

```
!pip install torch==1.13.1+cu116 torchvision==0.14.1+cu116 torchaudio==0.13.1 torchtext==0.14.1 torchdata==0.5.1 --extra-index-url https://
```

Looking in indexes: <https://pypi.org/simple>, <https://download.pytorch.org/whl/cu116>

Collecting torch==1.13.1+cu116

Downloading https://download.pytorch.org/whl/cu116/torch-1.13.1%2Bcu116-cp310-cp310-linux_x86_64.whl (1977.9 MB)
2.0/2.0 GB 599.8 kB/s eta 0:00:00

Collecting torchvision==0.14.1+cu116

Downloading https://download.pytorch.org/whl/cu116/torchvision-0.14.1%2Bcu116-cp310-cp310-linux_x86_64.whl (24.2 MB)
24.2/24.2 MB 18.6 MB/s eta 0:00:00

Collecting torchaudio==0.13.1

Downloading https://download.pytorch.org/whl/cu116/torchaudio-0.13.1%2Bcu116-cp310-cp310-linux_x86_64.whl (4.2 MB)
4.2/4.2 MB 57.5 MB/s eta 0:00:00

Collecting torchtext==0.14.1

Downloading torchtext-0.14.1-cp310-cp310-manylinux1_x86_64.whl (2.0 MB)
2.0/2.0 MB 22.0 MB/s eta 0:00:00

Collecting torchdata==0.5.1

Downloading torchdata-0.5.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (4.6 MB)
4.6/4.6 MB 61.4 MB/s eta 0:00:00

Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch==1.13.1+cu116) (4.11.0)

Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from torchvision==0.14.1+cu116) (1.25.2)

Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from torchvision==0.14.1+cu116) (2.31.0)

Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in /usr/local/lib/python3.10/dist-packages (from torchvision==0.14.1+cu116) (9.4)

Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from torchtext==0.14.1) (4.66.2)

Requirement already satisfied: urllib3>=1.25 in /usr/local/lib/python3.10/dist-packages (from torchdata==0.5.1) (2.0.7)

Collecting portalocker==2.0.0 (from torchdata==0.5.1)

Downloading portalocker-2.8.2-py3-none-any.whl (17 kB)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->torchvision==0.14.1+cu116) (3.2.0)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->torchvision==0.14.1+cu116) (3.10.1)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->torchvision==0.14.1+cu116) (2024.7.4)

Installing collected packages: torch, portalocker, torchvision, torchtext, torchdata, torchaudio

Attempting uninstall: torch

Found existing installation: torch 2.2.1+cu121

Uninstalling torch-2.2.1+cu121:

Successfully uninstalled torch-2.2.1+cu121

Attempting uninstall: torchvision

Found existing installation: torchvision 0.17.1+cu121

Uninstalling torchvision-0.17.1+cu121:

Successfully uninstalled torchvision-0.17.1+cu121

Attempting uninstall: torchtext

Found existing installation: torchtext 0.17.1

Uninstalling torchtext-0.17.1:

Successfully uninstalled torchtext-0.17.1

Attempting uninstall: torchdata

Found existing installation: torchdata 0.7.1

Uninstalling torchdata-0.7.1:

Successfully uninstalled torchdata-0.7.1

Attempting uninstall: torchaudio

Found existing installation: torchaudio 2.2.1+cu121

Uninstalling torchaudio-2.2.1+cu121:

Successfully uninstalled torchaudio-2.2.1+cu121

Successfully installed portalocker-2.8.2 torch-1.13.1+cu116 torchaudio-0.13.1+cu116 torchdata-0.5.1 torchtext-0.14.1 torchvision-0.14.1+cu116

```
# Installing the BLURR library and Bert-Score package
```

```
!pip install ohmeow-blurr -q
```

```
!pip install bert-score -q
```

```
!pip install wandb
```

```
!pip install sacrebleu
```

81.1/81.1 kB 2.8 MB/s eta 0:00:00
510.5/510.5 kB 12.8 MB/s eta 0:00:00
43.6/43.6 kB 5.7 MB/s eta 0:00:00

Preparing metadata (setup.py) ... done

Preparing metadata (setup.py) ... done

116.3/116.3 kB 16.5 MB/s eta 0:00:00
194.1/194.1 kB 25.0 MB/s eta 0:00:00
134.8/134.8 kB 18.9 MB/s eta 0:00:00

