Evaluation

(myenv) PS C:\Users\taofe\Desktop\cs6320 project\t-zero\evaluation> python run\_eval.py\

--dataset\_name super\_glue\

--dataset\_config\_name rte\

--template\_name "must be true"\

--model\_name\_or\_path bigscience/T0\_3B\

--output\_dir ./debug

05/03/2025 15:04:38 - INFO - \_\_main\_\_ - Distributed environment: NO

Num processes: 1

Process index: 0

Local process index: 0

Device: cuda

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

loading file spiece.model from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\spiece.model

loading file tokenizer.json from cache at None

loading file added\_tokens.json from cache at None

loading file special\_tokens\_map.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\special\_tokens\_map.json

loading file tokenizer\_config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\tokenizer\_config.json

loading file chat\_template.jinja from cache at None

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

You are using the default legacy behaviour of the <class 'transformers.models.t5.tokenization\_t5.T5Tokenizer'>. This is expected, and simply means that the `legacy` (previous) behavior will be used so nothing changes for you. If you want to use the new behaviour, set `legacy=False`. This should only be set if you understand what it means, and thoroughly read the reason why this was added as explained in https://github.com/huggingface/transformers/pull/24565

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

05/03/2025 15:04:39 - INFO - t0.model - Building EncoderDecoderModel

loading weights file model.safetensors from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\model.safetensors

Generate config GenerationConfig {

"decoder\_start\_token\_id": 0,

"eos\_token\_id": 1,

"pad\_token\_id": 0

}

All model checkpoint weights were used when initializing T5ForConditionalGeneration.

All the weights of T5ForConditionalGeneration were initialized from the model checkpoint at bigscience/T0\_3B.

If your task is similar to the task the model of the checkpoint was trained on, you can already use T5ForConditionalGeneration for predictions without further training.

Generation config file not found, using a generation config created from the model config.

05/03/2025 15:04:40 - INFO - \_\_main\_\_ - Sample 240 of the training set: {'input\_ids': [[9246, 24, 8595, 13334, 9, 2304, 6, 113, 7701, 8, 13052, 29, 12748, 828, 44, 8, 1775, 13, 1015, 6, 5468, 28, 301, 17694, 30, 192, 9194, 81, 8, 1941, 12, 428, 112, 520, 7, 8359, 7, 5, 4063, 6, 34, 398, 36, 1176, 24, 96, 439, 15, 2494, 13334, 9, 2304, 19, 3, 9, 2743, 13, 8, 1775, 13, 1015, 535, 58, 2163, 42, 150, 58], [9246, 24, 8595, 13334, 9, 2304, 6, 113, 7701, 8, 13052, 29, 12748, 828, 44, 8, 1775, 13, 1015, 6, 5468, 28, 301, 17694, 30, 192, 9194, 81, 8, 1941, 12, 428, 112, 520, 7, 8359, 7, 5, 4063, 6, 34, 398, 36, 1176, 24, 96, 439, 15, 2494, 13334, 9, 2304, 19, 3, 9, 2743, 13, 8, 1775, 13, 1015, 535, 58, 2163, 42, 150, 58]], 'attention\_mask': [[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[2163, 1], [465, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 0}.

05/03/2025 15:04:40 - INFO - \_\_main\_\_ - Sample 178 of the training set: {'input\_ids': [[9246, 24, 818, 8905, 47, 16, 8, 799, 6, 8, 1945, 1384, 47, 29766, 26, 6, 28, 26, 6, 28, 1 26, 6, 2 26, 6, 28, 186, 1652, 1180, 550, 45, 8, 12787, 30, 5022, 45, 9218, 1387, 4373, 5, 4063, 6, 34, 398, 36, 1176, 24, 96, 634, 412, 5, 134, 5, 2716, 29766, 26, 412, 5, 134, 5, 5169, 535, 58, 2163, 42, 150, 58], [9246, 24, 818, 8905, 47, 16, 8, 799, 6, 8, 1945, 1384, 4 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[2163, 1], [465, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 1}.

