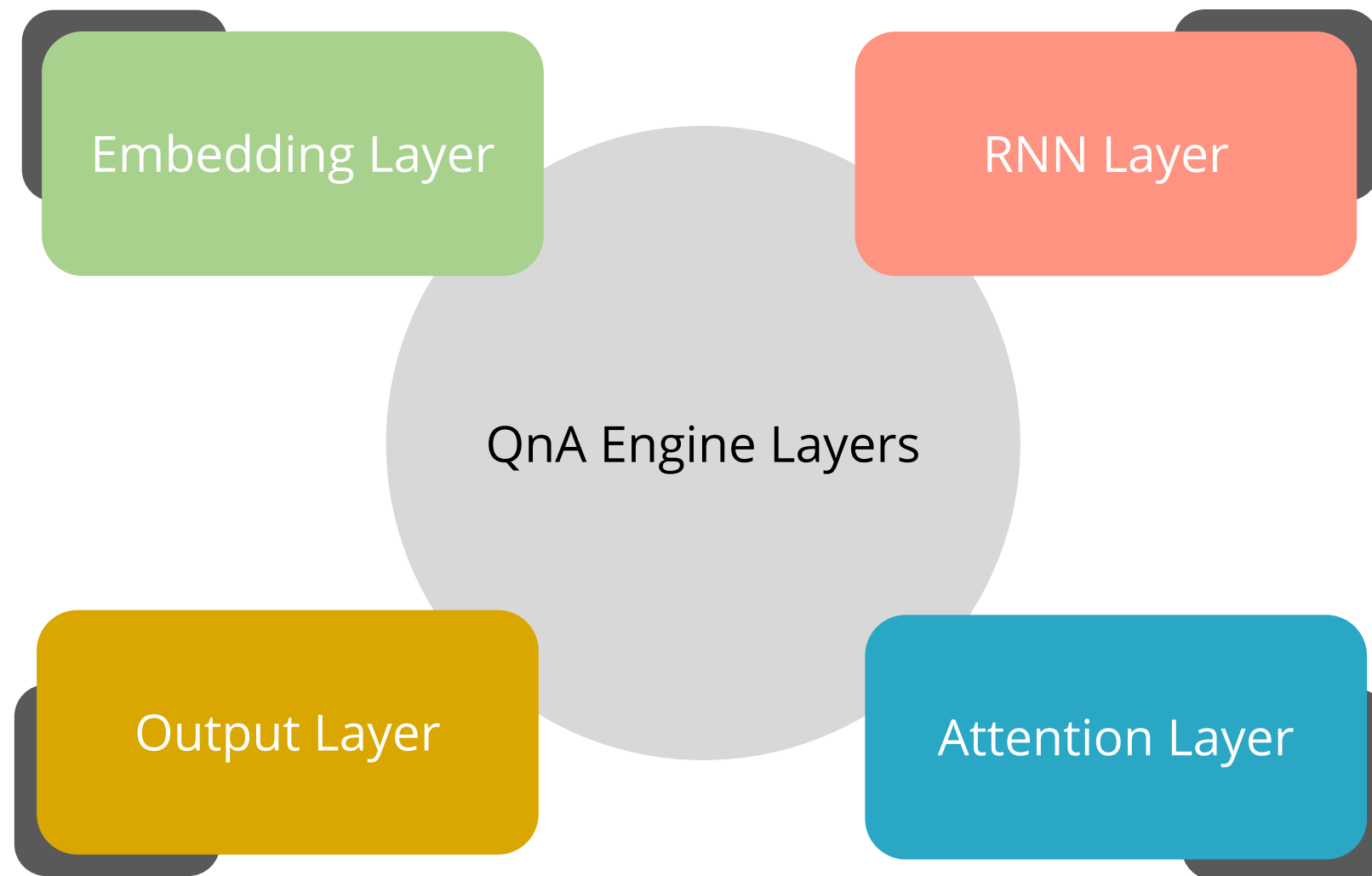
Question Answering Engine

Question Answering Engine: Introduction

- QnA engine is the system used to cater the responses for user's query.
- It was initially used for writing some rules but, when there is huge data, writing rule is not beneficial.
- The attention mechanism is helpful in QnA engine.
- The Question Answering Model will use attention mechanism to generate specific answer to question.
- It is neural network-based system with a focus on required information.



