Search models, datasets, users...

Hugging Face is way more fun with friends and colleagues! 🤗 Join an organization

Dismiss this message

NLP Course documentation

End-of-chapter quiz

End-of-chapter quiz

Ask a question

This chapter covered a lot of ground! Don’t worry if you didn’t grasp all the details; the next chapters will help you understand how things work under the hood.

First, though, let’s test what you learned in this chapter!

1. Explore the Hub and look for the roberta-large-mnli checkpoint. What task does it perform?

☐ Summarization

☑ Text classification

Correct! More precisely, it classifies if two sentences are logically linked across three labels (contradiction, neutral, entailment) — a task also called *natural language inference*.

☐ Text generation

Submit **You got all the answers!**

2. What will the following code return?

```
from transformers import pipeline

ner = pipeline("ner", grouped_entities=True)
ner("My name is Sylvain and I work at Hugging Face in Brooklyn.")
```

☐ It will return classification scores for this sentence, with labels "positive" or "negative".

☐ It will return a generated text completing this sentence.

☑ It will return the words representing persons, organizations or locations.

Correct! Furthermore, with `grouped_entities=True`, it will group together the words belonging to the same entity, like "Hugging Face".

Submit **You got all the answers!**

3. What should replace … in this code sample?

```
from transformers import pipeline

filler = pipeline("fill-mask", model="bert-base-cased")
result = filler("...")
```

□ This <mask> has been waiting for you.

☑ This [MASK] has been waiting for you.

Correct! Correct! This model's mask token is [MASK].

□ This man has been waiting for you.

Submit **You got all the answers!**

4. Why will this code fail?

```
from transformers import pipeline

classifier = pipeline("zero-shot-classification")
result = classifier("This is a course about the Transformers library")
```

☑ This pipeline requires that labels be given to classify this text.

Correct! Right — the correct code needs to include `candidate_labels=[...]`.

□ This pipeline requires several sentences, not just one.

□ The 🤗 Transformers library is broken, as usual.

□ This pipeline requires longer inputs; this one is too short.

Submit **You got all the answers!**

5. What does “transfer learning” mean?

□ Transferring the knowledge of a pretrained model to a new model by training it on the same dataset.

☑ Transferring the knowledge of a pretrained model to a new model by initializing the second model with the first model's weights.

Correct! Correct: when the second model is trained on a new task, it *transfers* the knowledge of the first model.

□ Transferring the knowledge of a pretrained model to a new model by building the second model with the same architecture as the first model.

Submit **You got all the answers!**

6. True or false? A language model usually does not need labels for its pretraining.

☑ True

Correct! The pretraining is usually *self-supervised*, which means the labels are created automatically from the inputs (like predicting the next word or filling in some masked words).

□ False

Submit **You got all the answers!**

7. Select the sentence that best describes the terms “model”, “architecture”, and “weights”.

□ If a model is a building, its architecture is the blueprint and the weights are the people living inside.

□ An architecture is a map to build a model and its weights are the cities represented on the map.

☑ An architecture is a succession of mathematical functions to build a model and its weights are those functions parameters.

Correct! The same set of mathematical functions (architecture) can be used to build different models by using different parameters (weights).

Submit **You got all the answers!**

8. Which of these types of models would you use for completing prompts with generated text?

☐ An encoder model

☑ A decoder model

Correct! Decoder models are perfectly suited for text generation from a prompt.

☐ A sequence-to-sequence model

Submit **You got all the answers!**

9. Which of those types of models would you use for summarizing texts?

☐ An encoder model

☐ A decoder model

☑ A sequence-to-sequence model

Correct! Sequence-to-sequence models are perfectly suited for a summarization task.

Submit **You got all the answers!**

10. Which of these types of models would you use for classifying text inputs according to certain labels?

☑ An encoder model

Correct! An encoder model generates a representation of the whole sentence which is perfectly suited for a task like classification.

☐ A decoder model

☐ A sequence-to-sequence model

Submit **You got all the answers!**

11. What possible source can the bias observed in a model have?

☑ The model is a fine-tuned version of a pretrained model and it picked up its bias from it.

Correct! When applying Transfer Learning, the bias in the pretrained model used persists in the fine-tuned model.

☑ The data the model was trained on is biased.

Correct! This is the most obvious source of bias, but not the only one.

☑ The metric the model was optimizing for is biased.