Building wheel for rouge-score (setup.py) ... done

Building wheel for seqeval (setup.py) ... done

61.1/61.1 kB 1.5 MB/s eta 0:00:00

Collecting wandb

Downloading wandb-0.16.6-py3-none-any.whl (2.2 MB)

2.2/2.2 MB 24.2 MB/s eta 0:00:00

Requirement already satisfied: Click!=8.0.0,>=7.1 in /usr/local/lib/python3.10/dist-packages (from wandb) (8.1.7)

Collecting GitPython!=3.1.29,>=1.0.0 (from wandb)

Downloading GitPython-3.1.43-py3-none-any.whl (207 kB)

207.3/207.3 kB 24.6 MB/s eta 0:00:00

Requirement already satisfied: requests<3,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (2.31.0)

Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (5.9.5)

Collecting sentry-sdk>=1.0.0 (from wandb)

Downloading sentry_sdk-1.45.0-py2.py3-none-any.whl (267 kB)

267.1/267.1 kB 32.2 MB/s eta 0:00:00

Collecting docker-pycreds>=0.4.0 (from wandb)

Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)

Requirement already satisfied: PyYAML in /usr/local/lib/python3.10/dist-packages (from wandb) (6.0.1)

Collecting setproctitle (from wandb)

Downloading setproctitle-1.3.3-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1 kB)

Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb) (67.7.2)

Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb) (1.4.4)

Requirement already satisfied: protobuf!=4.21.0,<5,>=3.19.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (3.20.3)

Requirement already satisfied: six>=1.4.0 in /usr/local/lib/python3.10/dist-packages (from docker-pycreds>=0.4.0->wandb) (1.16.0)

Collecting gitdb<5,>=4.0.1 (from GitPython!=3.1.29,>=1.0.0->wandb)

Downloading gitdb-4.0.11-py3-none-any.whl (62 kB)

62.7/62.7 kB 9.0 MB/s eta 0:00:00

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (3.7.7)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (2024.2.2)
Collecting smmap<6,>=3.0.1 (from gitdb<5,>=4.0.1->GitPython!=3.1.29,>=1.0.0->wandb)
 Downloading smmap-5.0.1-py3-none-any.whl (24 kB)
Installing collected packages: smmap, setproctitle, sentry-sdk, docker-pycreds, gitdb, GitPython, wandb
Successfully installed GitPython-3.1.43 docker-pycreds-0.4.0 gitdb-4.0.11 sentry-sdk-1.45.0 setproctitle-1.3.3 smmap-5.0.1 wandb-0.16.0
Collecting sacrebleu
 Downloading sacrebleu-2.4.2-py3-none-any.whl (106 kB)
106.7/106.7 kB 3.0 MB/s eta 0:00:00
Requirement already satisfied: portalocker in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (2.8.2)
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (2023.12.25)
Requirement already satisfied: tabulate>=0.8.9 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (0.9.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (1.25.2)
Collecting colorama (from sacrebleu)
 Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (4.9.4)
Installing collected packages: colorama, sacrebleu
Successfully installed colorama-0.4.6 sacrebleu-2.4.2

```
import pandas as pd
from fastai.text.all import *
from transformers import *
from blurr.text.data.all import *
from blurr.text.modeling.all import *
from fastai.callback.wandb import *
```

```
/usr/local/lib/python3.10/dist-packages/transformers/deepspeed.py:23: FutureWarning: t
warnings.warn(
WARNING:jax._src.xla_bridge:CUDA backend failed to initialize: Unable to use CUDA beca
Outdated cudNN installation found.
Version JAX was built against: 8906
Minimum supported: 8900
Installed version: 8302
The local installation version must be no lower than 8900. (Set TF_CPP_MIN_LOG_LEVEL=0
/usr/local/lib/python3.10/dist-packages/transformers/generation_utils.py:24: FutureWarni
warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation_tf_utils.py:24: FutureE
warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation_flax_utils.py:24: Futu
warnings.warn(
/usr/local/lib/python3.10/dist-packages/blurr/text/modeling/question_answering.py:31: l
squad_metric = load_metric("squad")
/usr/local/lib/python3.10/dist-packages/datasets/load.py:756: FutureWarning: The repos
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next m
warnings.warn(
Downloading builder script: 4.50k/? [00:00<00:00, 172kB/s]
Downloading extra modules: 3.29k/? [00:00<00:00, 120kB/s]
```

```
df=pd.read_csv("gpt2_result.csv")
```

```
new_df =df[['cleaned_lyrics','gpt_summaries']].copy()
```

```
train_df=pd.DataFrame(columns=['lyc', 'summ'])
test_df=pd.DataFrame(columns=['lyc', 'summ'])
```

```
lyc_tr=[]
lyc_ts=[]
sum_tr=[]
sum_ts=[]
```

```
i=0
for lyrics,summary in zip(new_df.cleaned_lyrics,new_df.gpt_summaries):
    if i<5500:
        lyc_tr.append(lyrics)
        sum_tr.append(summary)
    else:
        lyc_ts.append(lyrics)
        sum_ts.append(summary)
    i=i+1
```

```
train_df['lyc']=lyc_tr
train_df['summ']=sum_tr
test_df['lyc']=lyc_ts
test_df['summ']=sum_ts
```

```
train_df["lyc"]="summarize:"+train_df["lyc"]
```

```
pip install tiktoken
```