Question Answering Engine: Layers



Question Answering Engine: Layers

Embedding Layer

Context and corresponding questions lie within the training dataset for the model that can be broken into individual words. Then, these words can be converted into word embeddings using the pretrained vectors like GloVe.

Attention Layer

We have a hidden vector each for question and context. We need to look at them together in order to figure out the answer. This is where attention comes in. It decides, given the question, which words in the context should it “attend” to.



Question Answering Engine: Layers

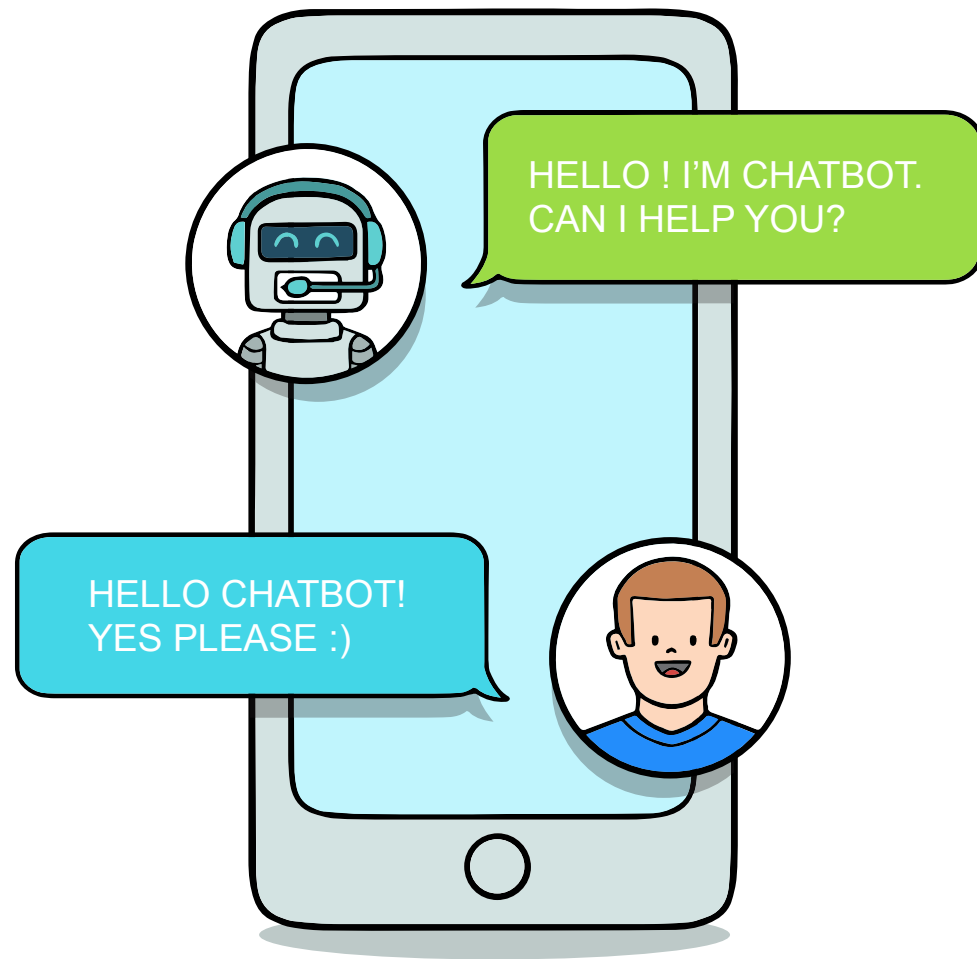
RNN Layer

A bidirectional GRU or LSTM helps to be aware of words before and after it.

Output Layer

Softmax output layer is the final layer of the model that helps to choose the start and end index for the answer span. It combines the context hidden states and the attention vector from the previous layer to create a blended response.

