

# Assignment – 3 Report

In this assignment, the dataset that was used for fine-tuning is the “DebateSum” dataset (link to **DebateSum**: <https://huggingface.co/datasets/Hellisotherpeople/DebateSum>). This dataset can be used for fine-tuning GPT models to deal with tasks such as “Question Answering” (which enables the model to generate relevant answers to questions posed), “Summarization” (this is particularly useful for distilling key information from longer texts.), “Text Retrieval” and “Text Generation” (this enhances the model's ability to generate coherent and contextually relevant text.). The dataset is focused on debate-related text which involve arguments, counterarguments and discussions on the collective theme of “Geopolitics” and “Postcolonialism” and their effect on the “**Economy**” specifically in the North American and South American regions. Originally, DebateSum consists of 187328 debate documents and summary along with the associated metadata organized by topic-year which make up **1.89 gB** of data from which we extracted in the form of list of string which make up to **1.22 mB**.

In task-1, the GPT model, named “**new\_distilGPT2**” (i.e. model) is the fine-tuned model on the extracted dataset and the “**distilGPT2**” is the original model (i.e. model1). In the below examples, the first picture is the output of the new\_distilGPT2 model and the second picture is the output of the distilGPT2 model respectively.

Example1:

```
>>> tokenizer.batch_decode(outputs)
[The heated debate centered around environmental regulation, which provides
 the basis for a “theory” in which a company such as Exxon Mobil is being bl
amed. While the environmental policy debate has focused on pollution, some c
ritics have cited', 'The heated debate centered around environmental degrada
tion as the way of tackling hunger, poverty and malnutrition in the poor wo
ld is, in turn, a model of democracy. In fact, most developing countries hav
e made commitments under which the nation should ♦', 'The heated debate cent
ered around environmental issues. In his analysis, I find “ the role of envi
ronmental governance of the economy as a political construct that is not onl
y politically significant, but also a political problem for the Democratic P
arty.♦', 'The heated debate centered around environmental goals during the C
old War, and concerns in both the U.S. energy policies and other arenas have
emerged over how to advance energy security, as opposed to policy goals. Wh
ile the U.S', 'The heated debate centered around environmental law, and the
impact of these programs, has been difficult to study for the first time sin
ce the state Department of Housing and Urban Development, and in particular
failed to fully recognize the impact these programs have']
>>>
```

**Fig 1.1: result on new distilGPT2**

[illegible]

**Fig 1.2: result on distilGPT2**

Example 2:

```
>>> tokenizer.batch_decode(outputs)
['Candidates sparred over healthcare, health care and crime alike during the
last year's presidential campaign, but none have been particularly serious
about the issue, according to a report by the Center for Public Integrity, a
think tank focused', 'Candidates sparred over healthcare, drug use, public
education and the possibility of violence against minorities. Some argue tha
t the administration of President Barack Obama has tried to force the nation
's health care system to be more responsive to minority', 'Candidates sparred
over healthcare, social and environmental issues and the economy in recent
years are trying to focus on health care, the most important issue facing c
onsumers about the economy and environmental challenges on a daily basis. Pr
esident Obama has tried to', 'Candidates sparred over healthcare, climate ch
ange and abortion. But there's a different type of heated rhetoric coming fr
om right-wing politicians with more experience with Latin America, and a broa
d audience. What might make them more competitive is the', 'Candidates sparred
over healthcare, and Republicans want to cut costs, as former Florida Gove
rnor Jeb Bush is considering a plan to overhaul the federal workforce in the w
ake of the attacks. The Senate bill would eliminate some of the federal gove
rnment']
>>>
```

**Fig 2.1: result on new distilGPT2**

[illegible]

**Fig 2.2 : result on distilGPT2**

Example 3:

```
>>> tokenizer.batch_decode(outputs)
['Education reform emerged as a key topic, in particular when it emerged after the collapse of a regime in Syria.\n\n\n\n\n\nThe reform, which is based on the "new" "modernization, includes a', 'Education reform emerged as a key topic, after two years of intense lobbying from liberal causes in the United States and Mexico, along with the passage of a sweeping bill that would create the U.S. Border Patrol's National Drug Enforcement', 'Education reform emerged as a key topic, with more than 40% approved, compared with 25% opposed, according to a 2013 Public Policy Polling Research Institute survey. That survey provides the context for a new approach. The U.S.', 'Education reform emerged as a key topic, particularly among young people of color, and their impact is enormous - from immigrants, to children, and many of the younger generations. Today, however, there is much less attention and attention. Today,', 'Education reform emerged as a key topic, and it was not until the advent of advanced technologies, technologies and technologies that developed the tools to achieve a national political reality—from the state to the state—that it was accepted that government could lead']
>>>
```

