

# Question answering example

This folder contains the `run_qa.py` script, demonstrating *question answering* with the 🤗 Transformers library. For straightforward use-cases you may be able to use this script without modification, although we have also included comments in the code to indicate areas that you may need to adapt to your own projects.

## Usage notes

Note that when contexts are long they may be split into multiple training cases, not all of which may contain the answer span.

As-is, the example script will train on SQuAD or any other question-answering dataset formatted the same way, and can handle user inputs as well.

## Multi-GPU and TPU usage

By default, the script uses a `MirroredStrategy` and will use multiple GPUs effectively if they are available. TPUs can also be used by passing the name of the TPU resource with the `--tpu` argument. There are some issues surrounding these strategies and our models right now, which are most likely to appear in the evaluation/prediction steps. We're actively working on better support for multi-GPU and TPU training in TF, but if you encounter problems a quick workaround is to train in the multi-GPU or TPU context and then perform predictions outside of it.

## Memory usage and data loading

One thing to note is that all data is loaded into memory in this script. Most question answering datasets are small enough that this is not an issue, but if you have a very large dataset you will need to modify the script to handle data streaming. This is particularly challenging for TPUs, given the stricter requirements and the sheer volume of data required to keep them fed. A full explanation of all the possible pitfalls is a bit beyond this example script and README, but for more information you can see the 'Input Datasets' section of [this document](#).

## Example command

```
python run_qa.py \
--model_name_or_path distilbert-base-cased \
--output_dir output \
--dataset_name squad \
--do_train \
--do_eval \
```