Question Answering Engine: Example



What's the name of Anakin Skywalker?

Darth Vader

Who destroyed the Death Star?

Luke Skywalker

Source

Spam-Ham Classification using Machine Learning



Problem Statement: Nowadays, digital contents have increased and people are moving toward internet. They communicate and share the information using emails. Companies target the audience to perform campaign and sell their products. Some people use this medium to commit fraud by sending false information through email. The task is to create a ML-based model to classify email text into spam and ham.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

Summarization of News



Problem Statement: Sometime during news reading, we must face very long news but due to lack of time, a person only focuses on certain important points. Create a script to summarize a news by finding only important information.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

Document Clustering for BBC News



Objective: To perform document clustering in order to assign similar articles under a single cluster label.

Problem Scenario: You are given a zip file which contains summaries of news from BBC. The data is taken from Kaggle.
(<https://www.kaggle.com/pariza/bbc-news-summary>).

The zip file contains a folder: BBC News Articles. This folder contains 5 sub-folders, named:

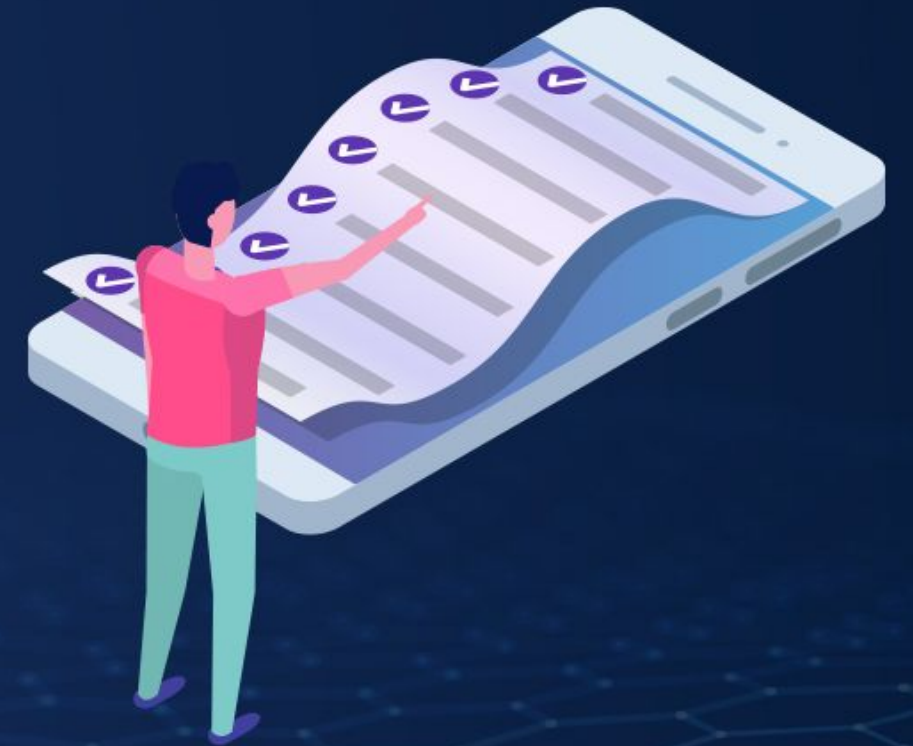
- Business
- Entertainment
- Politics
- Sports
- Tech

Each of these sub-folders contains text files which have summaries of different news articles.

Key Takeaways

You are now able to:

- 🕒 Explain Neural Machine Translation
- 🕒 Define Text Classification and Text Summarization
- 🕒 Explain document clustering, attention mechanism, and question answering engine
- 🕒 Demonstrate spam-ham classification using ML
- 🕒 Create a script to summarize a news y finding only important information



DATA AND ARTIFICIAL INTELLIGENCE



Knowledge Check

Knowledge Check

1

Which activation function is suitable for multiclass text classification in output layer?