Collecting tiktoken

Downloading tiktoken-0.6.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.8 MB)

1.8/1.8 MB 20.4 MB/s eta 0:00:00

Requirement already satisfied: regex<=2022.1.18 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2023.12.25)

Requirement already satisfied: requests<=2.26.0 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2.31.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<=2.26.0->tiktoken) (3.7)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<=2.26.0->tiktoken) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<=2.26.0->tiktoken) (2.0.7)

Requirement already satisfied: certifi<=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<=2.26.0->tiktoken) (2024.

Installing collected packages: tiktoken

Successfully installed tiktoken-0.6.0

```
import tiktoken
```

```
def nt(string: str, encoding_name: str) -> int:
    """Returns the number of tokens in a text string."""
    encoding = tiktoken.get_encoding(encoding_name)
    num_tokens = len(encoding.encode(string))
    return num_tokens
```

```
#Import the pretrained model
```

```
pretrained_model_name = "sshleifer/distilbart-cnn-12-6"
```

```
hf_arch, hf_config, hf_tokenizer, hf_model = get_hf_objects(pretrained_model_name,
                                                             model_cls=BartForConditionalGeneration)
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens).
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models.
  warnings.warn(
```

```
config.json: 100%                               1.80k/1.80k [00:00<00:00, 62.6kB/s]
```

```
tokenizer_config.json: 100%                      26.0/26.0 [00:00<00:00, 1.62kB/s]
```

```
vocab.json: 100%                                899k/899k [00:00<00:00, 1.13MB/s]
```

```
merges.txt: 100%                                456k/456k [00:00<00:00, 767kB/s]
```

```
pytorch_model.bin: 100%                         1.22G/1.22G [00:13<00:00, 87.4MB/s]
```

```
from transformers import BartConfig
```

```
# Assuming hf_config is your original PEGASUS configuration
model_hf_config = hf_config.to_dict()
```

```
# Check if the key exists before attempting to delete it
if "summarization" in model_hf_config:
    if "prefix" in model_hf_config["summarization"]:
        del model_hf_config["summarization"]["prefix"]
```

```
# Create a new configuration object from the modified dictionary
hf_config = BartConfig.from_dict(model_hf_config)
```

```
preprocessor = SummarizationPreprocessor(
    hf_tokenizer,
    id_attr="id",
    text_attr="lyc",
    target_text_attr="summ",
    max_input_tok_length=1024,
    max_target_tok_length=128,
    min_summary_char_length=30
)
```

```
from blurr.text.data.seq2seq.summarization import SummarizationPreprocessor
```

```
import warnings
warnings.filterwarnings('ignore')
# Running the preprocssor declared above
proc_df = pd.DataFrame()
for g, batch_df in train_df.groupby(np.arange(len(train_df)) // preprocessor.batch_size):
    proc_df = pd.concat([proc_df, preprocessor._process_df_batch(batch_df)], ignore_index=True)