05/03/2025 15:04:40 - INFO - \_\_main\_\_ - Sample 243 of the training set: {'input\_ids': [[9246, 24, 10508, 31, 7, 2314, 5562, 19, 28788, 84, 3213, 8, 66, 11097, 663, 13, 167, 12806, 151, 5, 2940, 2779, 4810, 1330, 966, 16, 7, 15, 1893, 3834, 1413, 17, 3757, 15, 26, 28, 20687, 10822, 5, 37, 151, 1032, 12702, 28788, 84, 963, 11, 16224, 358, 13, 3564, 7373, 5, 4063, 6, 34, 398, 36, 1176, 24, 96, 634, 2314, 5562, 13, 10508, 19, 37, 52, 9, 16716, 28788, 535, 58, 2163, 42, 150, 58], [9246, 24, 10508, 31, 7, 2314, 5562, 19, 28788, 84, 3213, 8, 66, 11097, 663, 13, 167, 12806, 151, 5, 2940, 2779, 4810, 1330, 966, 16, 7, 15, 1893, 3834, 1413, 17, 3757, 15, 26, 28, 20687, 10822, 5, 37, 151, 1032, 12702, 28788, 84, 963, 11, 16224, 358, 13, 3564, 7373, 5, 4063, 6, 34, 398, 36, 1176, 24, 96, 634, 2314, 5562, 13, 10508, 19, 37, 52, 9, 16716, 28788, 535, 58, 2163, 42, 150, 58]], 'attention\_mask': [[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[2163, 1], [465, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 1}.

Downloading builder script: 100%|█████████████████████████████████████████████████████████████████████████████████████████████████████| 4.20k/4.20k [00:00<?, ?B/s]

05/03/2025 15:04:51 - INFO - \_\_main\_\_ - \*\*\*\*\* Running evaluation \*\*\*\*\*

05/03/2025 15:04:51 - INFO - \_\_main\_\_ - Num examples = 277

05/03/2025 15:04:51 - INFO - \_\_main\_\_ - Instantaneous batch size per device = 8

05/03/2025 15:04:51 - INFO - \_\_main\_\_ - Total eval batch size (w. parallel, distributed) = 8

0%| | 0/35 [00:00<?, ?it/s]You're using a T5TokenizerFast tokenizer. Please note that with a fast tokenizer, using the `\_\_call\_\_` method is faster than using a method to encode the text followed by a call to the `pad` method to get a padded encoding.

Passing a tuple of `past\_key\_values` is deprecated and will be removed in Transformers v4.48.0. You should pass an instance of `EncoderDecoderCache` instead, e.g. `past\_key\_values=EncoderDecoderCache.from\_legacy\_cache(past\_key\_values)`.

100%|███████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 35/35 [1:25:35<00:00, 146.73s/it]

Result/Output

{

    "dataset\_name": "super\_glue",

    "dataset\_config\_name": "rte",

    "template\_name": "must be true",

    "evaluation": {

        "accuracy": 0.7003610108303249

    }

}

python run\_eval.py --dataset\_name winogrande --dataset\_config\_name winogrande\_xl --template\_name "Replace" --model\_name\_or\_path bigscience/T0\_3B --output\_dir ./debug

05/03/2025 16:48:01 - INFO - \_\_main\_\_ - Distributed environment: NO

Num processes: 1

Process index: 0

Local process index: 0

Device: cuda

README.md: 100%|█████████████████████████████████████████████████████████████████████████████████████████████████| 9.97k/9.97k [00:00<00:00, 2.74MB/s]

C:\Users\taofe\Desktop\cs6320 project\myenv\lib\site-packages\huggingface\_hub\file\_download.py:144: UserWarning: `huggingface\_hub` cache-system uses symlinks by default to efficiently store duplicated files but your machine does not support them in C:\Users\taofe\.cache\huggingface\hub\datasets--winogrande. Caching files will still work but in a degraded version that might require more space on your disk. This warning can be disabled by setting the `HF\_HUB\_DISABLE\_SYMLINKS\_WARNING` environment variable. For more details, see https://huggingface.co/docs/huggingface\_hub/how-to-cache#limitations.

To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In order to activate developer mode, see this article: https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development

warnings.warn(message)

winogrande.py: 100%|█████████████████████████████████████████████████████████████████████████████████████████████| 5.65k/5.65k [00:00<00:00, 1.41MB/s]

Downloading data: 100%|██████████████████████████████████████████████████████████████████████████████████████████| 3.40M/3.40M [00:00<00:00, 13.7MB/s]

Generating train split: 100%|█████████████████████████████████████████████████████████████████████████| 40398/40398 [00:00<00:00, 64560.65 examples/s]

Generating test split: 100%|████████████████████████████████████████████████████████████████████████████| 1767/1767 [00:00<00:00, 32418.13 examples/s]

Generating validation split: 100%|██████████████████████████████████████████████████████████████████████| 1267/1267 [00:00<00:00, 30397.16 examples/s]