- a. Sigmoid
- b. Relu
- c. Softmax
- d. Linear



Knowledge Check

1

Which activation function is suitable for multiclass text classification in output layer?

- a. Sigmoid
- b. Relu
- c. Softmax
- d. Linear



The correct answer is **c.**

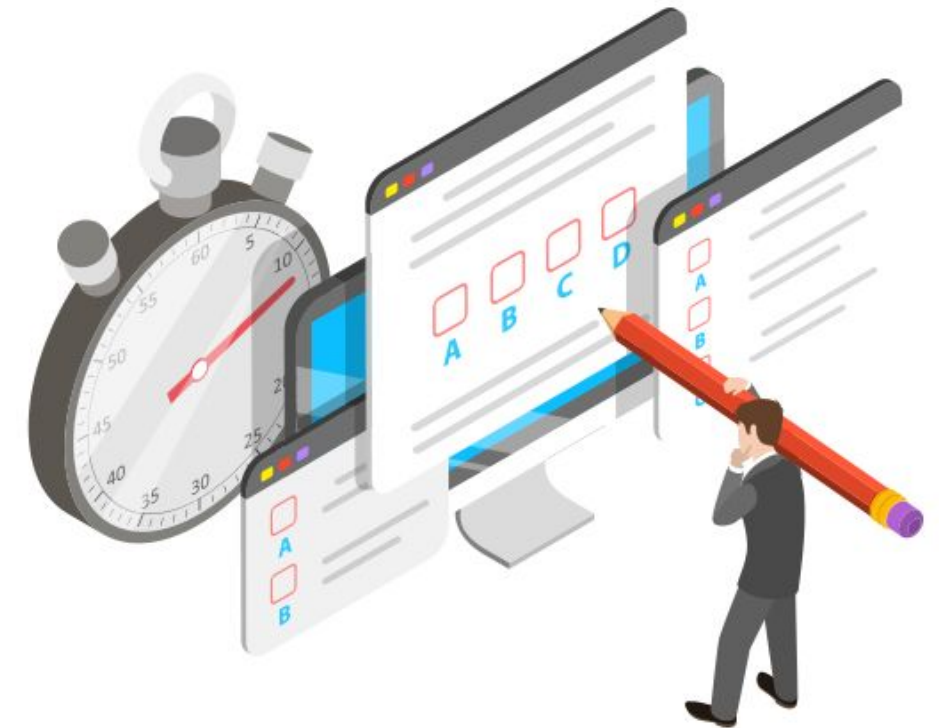
Softmax is suitable for multiclass text classification in output layer.

Knowledge Check

2

What are the applications of sequence-to-sequence generator?

- a. Machine Translation
- b. Question Answering
- c. Automatic Text Generation
- d. All of the above