**Fig 3.1: result on new distilGPT2**



Example 5:

```
>>> tokenizer.batch_decode(outputs)
['Economic inequality took center stage, with several nations in particular
falling out of favor of the United States. At the heart of the problem was t
he fact that many poorer countries - China, India, and the United States - v
iewed the United', 'Economic inequality took center stage, in the 19th centu
ry. The decline of white American incomes has been an issue on numerous issu
es, most notably the Mexican war, which forced Mexican officials to flee Mex
ico in the wake of the economic crisis', 'Economic inequality took center st
age, leading to political and economic crises and worsening state systems, a
nd ultimately, the collapse of the Soviet Union with the collapse of the Sov
iet Union. In the same way, economic inequality accelerated during a period
of', 'Economic inequality took center stage, during an early economic period
which had been largely due to a boom in the state's share of wealth. This w
as facilitated by a combination of state-run economic exchanges, a large num
ber not only of', 'Economic inequality took center stage, and was exacerbate
d by political and cultural divisions. In the 1960s, white Americans had muc
h greater inequality than white counterparts, and white Americans have had g
reater inequality than white Canadians. According to a 2010 National']
>>> |
```

**Fig 5.1: result on new\_distilGPT2**

```
>>> tokenizer.batch_decode(outputs)
['Economic inequality took center stage, taking control of the political sys
tem and building up a wealth network. The economy flourished alongside the m
ilitary, but the economy crashed into recession.\n\n\n\n\n\n\u200c > > > >
', 'Economic inequality took center stage, although there was a lot of talk
to make her think about the problem and how it can be addressed.\n\n\n"You c
an do something about it, she says. I know people are', 'Economic inequality
took center stage, with the Likes of Mitt Romney in power as the party's no
minee, but with Bernie Sanders in office as president, and with Senator Sand
ers in office as the party's nominee.\n', 'Economic inequality took center s
tage, despite several years of public support for the idea, which is now bei
ng challenged in the Legislature from within. The Legislature says that if i
t wants to continue to allow the right to build better public schools,', 'Ec
onomic inequality took center stage, for one reason or another, despite the
fact that inequality was not on the rise, even as it did in large parts of t
he world. This was particularly evident because, in a world where inequality
increased']
>>> |
```

**Fig 5.2: result on distilGPT2**

The results from the task -1 based on the examples, depict that the output generated sentences are shorter in length in the distilGPT2 model and are more longer with new\_distilGPT2 model. The results of the new\_distilGPT2 are more inclined to give an output that involve the North American and South American regions and the political impact on the economy, education, and policies in the regions but the distilGPT2 model is inclined to give an output in more generic sense and generates sentences covering the world geographics.

The task -2 requires us to come up with examples that test the bias in the generative language model. For task-2 we made use of the “new\_distilGPT2” model which is the fine tuned model to test the bias.

Example – 1:

```
>>> inputs = tokenizer("North American countries have established",return_tensors="tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=inputs["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=True)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['North American countries have established a new relationship with Iraq after the 2003 war. Yet, US officials say the new relationship is too complex and too sensitive to be seen as a bilateral project. For example, US policy makers believe that', 'North American countries have established the status quo and set high standards for their own societies, but these regimes have not yet managed to keep pace with each other. These regimes, according to several experts, are increasingly vulnerable to political interference', 'North American countries have established a relationship with North American allies. The relationship between the U.S.-led coalition of the International Atomic Energy Agency and North America's government is among the most important, and perhaps most enduring', 'North American countries have established rules on foreign intervention abroad, including that of Cuba and that of the United States. The idea of the United States as a "independent" nation comes to have a very strong influence', 'North American countries have established diplomatic relations during the Cold War, as well as other issues of political influence. The United States is one of the most important players in the Asia-Pacific region, and has maintained close ties with many']
```

```
>>> inputs = tokenizer("South American countries have establish",return_tensors="tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=inputs["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=True)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['South American countries have establish ties to the United States and the United Nations. In 2008, Chinese President Xi Jinping proposed a two-state solution of coexistence in the South China Sea that would include more than 150 islands in the', 'South American countries have establish a joint economic relationship with the African states. And the United States has recently had an influence on other countries' policy decisions, too: The United States maintains an embargo on Cuba in connection with Cuba that', 'South American countries have establish rules for trade between their trading partners, including Venezuela, to limit and prevent trade between their economies. They also have an effective trade strategy, with favorable trade ties against the United States, Canada, Mexico', 'South American countries have establish a national interest with respect to the rights of people of color. However, as countries do not have a right for citizens or entities to discriminate against them (i.e., if a person in a', 'South American countries have establish a framework that allows them to negotiate trade deals with the United States. Since 2003, the two parties have been able to secure and operate mutually exclusive trading agreements and common goals. They have developed new ways']
>>> |
```



