

DeepSeek-R1-Distill-Qwen-1.5B-GPTQ-Int4

1. [Click to download the model](#) and upload it to raspberrypi5.

File Description

```
m5stack@raspberrypi:~/rsp/DeepSeek-R1-Distill-Qwen-1.5B-GPTQ-Int4$ ls -lh
total 2.9M
-rw-rw-r-- 1 m5stack m5stack    0 Aug 12 10:56 config.json
drwxrwxr-x 2 m5stack m5stack 4.0K Aug 12 11:00 deepseek-r1-1.5b-gptq-int4-ax650
drwxrwxr-x 2 m5stack m5stack 4.0K Aug 12 10:56 deepseek-r1_tokenizer
-rw-rw-r-- 1 m5stack m5stack 4.2K Aug 12 10:56 deepseek-r1_tokenizer.py
-rw-rw-r-- 1 m5stack m5stack 976K Aug 12 10:58 main_axcl_aarch64
-rw-rw-r-- 1 m5stack m5stack 999K Aug 12 10:58 main_axcl_x86
-rw-rw-r-- 1 m5stack m5stack 932K Aug 12 10:58 main_prefill
-rw-rw-r-- 1 m5stack m5stack  277 Aug 12 10:56 post_config.json
-rw-rw-r-- 1 m5stack m5stack 7.4K Aug 12 10:56 README.md
-rw-rw-r-- 1 m5stack m5stack  533 Aug 12 10:56 run_deepseek-r1_1.5b_gptq_int4_ax650.sh
-rw-rw-r-- 1 m5stack m5stack  538 Aug 12 10:56 run_deepseek-r1_1.5b_gptq_int4_axcl_aarch64.sh
-rw-rw-r-- 1 m5stack m5stack  534 Aug 12 10:56 run_deepseek-r1_1.5b_gptq_int4_axcl_x86.sh
```

2. Create a virtual environment

```
python -m venv deepseek
```

3. Activate the virtual environment

```
source deepseek/bin/activate
```

4. Install dependencies

```
pip install transformers jinja2
```

5. Start the tokenizer parser

```
python deepseek-r1_tokenizer.py --port 12345
```

6. Run the tokenizer service — Host IP defaults to `localhost`, port should be set to `12345`.

Once running, you should see something like this:

```
(deepseek) m5stack@raspberrypi:~/rsp/DeepSeek-R1-Distill-Qwen-1.5B-GPTQ-Int4 $ python deepseek-r1_tokenizer.
None of PyTorch, TensorFlow >= 2.0, or Flax have been found. Models won't be available and only tokenizers,
151646 < |begin_of_sentence| > 151643 < |end_of_sentence| >
< |begin_of_sentence| > You are DeepSeek-R1, You are a helpful assistant.< |User| > hello world< |Assistant| >
[151646, 151646, 2610, 525, 18183, 39350, 10911, 16, 11, 1446, 525, 264, 10950, 17847, 13, 151644, 14990, 18
http://localhost:12345
```

Tip

The following steps should be done in a **new terminal** on the raspberrypi.

