1. Identify the correct order of the gates that information flows through in an LSTM unit. 1 / 1 point

- ○ Input gate, forget gate, output gate.
- ◉ Forget gate, input gate, output gate.
- ○ Output gate, forget gate, input gate.
- ○ Forget gate, output gate, input gate

Correct
Correct.

2. Which are some applications of LSTMs? 1 / 1 point

- ☑ Image captioning

Correct
Correct

- ☑ Speech recognition

Correct
Correct

- ☑ Next character prediction

Correct
Correct

- ☑ Chatbots

Correct
Correct

- ☑ Music composition

Correct
Correct

3. The tanh layer ensures the values in your network stay numerically stable, by squeezing all values between -1 and 1. This prevents any of the values from the current inputs from becoming so large that they make the other values insignificant. 1 / 1 point

3. The tanh layer ensures the values in your network stay numerically stable, by squeezing all values between -1 and 1. This prevents any of the values from the current inputs from becoming so large that they make the other values insignificant. 1 / 1 point

- ○ False
- ◉ True

Correct
Correct.

4. What type of architecture is a named entity recognition using? 1 / 1 point

- ◉ Many to many
- ○ One to many
- ○ Many to one

Correct
Correct.

5. Extract the named entities from the following sentence: 1 / 1 point

Younes, a Moroccan artificial intelligence engineer, travelled to France for a conference.

- ○ Younes, Moroccan, conference.
- ◉ Younes, Moroccan, France.
- ○ Younes, Moroccan, engineer.
- ○ Younes, Moroccan engineer, France.

Correct
Correct.

6. In a vectorized representation of your data, equal sequence length allows more efficient batch processing. 1 / 1 point

- ◉ True.
- ○ False

Correct

7. Why is it important to mask padded tokens when computing the loss? 1 / 1 point

- ○ We add the loss of the padded tokens independently.
- ◉ Padded tokens are not part of the data and are just used to help us keep the same sequence length for more efficient batch processing. We should not include their loss.

✓ **Correct**
Correct.

8. In which of the following orders should we train an Named Entity Recognition with an LSTM? 1 / 1 point

- ○
  1. Create a tensor for each input and its corresponding number
  2. Put them in a batch => 64, 128, 256, 512 ...
  3. Run the output through a dense layer
  4. Feed it into an LSTM unit
  5. Predict using a log softmax over K classes
- ○
  1. Create a tensor for each input and its corresponding number
  2. Put them in a batch => 64, 128, 256, 512 ...
  3. Run the output through a dense layer
  4. Predict using a log softmax over K classes
  5. Feed it into an LSTM unit
- ◉
  1. Create a tensor for each input and its corresponding number
  2. Put them in a batch => 64, 128, 256, 512 ...
  3. Feed it into an LSTM unit
  4. Run the output through a dense layer
  5. Predict using a log softmax over K classes

✓ **Correct**
Correct.

9. LSTMS solve vanishing/exploding gradient problems when compared to basic RNNs. 1 / 1 point

- ◉ True
- ○ False

✓ **Correct**

3. Run the output through a dense layer
4. Predict using a log softmax over K classes
5. Feed it into an LSTM unit

◉ 1. Create a tensor for each input and its corresponding number
2. Put them in a batch => 64, 128, 256, 512 ...
3. Feed it into an LSTM unit
4. Run the output through a dense layer
5. Predict using a log softmax over K classes

Correct
Correct.

9. LSTMS solve vanishing/exploding gradient problems when compared to basic RNNs. 1 / 1 point

◉ True

○ False

Correct
Correct.

10. Which of the following are true about LSTMs and vanilla RNNs? 1 / 1 point

☐ LSTMs are typically trained faster than vanilla RNNs.

☑ LSTMs can better retain information from earlier parts of the sentence.

Correct

☐ LSTMs suffer from vanishing gradients, but RNNs don't.

☐ LSTMs suffer from exploding gradients, but RNNs don't.

☑ A single LSTM cell is more complex than a single cell in vanilla RNN.

Correct
Correct. LSTMs use input, output and forget gates to propagate information in a more sophisticated way than vanilla RNNs.