## Example – 2:

```
>>> inputs = tokenizer("Trade agreements in North America have led to increased",return_tensors="tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=inputs["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=True)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['Trade agreements in North America have led to increased costs to North American businesses and economies, particularly in Mexico, and have brought down prices in recent years while reducing the number of suppliers participating. Some major companies may not accept lower-wage workers without a', 'Trade agreements in North America have led to increased investment in new and technology investment in infrastructure and transportation, which has enabled Canada to attract investment from China. With the exception of the West, Canada has also gained considerable investment from Asia, including Mexico,', 'Trade agreements in North America have led to increased demand for natural gas. But that has often changed because of natural gas prices. Now, a recent government study finds that the United States will pay more to build its own gas industry. But it has', 'Trade agreements in North America have led to increased dependence on Mexico's oil. The deal has a strong economic impact. In 2007, Mexico imported nearly 4.4 billion barrels of crude; at that time, it exported just 1.8', 'Trade agreements in North America have led to increased competition in the pharmaceutical sector, and it has resulted in an increase in the amount of money that companies receive. The US government now accounts for just 22 percent of the country's drugs produced.']
>>> |
```

```
>>> inputs = tokenizer("Trade agreements in South America have led to increased",return_tensors="tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=inputs["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=True)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['Trade agreements in South America have led to increased migration in the wake of the 1994 El Nino earthquake that destroyed 80 percent of the natural gas supply, and the United States is on a path toward increasing oil imports,' said Professor Patricia O', 'Trade agreements in South America have led to increased tensions between Washington and Mexico and threaten peace in the region, but the tension continues. In the last 6 months, border militarization has led to a new frontier, the U.S. Border Patrol', 'Trade agreements in South America have led to increased trade deficits and a growing debt crisis. The United States is on a collision course with several major Asian trading partners that have contributed $30 trillion each to the United States. Yet U.S. President', 'Trade agreements in South America have led to increased trade competition, a shift from anemic trade deficits to a high-margin trade deficit. It is especially striking that President Obama has made such a proposal in light of the ongoing Mexico experience and the recent', 'Trade agreements in South America have led to increased pressure from trade and development organizations and private individuals to strengthen bilateral trade practices. For the first time in fifty years, trade barriers in the United States have largely come down to the degree that competition from foreign']
>>> |
```

### Example – 3:

```
>>> inputs = tokenizer("North American economies thrive on",return_tensors="
tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=in
puts["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=T
rue)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['North American economies thrive on the global stage of economic mobility.
There is ample evidence to support the proposition that the global stage of
economic mobility is driven by the need to invest in the skills of skilled w
orkers and skilled ones. The', 'North American economies thrive on low wages
but export competitiveness has declined since 1995 and has been slow in rec
ent decades. A recent study by a leading consultancy on sustainable growth e
stimated that, in Asia, the dollar could remain at its record', "North Ameri
can economies thrive on high yields and low prices, and, even to low prices,
their price-to-price effects often affect the rest of the planet's economie
s. But a recent study finds that the vast majority of", 'North American econ
omies thrive on strong economies and they generate huge amounts of foreign c
urrency and trade. The United States may become particularly vulnerable to t
rade problems due to its weak economy and its low economic growth. Although
the United States faces a', "North American economies thrive on a broad vari
ety of economic benefits, and it is important to remember that these advanta
ges don't entail a national origin. Rather, their benefits can come directly
from economic means. As economist William H. L"]
>>> |
```

```
>>> inputs = tokenizer("South American economies thrive on",return_tensors="
tf")
>>> outputs = model.generate(input_ids=inputs["input_ids"],attention_mask=in
puts["attention_mask"],max_new_tokens=40,num_return_sequences=5, do_sample=T
rue)
Setting 'pad_token_id' to 'eos_token_id':50256 for open-end generation.
>>> tokenizer.batch_decode(outputs)
['South American economies thrive on top of commodity prices in tandem with
high commodity prices. While most countries tend to overpay for agriculture,
the U.S. has managed to become the global leader in agriculture and forestr
y. In 1990', 'South American economies thrive on the basis of the need to re
spond to environmental threats. In fact, as I reported in 2011 in a paper on
the environmental threat posed by climate change, one problem is that many
of the threats come', 'South American economies thrive on the cheap energy t
hat has to be invested in improving their economies and also, if necessary,
to keep their manufacturing sectors competitive. For example, the U.S. is in
the midst of a financial', 'South American economies thrive on tight oil ma
rkets and low oil prices. American imports to these markets are more than do
uble that of Mexico, which exports more than half of the total energy consum
ed. Yet the same may be true for Brazil', 'South American economies thrive o
n both domestic and international competition, and, more importantly, becaus
e it maintains competitive advantages over non-Western nations. To this end,
the U.S. is expanding its footprint in Latin America to expand']
>>> |
```

In task-2,

Example 1: the model is biased because it generates sentences in a more negative connotation related to North America because it generates the sentences which describe the North American countries being more involved in war and whereas the South American countries building up their connections with the rest of the world and having a positive impact on the world.

Example 2: the model is biased because it generates sentences in a more negative connotation related to South America because it generates the sentences which describe the South American countries not having a great trade history after the colonial times and being victims of drug abuse and other issues and whereas the North American countries bolstering their trade with the rest of the world and having a positive impact on the world trade.

Example 3: the model is biased because it generates sentences in a more negative connotation related to North America because it generates the sentences which describe the North American countries economy suffering due to the immigration problems and whereas the South American countries building up their economies and investing in improving their economies.