Correct! A less obvious source of bias is the way the model is trained. Your model will blindly optimize for whatever metric you chose, without any second thoughts.

Submit **You got all the answers!**

← Summary

✓ Complete Chapter

Search models, datasets, users...

Hugging Face is way more fun with friends and colleagues! 🤗 Join an organization

Dismiss this message

Pytorch TensorFlow

Ask a question

End-of-chapter quiz

1. What is the order of the language modeling pipeline?

□ First, the model, which handles text and returns raw predictions. The tokenizer then makes sense of these predictions and converts them back to text when needed.

□ First, the tokenizer, which handles text and returns IDs. The model handles these IDs and outputs a prediction, which can be some text.

☑ The tokenizer handles text and returns IDs. The model handles these IDs and outputs a prediction. The tokenizer can then be used once again to convert these predictions back to some text.

Correct! Correct! The tokenizer can be used for both tokenizing and de-tokenizing.

Submit **You got all the answers!**

2. How many dimensions does the tensor output by the base Transformer model have, and what are they?

□ 2: The sequence length and the batch size

□ 2: The sequence length and the hidden size

☑ 3: The sequence length, the batch size, and the hidden size

Correct! Correct!

Submit **You got all the answers!**

3. Which of the following is an example of subword tokenization?

☑ WordPiece

Correct! Yes, that's one example of subword tokenization!

□ Character-based tokenization

□ Splitting on whitespace and punctuation

☑ BPE

Correct! Yes, that's one example of subword tokenization!

☑ Unigram

Correct! Yes, that's one example of subword tokenization!

□ None of the above

Submit **You got all the answers!**

4. What is a model head?

□ A component of the base Transformer network that redirects tensors to their correct layers

□ Also known as the self-attention mechanism, it adapts the representation of a token according to the other tokens of the sequence

☑ An additional component, usually made up of one or a few layers, to convert the transformer predictions to a task-specific output

Correct! That's right. Adaptation heads, also known simply as heads, come up in different forms: language modeling heads, question answering heads, sequence classification heads...

Submit **You got all the answers!**

5. What is an AutoModel?

□ A model that automatically trains on your data

☑ An object that returns the correct architecture based on the checkpoint

Correct! Exactly: the `AutoModel` only needs to know the checkpoint from which to initialize to return the correct architecture.

☐ A model that automatically detects the language used for its inputs to load the correct weights

Submit **You got all the answers!**

6. What are the techniques to be aware of when batching sequences of different lengths together?

☑ Truncating

Correct! Yes, truncation is a correct way of evening out sequences so that they fit in a rectangular shape. Is it the only one, though?

☐ Returning tensors

☑ Padding

Correct! Yes, padding is a correct way of evening out sequences so that they fit in a rectangular shape. Is it the only one, though?

☑ Attention masking

Correct! Absolutely! Attention masks are of prime importance when handling sequences of different lengths. That's not the only technique to be aware of, however.

Submit **You got all the answers!**

7. What is the point of applying a SoftMax function to the logits output by a sequence classification model?

☐ It softens the logits so that they're more reliable.

☑ It applies a lower and upper bound so that they're understandable.

Correct! Correct! The resulting values are bound between 0 and 1. That's not the only reason we use a SoftMax function, though.

☑ The total sum of the output is then 1, resulting in a possible probabilistic interpretation.

Correct! Correct! That's not the only reason we use a SoftMax function, though.

Submit **You got all the answers!**

8. What method is most of the tokenizer API centered around?

□ `encode`, as it can encode text into IDs and IDs into predictions

☑ Calling the tokenizer object directly.

Correct! Exactly! The `__call__` method of the tokenizer is a very powerful method which can handle pretty much anything. It is also the method used to retrieve predictions from a model.

□ `pad`

□ `tokenize`

9. What does the result variable contain in this code sample?

```
from transformers import AutoTokenizer

tokenizer = AutoTokenizer.from_pretrained("bert-base-cased")
result = tokenizer.tokenize("Hello!")
```

☑ A list of strings, each string being a token

Correct! Absolutely! Convert this to IDs, and send them to a model!

□ A list of IDs

□ A string containing all of the tokens

Submit **You got all the answers!**

10. Is there something wrong with the following code?

```
from transformers import AutoTokenizer, AutoModel

tokenizer = AutoTokenizer.from_pretrained("bert-base-cased")
model = AutoModel.from_pretrained("gpt2")

encoded = tokenizer("Hey!", return_tensors="pt")
result = model(**encoded)
```

□ No, it seems correct.