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

loading file spiece.model from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\spiece.model

loading file tokenizer.json from cache at None

loading file added\_tokens.json from cache at None

loading file special\_tokens\_map.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\special\_tokens\_map.json

loading file tokenizer\_config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\tokenizer\_config.json

loading file chat\_template.jinja from cache at None

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

You are using the default legacy behaviour of the <class 'transformers.models.t5.tokenization\_t5.T5Tokenizer'>. This is expected, and simply means that the `legacy` (previous) behavior will be used so nothing changes for you. If you want to use the new behaviour, set `legacy=False`. This should only be set if you understand what it means, and thoroughly read the reason why this was added as explained in https://github.com/huggingface/transformers/pull/24565

loading configuration file config.json from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\config.json

Model config T5Config {

"architectures": [

"T5ForConditionalGeneration"

],

"classifier\_dropout": 0.0,

"d\_ff": 5120,

"d\_kv": 64,

"d\_model": 2048,

"decoder\_start\_token\_id": 0,

"dense\_act\_fn": "gelu\_new",

"dropout\_rate": 0.1,

"eos\_token\_id": 1,

"feed\_forward\_proj": "gated-gelu",

"gradient\_checkpointing": false,

"initializer\_factor": 1.0,

"is\_encoder\_decoder": true,

"is\_gated\_act": true,

"layer\_norm\_epsilon": 1e-06,

"model\_type": "t5",

"num\_decoder\_layers": 24,

"num\_heads": 32,

"num\_layers": 24,

"output\_past": true,

"pad\_token\_id": 0,

"relative\_attention\_max\_distance": 128,

"relative\_attention\_num\_buckets": 32,

"tie\_word\_embeddings": false,

"transformers\_version": "4.51.3",

"use\_cache": true,

"vocab\_size": 32128

}

05/03/2025 16:48:04 - INFO - t0.model - Building EncoderDecoderModel

loading weights file model.safetensors from cache at C:\Users\taofe\.cache\huggingface\hub\models--bigscience--T0\_3B\snapshots\032a20e775dd500df0a5a7f404466183d67f172b\model.safetensors

Generate config GenerationConfig {

"decoder\_start\_token\_id": 0,

"eos\_token\_id": 1,

"pad\_token\_id": 0

}

All model checkpoint weights were used when initializing T5ForConditionalGeneration.

All the weights of T5ForConditionalGeneration were initialized from the model checkpoint at bigscience/T0\_3B.

If your task is similar to the task the model of the checkpoint was trained on, you can already use T5ForConditionalGeneration for predictions without further training.

Generation config file not found, using a generation config created from the model config.

Map: 100%|████████████████████████████████████████████████████████████████████████████████████████████████| 1267/1267 [00:02<00:00, 525.42 examples/s]

05/03/2025 16:48:07 - INFO - \_\_main\_\_ - Sample 598 of the training set: {'input\_ids': [[21203, 3, 17, 13296, 550, 66, 8, 12495, 11, 11797, 542, 12, 1428, 622, 1896, 31, 7, 24264, 30, 23363, 4371, 250, 3, 834, 124, 7, 21, 112, 533, 5, 28035, 8, 3, 834, 16, 8, 756, 7142, 28, 8, 2024, 1182, 10, 3, 18, 21203, 3, 18, 622, 1896], [21203, 3, 17, 13296, 550, 66, 8, 12495, 11, 11797, 542, 12, 1428, 622, 1896, 31, 7, 24264, 30, 23363, 4371, 250, 3, 834, 124, 7, 21, 112, 533, 5, 28035, 8, 3, 834, 16, 8, 756, 7142, 28, 8, 2024, 1182, 10, 3, 18, 21203, 3, 18, 622, 1896]], 'attention\_mask': [[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[21203, 1], [622, 1896, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1, 1]], 'targets': 0}.

05/03/2025 16:48:07 - INFO - \_\_main\_\_ - Sample 917 of the training set: {'input\_ids': [[19454, 1114, 12, 1431, 12, 14411, 334, 239, 6, 250, 3, 834, 2994, 5657, 5487, 12, 80, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, , 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[19454, 1], [14411, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 0}.

, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[19454, 1], [14411, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 0}.

05/03/2025 16:48:07 - INFO - \_\_main\_\_ - Sample 581 of the training set: {'input\_ids': [[2549, 16340, 47, 2697, 1096, 8, 20717, 5, 216, 470, 2124, 8, 3, 834, 19, 24, 600, 5, 28035, 8, 3, 834, 16, 8, 756, 7142, 28, 8, 1], [1, 1]], 'targets': 0}.