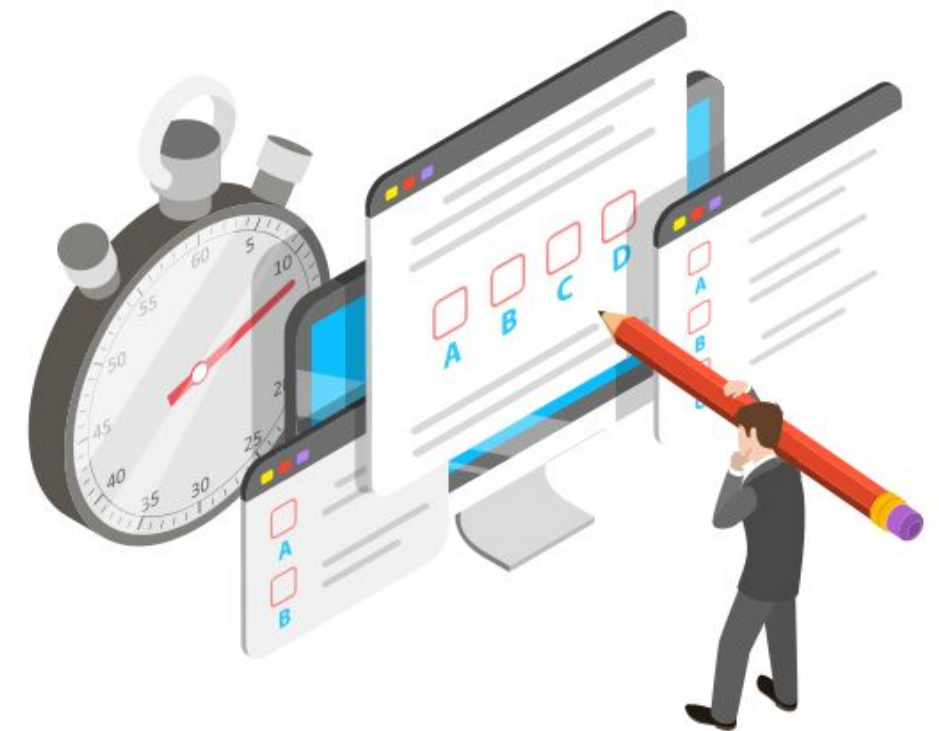


Knowledge Check

2

What are the applications of sequence-to-sequence generator?

- a. Machine Translation
- b. Question Answering
- c. Automatic Text Generation
- d. All of the above



The correct answer is **d.**

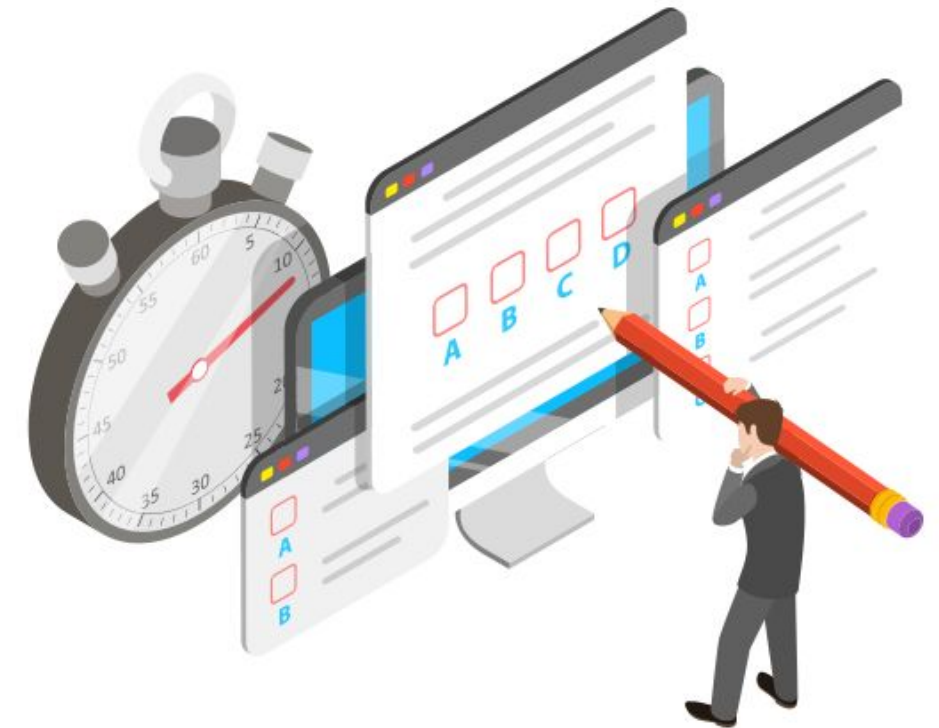
Machine Translation, Question Answering, and Automatic Text Generation are the applications of sequence-to-sequence generator.

Knowledge Check

3

Information to forget is decided by _____ in GRU.

