

Part 3: LLM Integration and Development

- Text Processing (with a language model):
 - Load a pre-trained LLM (e.g., from Hugging Face Transformers).
 - Do the following:
 - Perform text summarization on a given news article.
 - Generate a few creative text variations based on a seed sentence.

```

1.py X
C:\Users> 91996 > Desktop > Python > 1.py > ...
1 from transformers import pipeline
2
3 # Load pre-trained LLM for text summarization and text generation
4 summarization_pipeline = pipeline("summarization")
5 text_generation_pipeline = pipeline("text-generation")
6
7 # Given news article
8 news_article = """
9 According to recent reports, scientists have discovered a new species of butterfly in the Amazon rainforest. The butterfly, named Morpho amazonicus,
10 """
11
12 # Perform text summarization
13 summary = summarization_pipeline(news_article, max_length=150, min_length=30, do_sample=False)
14 print("Summary of the news article:")
15 print(summary[0]['summary_text'])
16
17 # Seed sentence for text generation
18 seed_sentence = "Once upon a time in a faraway land"
19
20 # Generate creative text variations based on the seed sentence
21 generated_text = text_generation_pipeline(seed_sentence, max_length=100, num_return_sequences=3)
22 print("\nCreative text variations based on the seed sentence:")
23 for i, text in enumerate(generated_text):
24     print(f"Variation {i + 1}: {text['generated_text']}")
25

```

```
PS C:\Users\91996\Desktop\WEB DEV> c:\python311\python.exe c:\Users\91996\Desktop\Python\1.py
No model was supplied, defaulted to google-t5/t5-small and revision d769bba (https://huggingface.co/google-t5/t5-small).
Using a pipeline without specifying a model name and revision in production is not recommended.
model.safetensors: 100% | 242M/242M [12:56<00:00, 312kB/s]
c:\python311\libs\site-packages\huggingface_hub\file_download.py:149: UserWarning: 'huggingface_hub' cache-system uses symlinks by default to efficiently store duplicated files b
ut your machine does not support them in C:\Users\91996\cache\huggingface\hub\models--google-t5/t5-small. Caching files will still work but in a degraded version that might re
quire 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/d
ocs/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 see activate developer mode, see this article: https
://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development
warnings.warn(message)
2024-03-10 22:42:39.099764: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-critical op
erations.
To enable the following instructions: SSE SSE2 SSE3 SSE4.1 SSE4.2 AVX AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
All Pytorch model weights were used when initializing TFT5ForConditionalGeneration.

All the weights of TFT5ForConditionalGeneration were initialized from the Pytorch model.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFT5ForConditionalGeneration for predictions without further training.
tokenizer_config.json: 100% | 2.32k/2.32k [00:00<, 78/s]
spiece.model: 792k/792k [00:00<00:00, 887kB/s]
tokenizer.json: 100% | 1.39M/1.39M [00:07<00:00, 183kB/s]

No model was supplied, defaulted to openai-community/gpt2 and revision 6c0be608 (https://huggingface.co/openai-community/gpt2).
Using a pipeline without specifying a model name and revision in production is not recommended.
config.json: 100% | 665/665 [00:00<, 78/s]
C:\python311\libs\site-packages\huggingface_hub\file_download.py:149: UserWarning: 'huggingface_hub' cache-system uses symlinks by default to efficiently store duplicated files b
ut your machine does not support them in C:\Users\91996\cache\huggingface\hub\models--openai-community-gpt2. Caching files will still work but in a degraded version that might re
quire 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 see activate developer mode, see this article: https
://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development
warnings.warn(message)
model.safetensors: 100% | 548M/548M [10:43<00:00, 852kB/s]
All Pytorch model weights were used when initializing TFGPT2LMHeadModel.

All the weights of TFGPT2LMHeadModel were initialized from the Pytorch model.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFGPT2LMHeadModel for predictions without further training.
tokenizer_config.json: 100% | 26.0/26.0 [00:00<, 78/s]
vocab.json: 100% | 1.04M/1.04M [00:01<00:00, 926kB/s]
merges.txt: 100% | 456k/456k [00:00<00:00, 490kB/s]
tokenizer.json: 100% | 1.36M/1.36M [00:01<00:00, 918kB/s]
```

```
tokenizer.json: 108% | 1.36M/1.36M [00:01:00:00, 918kB/s]
Your max_length is set to 150, but your input_length is only 104. Since this is a summarization task, where outputs shorter than the input are typically wanted, you might consider decreasing max_length manually, e.g. summarizer('...', max_length=52)
2024-03-10 22:53:46.280567: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0xf56208be0 initialized for platform Host (this does not guarantee that XLA will be used). Devices:
2024-03-10 22:53:46.280982: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version
2024-03-10 22:53:46.351213: I tensorflow/compiler/mlir/tensorflow/utils/dump_mlir_util.cc:255] disabling MLIR crash reproducer, set env var 'MLIR_CRASH_REPRODUCER_DIRECTORY' to enable.
2024-03-10 22:53:46.627587: I .\tensorflow\compiler\jit\device_compiler.h:186] Compiled cluster using XLA! This line is logged at most once for the lifetime of the process.
Summary of the news article:
the butterfly, named Morpho amazonicus, has stunning iridescent blue wings . it is believed to be native to the region .
Truncation was not explicitly activated but 'max_length' is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.

Creative text variations based on the seed sentence:
Variation 1: Once upon a time in a faraway land, I met a young boy from a small village.

"This is your mom" I said with a serious tone. He nodded and said. I started walking and when I finally stopped, he asked "How old you are"

When we walked on his land, his mother gave him a hug and asked him not to worry about me. However, I only got one hug. I felt as bad as it ever felt to me.
Variation 2: Once upon a time in a faraway land, the Sun rises with its power.

The Sun appears to be a giant cloud cloud from the east. The sun can be seen as a single star above the horizon.

The sun can be said to reflect energy from all directions. While traveling through space the sun can appear to radiate energy from all directions.

The Sun is the Sun to its Sun as seen from the heavens.

The moon is a glowing orb which
Variation 3: Once upon a time in a faraway land, the ancient people of Tamrisha, and the two peoples who formed the new nation, were united and unified together. This was one of the great moments in Tamrisha. The last years of the Second Crusade were tumultuous events on this continent. Even the very name and organization of Tamrisha is contested, and a multitude of issues were quickly forgotten.

The new king, Emperor Uriel Septim VII, arrived in the capital of the
PS C:\Users\91996\Desktop\WEB DEV> & c:/python311/python.exe c:/Users/91996/Desktop/python/1.py
No model was supplied, defaulted to google-t5/t5-small and revision 6c0e680 (https://huggingface.co/google-t5/t5-small).
Using a pipeline without specifying a model name and revision in production is not recommended.
2024-03-10 23:08:37.369994: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-critical op
```

```
All the weights of TFT5ForConditionalGeneration were initialized from the PyTorch model.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFT5ForConditionalGeneration for predictions without further training.
No model was supplied, defaulted to openai-community/gpt2 and revision 6c0e680 (https://huggingface.co/openai-community/gpt2).
Using a pipeline without specifying a model name and revision in production is not recommended.
All PyTorch model weights were used when initializing TFGPT2LMHeadModel.

All the weights of TFGPT2LMHeadModel were initialized from the PyTorch model.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFGPT2LMHeadModel for predictions without further training.
Your max_length is set to 150, but your input_length is only 104. Since this is a summarization task, where outputs shorter than the input are typically wanted, you might consider decreasing max_length manually, e.g. summarizer('...', max_length=52)
2024-03-10 23:08:42.330127: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0x23feac0f70 initialized for platform Host (this does not guarantee that XLA will be used). Devices:
2024-03-10 23:08:42.330325: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version
2024-03-10 23:08:42.334352: I tensorflow/compiler/mlir/tensorflow/utils/dump_mlir_util.cc:255] disabling MLIR crash reproducer, set env var 'MLIR_CRASH_REPRODUCER_DIRECTORY' to enable.
2024-03-10 23:08:42.353904: I .\tensorflow\compiler\jit\device_compiler.h:186] Compiled cluster using XLA! This line is logged at most once for the lifetime of the process.
Summary of the news article:
the butterfly, named Morpho amazonicus, has stunning iridescent blue wings . it is believed to be native to the region .
Truncation was not explicitly activated but 'max_length' is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.

Creative text variations based on the seed sentence:
Variation 1: Once upon a time in a faraway land and some far away land, a strange, ancient, and strange entity fell from the skies, appearing with its tail long and sharp.

From the ancient, and the ancient -

Shen was like so, and to that day she could not see, nor hear, or even think of anything that was a distant land, but of only a small distance. For then one day one year old Shen, the infant. A child born
Variation 2: Once upon a time in a faraway land called The Great, a man who had long lived in the valley he had come to call home lived. Thereupon, while he was with his wife at the time of their marriage, he came home and saw a white man standing near a well, at which he went up and blessed him. The white man looked at him and said, "Where do you think he is?"

The white man looked down at the man and said, "I
Variation 3: Once upon a time in a faraway land this is the way you want to be, you will always be going back and forth, like a snake in a water of danger. You are always with those in danger, because you will bring their death in the same moment." (Sara 3:5 ESV)

And this is the way God created the world in his time. (see Genesis 4:4-5).
```