```

```
text_gen_kwargs = default_text_gen_kwargs(hf_config, hf_model, task="summarization")
```

```
batch_tokenize_transform = Seq2SeqBatchTokenizeTransform(
    hf_arch, hf_config, hf_tokenizer, hf_model, text_gen_kwargs=text_gen_kwargs
)
```

```
# Creation of Seq2Seq Text Block
blocks = (Seq2SeqTextBlock(batch_tokenize_tfm = batch_tokenize_transform), noop)
```

```
# Creating DataBlock for the blocks declared above
dblock = DataBlock(blocks=blocks, get x=ColReader("proc lyc"), get y=ColReader("proc summ"), splitter=RandomSplitter())
```

```
dls = dblock.dataloaders(proc_df, bs=4)
```

```
summarization_metrics = {
    "rouge": {
        "compute_kwargs": {"rouge_types": ["rouge1", "rouge2", "rougeL"],
            "use_stemmer": True
        },
        "returns": ["rouge1", "rouge2", "rougeL"],
    },
    "bertscore": {"compute_kwargs": {"lang": "en"},
        "returns": ["precision", "recall", "f1"]
    },
    "bleu": {"returns": "bleu"},
    "meteor": {"returns": "meteor"},
    "sacrebleu": {"returns": "score"}
}
```

```
!pip install sacrebleu
```

```
Requirement already satisfied: sacrebleu in /usr/local/lib/python3.10/dist-packages (2.4.2)
Requirement already satisfied: portalocker in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (2.8.2)
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (2023.12.25)
Requirement already satisfied: tabulate>=0.8.9 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (0.9.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (1.25.2)
Requirement already satisfied: colorama in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (0.4.6)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (4.9.4)
```

```
# Intialize Model for hf_model using BaseModelWrapper
model = BaseModelWrapper(hf_model)
# Learner Callbacks
learn_cbs = [BaseModelCallback]

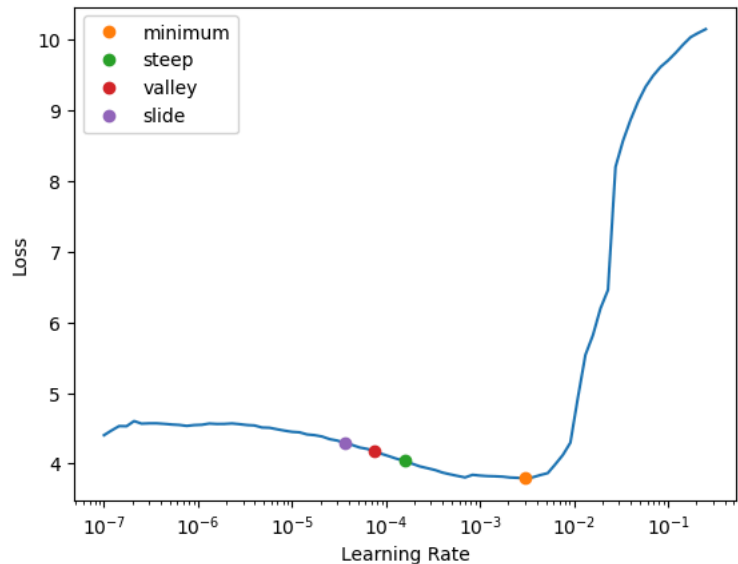
# Callback for calculating metrics
fit_cbs = [Seq2SeqMetricsCallback(custom_metrics=summarization_metrics, calc_every="epoch")]

# Intialize Fast.AI Learner
learn = Learner(
    dls,
    model,
    opt_func=partial(Adam),
    loss_func=PreCalculatedCrossEntropyLoss(),
    cbs=learn_cbs,
    splitter=partial(blurr_seq2seq_splitter, arch=hf_arch)
)

# # Setting precision to 16-bit floating
learn = learn.to_fp16()
learn.freeze()
```

```
Downloading builder script: 5.65k/? [00:00<00:00, 84.0kB/s]
Downloading builder script: 8.10k/? [00:00<00:00, 447kB/s]
Downloading builder script: 6.06k/? [00:00<00:00, 101kB/s]
Downloading extra modules: 4.07k/? [00:00<00:00, 204kB/s]
Downloading builder script: 5.34k/? [00:00<00:00, 321kB/s]
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
Downloading builder script: 7.65k/? [00:00<00:00, 146kB/s]
```

```
# Finding Optimal Learning Rate
res = learn.lr_find(suggest_funcs=[minimum, steep, valley, slide])
```



```
# Installing nltk package
import nltk
nltk.download('punkt')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
True
```

```
df = learn.fit_one_cycle(3, lr_max=res.valley, cbs=fit_cbs)
```