- a. Forget Gate
- b. Update Gate
- c. Reset Gate
- d. Output Gate

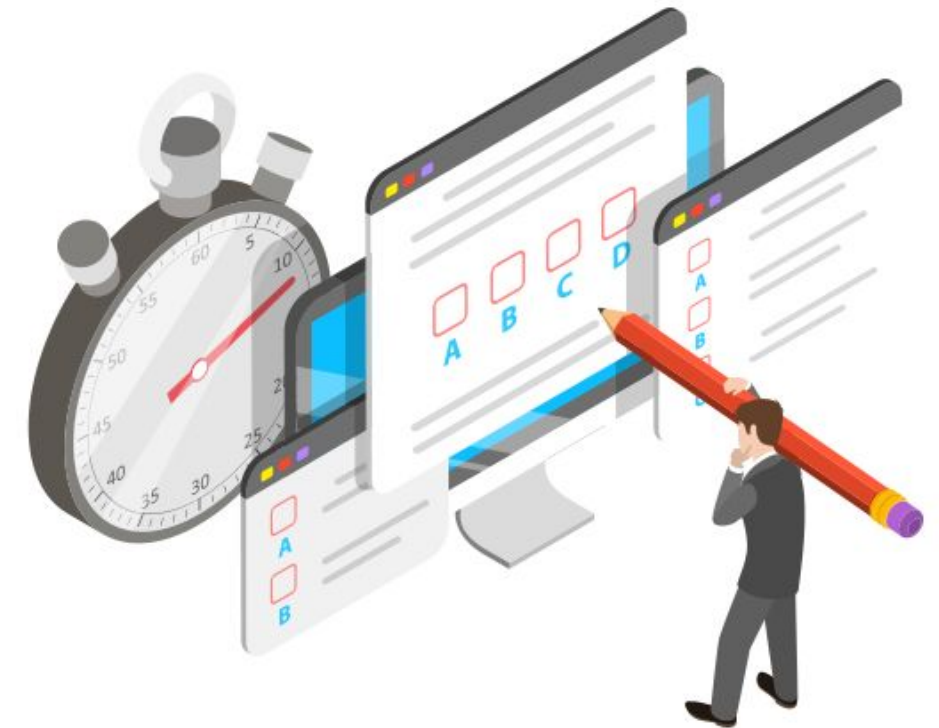


**Knowledge
Check**

3

Information to forget is decided by _____ in GRU.

- a. Forget Gate
- b. Update Gate
- c. Reset Gate
- d. Output Gate



The correct answer is **c.**

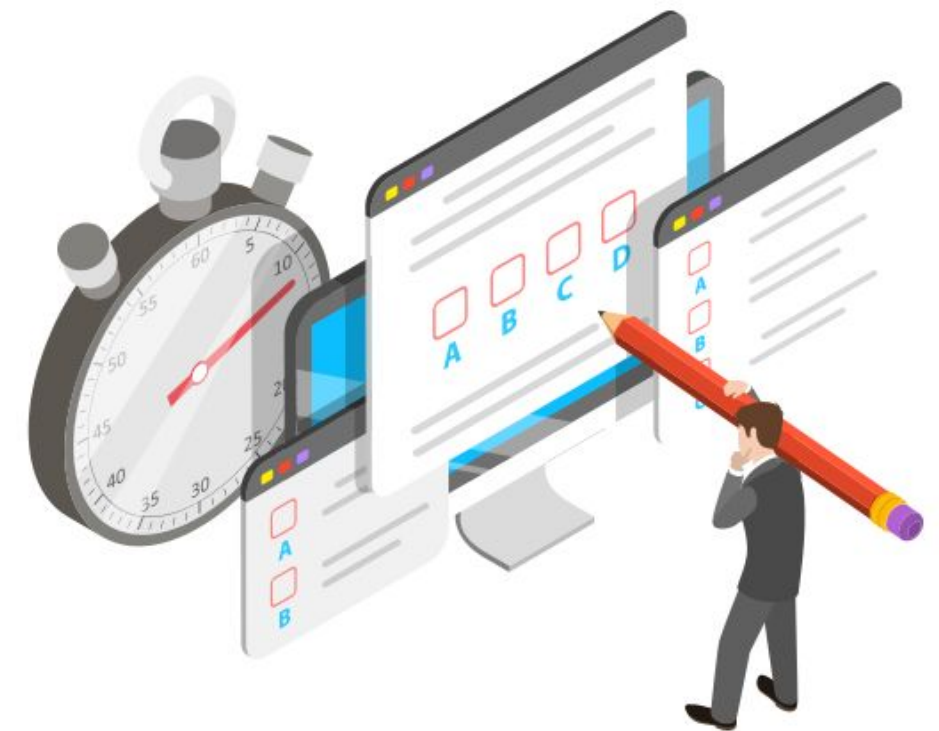
In GRU, information to forget is decided by reset gate.

