

## Parameters in Generate function

I tried out different parameters and an input of 10 random consecutive messages. (I could not put more because there is a limit of 512/1024 input token in the function). My findings for each of the parameter are as follows:

1. **max\_length**  
The best length I found was 30 more than the input token list size. This way repetition is also avoided and only 2-3 new messages get printed, which is optimal.
2. **num\_beams**  
Changing this parameter does not improve the output string and in certain cases decreases its quality. So, I prefer not to use it.
3. **no\_repeat\_ngrams**  
This parameter also cannot be used as each new message start with the name followed by a colon. On using no\_repeat\_ngrams, every message starts with a different name, and so this cannot be used.
4. **top\_k and top\_p**  
After some experimenting, I found the best top\_k and top\_p values to be 10 and 0.1 respectively. With do\_sample set to True. It produced meaningful and diverse results with very less repetition.
5. **temperature**  
Changing it doesn't affect the output

Final parameters

```
output = model.generate(  
    encoded_input['input_ids'],  
    max_length=maxl,  
    top_k=10,  
    top_p=0.10,  
    do_sample=True,  
)
```