epoch	train_loss	valid_loss	rouge1	rouge2	rougeL	bertscore_precision	bertscore
0	1.984028	1.873365	0.449425	0.159410	0.291924	0.895606	
1	1.657823	1.672373	0.476578	0.178085	0.305521	0.899748	
2	1.264763	1.660228	0.481008	0.180675	0.307208	0.900172	

tokenizer_config.json: 100%25.0/25.0 [00:00<00:00, 1.83kB/s]

config.json: 100%482/482 [00:00<00:00, 26.4kB/s]

vocab.json: 100%899k/899k [00:00<00:00, 1.13MB/s]

merges.txt: 100%456k/456k [00:00<00:00, 765kB/s]

tokenizer.json: 100%1.36M/1.36M [00:00<00:00, 1.70MB/s]

model.safetensors: 100%1.42G/1.42G [01:13<00:00, 20.6MB/s]

```
res = learn.show_results(learner=learn)
```

text	target	prediction
<pre>summarize:daz dillinger yeah daz dillinger finna take gangsta shit niggas even know finsta know call gangsta rap niggas day one tray deee lookie lookie motherfuckers advise ghetto livin struggling day day tryin flip fifty dollar double buy bag weed fresh pair khakis take bitch movie ass niggas get money man get paid motherfuckers cut welfare goin eat cheat sons bitches done even made game young black niggas lives striking left right motherfucking end sight fuck sell dope get jack get sag put rag get motherfucking thang niggas cause goddammit hell bout youngster techniec ride nigga bout get homies nigga right tray deee yeah motherfucker techniec come bad azz let dip nigga nigga cuzz nigga nigga nigga nigga homie called big shot today</pre>		[\n\nThe song "Gangsta Rap" by Daz Dillinger is a gritty and aggressive track that showcases the rapper's lyrical prowess and street mentality. The lyrics depict a life immersed in gangsta rap, with references to violence, drug dealing, and criminal activities. Dillinger asserts his dominance in the rap
<pre>learn.metrics = None learn = learn.to_fp32() learn.export(fname="bert_summary_export.pkl")</pre>		
<pre>infer = load_learner(fname="bert_summary_export.pkl")</pre>	Baggin Ass Criminal" by	unafraid to take on any challenge that comes his
<pre>song=test_df.lyc[5] ground_truth=test_df.summ[5]</pre>	glorifies the gangsta	celebrates the lifestyle of a pimp. The lyrics boast
<pre>song</pre>	the struggles of living in	staying true to oneself and not succumbing to
<pre>'intro rza hold stop stop goin goin goin goin holdin goin goin goin holdin goin goin goin holdin verse rza could soar sky like bird w ould disturbed see brothers curb smoking herb cops third precinct rush like herd charge delinquent possession greets arrest strip sea rch undress make youngster come sign confession would watch cop chopper high powered binocular cameras attached bottom aircraft takin photographs lab pointin millimetre shot magazine full clip gat back hovering projects spotlight shining bright moving targets infrare d light night dark objects harassing citizens charged yet would order flock pelicans jam propellers top see forfeit death system fail ure take devils eyewitness tom brokaw mentally fry brain blow smoke killa bee presidential suite bill hillary bow kiss feet auxillary security retreat heavy artillery sun heat meet agony defeats cavalry six pack battery back anti gravity raps snaps head inside bear t rap cannot adapt habitat bigger jigsaw puzzle world map niggas think...' daz dillinger shit known dogg pound criminal verse four snopafly spot hot hit switch</pre>	showcases the bravado,	on themes of success, staying true to oneself, and
<pre>infer.blurr_summarize(song)</pre>		
<pre>infer.blurr_summarize("Greetings loved ones Let's take a journey I know a place Where the grass is really greener Warm, wet n' wild There</pre>		
<pre>ground_truth</pre>		
<pre>'\n\nThe song "What's Goin On" by Gravediggaz delves into various societal issues and questions the state of the world. The lyrics t ouch on themes of police brutality, racial profiling, government surveillance, and the manipulation of power. The verses paint a vivi d picture of a corrupt system that preys on the vulnerable and perpetuates injustice. The song also emphasizes the need for awareness and resistance against oppressive forces. Overall, it serves as a call to action, urging listeners to question the status quo and see k change in order to create a more iust and equitable societv.\n\n\n'</pre>		