Knowledge Check

4

Vanishing Gradient is problematic in _____.

- a. RNN
- b. LSTM
- c. GRU
- d. None of the above

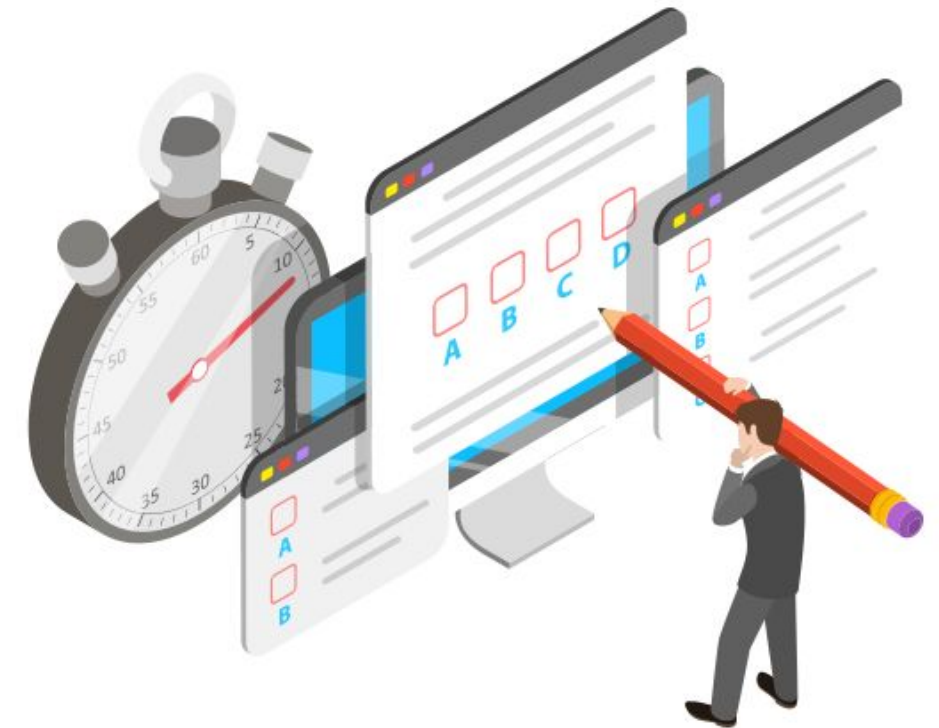


**Knowledge
Check**

4

Vanishing Gradient is problematic in _____.