☑ The tokenizer and model should always be from the same checkpoint.

Correct! Right!

□ It's good practice to pad and truncate with the tokenizer as every input is a batch.

Submit **You got all the answers!**

← Basic usage completed!

✓ Complete Chapter

Search models, datasets, users...

Hugging Face is way more fun with friends and colleagues! 🤗 Join an organization

Dismiss this message

Pytorch TensorFlow

Ask a question

End-of-chapter quiz

Test what you learned in this chapter!

1. The emotion dataset contains Twitter messages labeled with emotions. Search for it in the Hub , and read the dataset card. Which of these is not one of its basic emotions?

□ Joy

□ Love

☑ Confusion

Correct! Correct! Confusion is not one of the six basic emotions.

□ Surprise

Submit **You got all the answers!**

2. Search for the ar_sarcasm dataset in the Hub . Which task does it support?

☑ Sentiment classification

Correct! That's right! You can tell thanks to the tags.

□ Machine translation

□ Named entity recognition

□ Question answering

Submit **You got all the answers!**

3. How does the BERT model expect a pair of sentences to be processed?

☐ Tokens_of_sentence_1 [SEP] Tokens_of_sentence_2

☐ [CLS] Tokens_of_sentence_1 Tokens_of_sentence_2

☑ [CLS] Tokens_of_sentence_1 [SEP] Tokens_of_sentence_2 [SEP]

Correct! That's correct!

☐ [CLS] Tokens_of_sentence_1 [SEP] Tokens_of_sentence_2

Submit **You got all the answers!**

4. What are the benefits of the Dataset.map() method?

☐ The results of the function are cached, so it won't take any time if we re-execute the code.

☑ It can apply multiprocessing to go faster than applying the function on each element of the dataset.

Correct! This is a neat feature of this method, but it's not the only one!

☑ It does not load the whole dataset into memory, saving the results as soon as one element is processed.

Correct! That's one advantage of this method. There are others, though!

Submit **You didn't select all the correct answers, there's more!**

5. What does dynamic padding mean?

☐ It's when you pad the inputs for each batch to the maximum length in the whole dataset.

☑ It's when you pad your inputs when the batch is created, to the maximum length of the sentences inside that batch.

Correct! That's correct! The "dynamic" part comes from the fact that the size of each batch is determined at the time of creation, and all your batches might have different shapes as a result.

☐ It's when you pad your inputs so that each sentence has the same number of tokens as the previous one in the dataset.

Submit **You got all the answers!**

6. What is the purpose of a collate function?

☐ It ensures all the sequences in the dataset have the same length.

☑ It puts together all the samples in a batch.

Correct! Correct! You can pass the collate function as an argument of a `DataLoader`. We used the `DataCollatorWithPadding` function, which pads all items in a batch so they have the same length.

☐ It preprocesses the whole dataset.

☐ It truncates the sequences in the dataset.

Submit **You got all the answers!**

7. What happens when you instantiate one of the AutoModelForXxx classes with a pretrained language model (such as bert-base-uncased) that corresponds to a different task than the one for which it was trained?

☐ Nothing, but you get a warning.

NLP Course documentation

End-of-chapter quiz ⌄

Correct! Correct. For example, when we used `AutoModelForSequenceClassification` with `bert-base-uncased`, we got warnings when instantiating the model. The pretrained head is not used for the sequence classification task, so it's discarded and a new head is instantiated with random weights.

☐ The head of the pretrained model is discarded.

☐ Nothing, since the model can still be fine-tuned for the different task.

Submit **You got all the answers!**

8. What’s the purpose of TrainingArguments ?

☑ It contains all the hyperparameters used for training and evaluation with the `Trainer`.

Correct! Correct!

☐ It specifies the size of the model.

☐ It just contains the hyperparameters used for evaluation.

☐ It just contains the hyperparameters used for training.

Submit **You got all the answers!**

9. Why should you use the 🤗 Accelerate library?

□ It provides access to faster models.

□ It provides a high-level API so I don't have to implement my own training loop.

☑ It makes our training loops work on distributed strategies.

Correct! Correct! With 🤗 Accelerate, your training loops will work for multiple GPUs and TPUs.

□ It provides more optimization functions.

Submit **You got all the answers!**

← Fine-tuning, Check!

✓ Complete Chapter