05/03/2025 16:48:07 - INFO - \_\_main\_\_ - Sample 581 of the training set: {'input\_ids': [[2549, 16340, 47, 2697, 1096, 8, 20717, 5, 216, 470, 2124, 8, 3, 834, 19, 24, 600, 5, 28035, 8, 3, 834, 16, 8, 756, 7142, 28, 8, 2024, 1182, 10, 3, 18, 16340, 3, 18, 20717], [2549, 16340, 47, 2697, 1096, 8, 20717, 5, 216, 470, 2124, 8, 3, 834, 19, 24, 600, 5, 28035, 8, 3, 834, 16, 8, 756, 7142, 28, 8, 2024, 1182, 10, 3, 18, 16340, 3, 18, 20717]], 'attention\_mask': [[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]], 'labels': [[16340, 1], [20717, 1]], 'labels\_attention\_mask': [[1, 1], [1, 1]], 'targets': 1}.

05/03/2025 16:48:18 - INFO - \_\_main\_\_ - \*\*\*\*\* Running evaluation \*\*\*\*\*

05/03/2025 16:48:18 - INFO - \_\_main\_\_ - Num examples = 1267

05/03/2025 16:48:18 - INFO - \_\_main\_\_ - Instantaneous batch size per device = 8

05/03/2025 16:48:18 - INFO - \_\_main\_\_ - Total eval batch size (w. parallel, distributed) = 8

0%| | 0/159 [00:00<?, ?it/s]You're using a T5TokenizerFast tokenizer. Please note that with a fast tokenizer, using the `\_\_call\_\_` method is faster than using a method to encode the text followed by a call to the `pad` method to get a padded encoding.

Passing a tuple of `past\_key\_values` is deprecated and will be removed in Transformers v4.48.0. You should pass an instance of `EncoderDecoderCache` instead, e.g. `past\_key\_values=EncoderDecoderCache.from\_legacy\_cache(past\_key\_values)`.

100%|█████████████████████████████████████████████████████████████████████████████████████████████████████████████| 159/159 [3:36:33<00:00, 81.72s/it]

Result/output

{

    "dataset\_name": "winogrande",

    "dataset\_config\_name": "winogrande\_xl",

    "template\_name": "Replace",

    "evaluation": {

        "accuracy": 0.5248618784530387

    }

}

Inference

(myenv) PS C:\Users\taofe\Desktop\cs6320 project> python inference.py

You are using the default legacy behaviour of the <class 'transformers.models.t5.tokenization\_t5.T5Tokenizer'>. This is expected, and simply means that the `legacy` (previous) behavior will be used so nothing changes for you. If you want to use the new behaviour, set `legacy=False`. This should only be set if you understand what it means, and thoroughly read the reason why this was added as explained in https://github.com/huggingface/transformers/pull/24565

🔹 Example 1: Task - Natural Language Inference (RTE/ANLI)

📝 Prompt: Premise: The man is walking his dog. Hypothesis: A man is out with his pet. Is the hypothesis true?

🔸 Model Output: Yes

🔹 Example 2: Task - Question Answering (BoolQ)

📝 Prompt: Passage: Lightning can strike the same place twice. Question: Can lightning strike the same place twice?

🔸 Model Output: yes

🔹 Example 3: Task - Coreference Resolution (WSC)

📝 Prompt: Sentence: Emma did not pass the ball to Janie although she saw that she was open. Question: Who was open?

🔸 Model Output: Janie

🔹 Example 4: Task - Causal Reasoning (COPA)

📝 Prompt: Premise: The man broke his toe. What was the cause? Choice1: He got a hole in his sock. Choice2: He dropped a hammer on his foot.

🔸 Model Output: He dropped a hammer on his foot.

🔹 Example 5: Task - Sentence Completion (HellaSwag)

📝 Prompt: First, I opened the cupboard. Then, I reached for the cereal. Next, I poured it into the bowl. Finally,

🔸 Model Output: I ate it.

🔹 Example 6: Task - Multiple-Choice QA (ARC/OpenBookQA)

📝 Prompt: Question: Which part of the plant makes seeds? Options: A. root B. stem C. flower D. leaf

🔸 Model Output: C

🔹 Example 7: Task - Word Sense Disambiguation (WiC)

📝 Prompt: Sentence 1: The bank will not be open on Sunday. Sentence 2: He sat down on the river bank. Question: Do the two sentences use the word 'bank' in the same way?

🔸 Model Output: No