- a. RNN
- b. LSTM
- c. GRU
- d. None of the above



The correct answer is **a.**

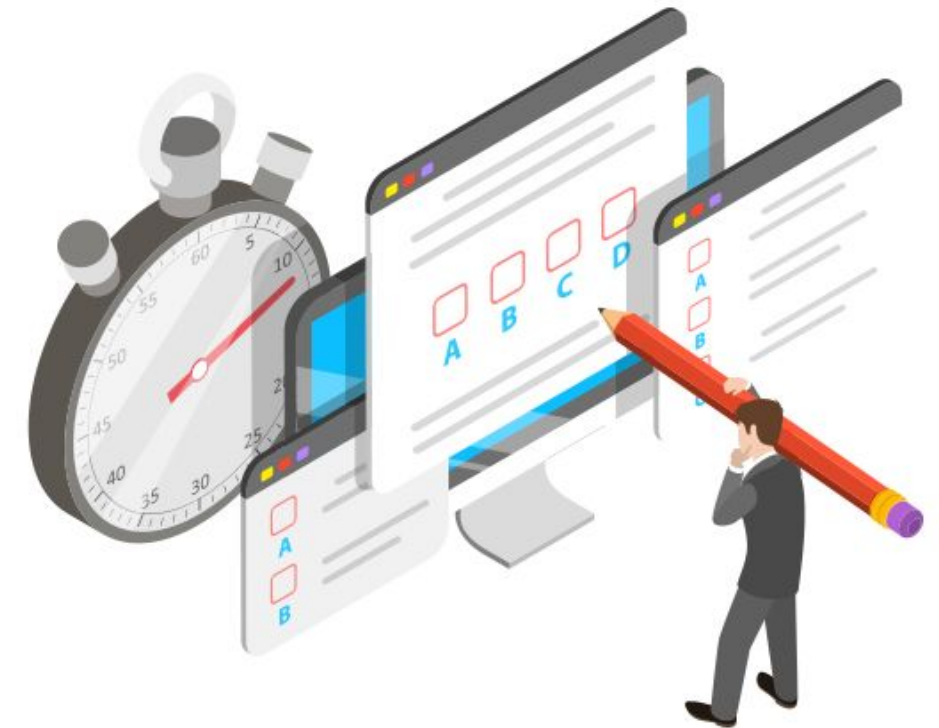
Vanishing gradient is problematic in RNN.

Knowledge Check

5

Which layer understands the answer for questions asked in Question and Answering?

- a. RNN Layer
- b. Embedding Layer
- c. Attention Layer
- d. None of the above

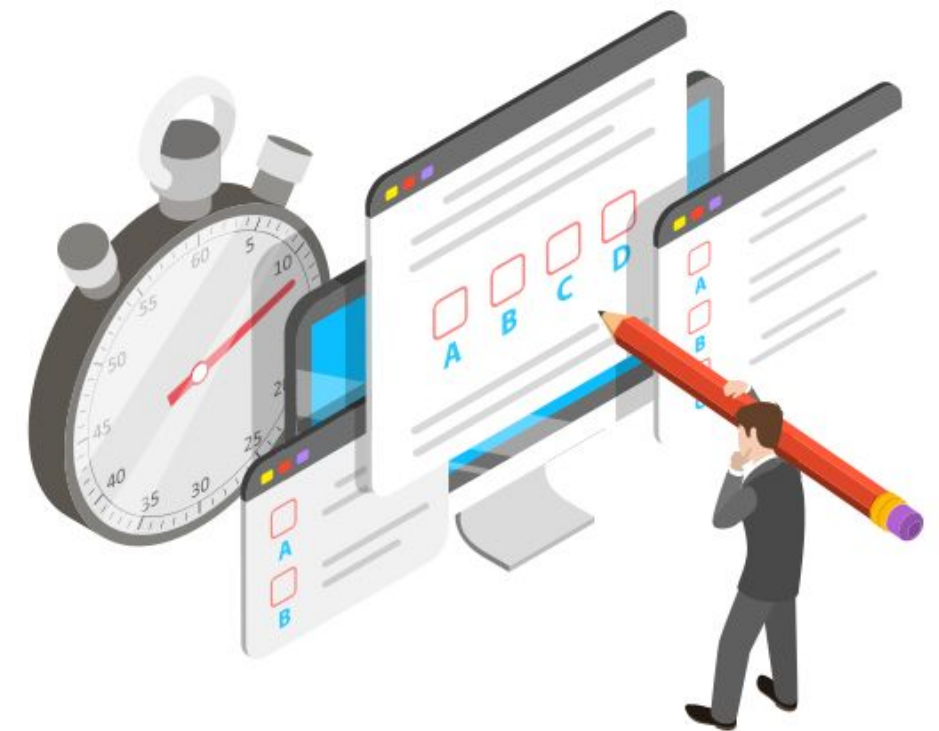


**Knowledge
Check**

5

Which layer understands the answer for questions asked in Question and Answering?

- a. RNN Layer
- b. Embedding Layer
- c. Attention Layer
- d. None of the above



The correct answer is **c.**

Attention layer understands the answer for questions asked in Question and Answering.