- Custom Prompt Engineering:

- o Demonstrate designing clear and effective prompts to leverage the LLM's capabilities. Showcase how prompts can steer the output for a specific use case within your AI solution concept.

Custom Prompt Engineering for AI Solution Concept

AI Solution Concept: Personalized Recipe Recommendation System

Prompt Design:

Prompt 1:

"Generate a personalized recipe recommendation based on the following criteria: I am in the mood for a healthy dinner option that includes chicken and vegetables. Please ensure the recipe is easy to prepare within 30 minutes."

Prompt 2:

"Craft a recipe suggestion that caters to my preference for a vegetarian dish rich in protein. Consider incorporating beans or tofu as the main ingredient and keep the cooking time under 45 minutes."

Explanation:

Clear Criteria: The prompts begin with clear criteria outlining the user's preferences, such as dietary preferences (e.g., vegetarian), main ingredients (e.g., chicken, vegetables, beans, tofu), and time constraints (e.g., within 30 minutes, under 45 minutes).

Specific Ingredients: By specifying ingredients like chicken, vegetables, beans, or tofu, the prompts guide the AI to focus on recipes that align with the user's dietary choices and main ingredient preferences.

Nutritional Focus: The prompts emphasize health and nutrition by highlighting preferences for healthy options and protein-rich meals, ensuring that the AI suggests recipes that meet the user's dietary goals.

Time Constraints: Including time constraints ensures that the recommended recipes are practical and feasible for the user's schedule, enhancing the relevance and usability of the AI-generated suggestions.

Flexibility: While the prompts provide specific guidelines, they also offer some flexibility for creativity within the given constraints. This allows the AI to generate diverse recipe suggestions that cater to the user's preferences while considering variations in ingredients and cooking methods.

Impact:

By leveraging these well-crafted prompts, the AI solution can effectively steer the output towards personalized recipe recommendations that align with the user's dietary preferences, nutritional goals, and time constraints. This enhances the user experience by providing relevant and practical suggestions tailored to their individual needs, ultimately increasing user satisfaction and engagement with the AI solution.