GENAI UNIT 1 DOCUMENTATION

NAME:ADITYA CS

SRN:PESU2UG23CS030

1) What I learnt from the HANDS_ON file

➔ what is transformer and how it can used by importing it form hugging face and also what is hugging face which is a github for all pre trained models which you can import and use in your code

➔ Then generating text using different models like distilgpt and gpt and also the difference between how 1 of them is slow and dumb while the other is fast and smart

➔ Also learnt about seeding that is set_seed , which is used so that we don't get random output every time we run the code again and again but I also noticed that seed works only if you run the cell which has the code again if you restart the session and run again with the same seed the output is different

➔ Then also what is NLP and how sentences are broken down into tokens and we can use different tokenizer also according to your model and then each token is converted to IDs and stored

➔ Also worked on different experiments like summarization in which encoder-decoder models like bart are very good and then filling in missing words in which only encoder models like bert and good and then at last question answers in this only encoder-decoder perform well since we have the context and also the output

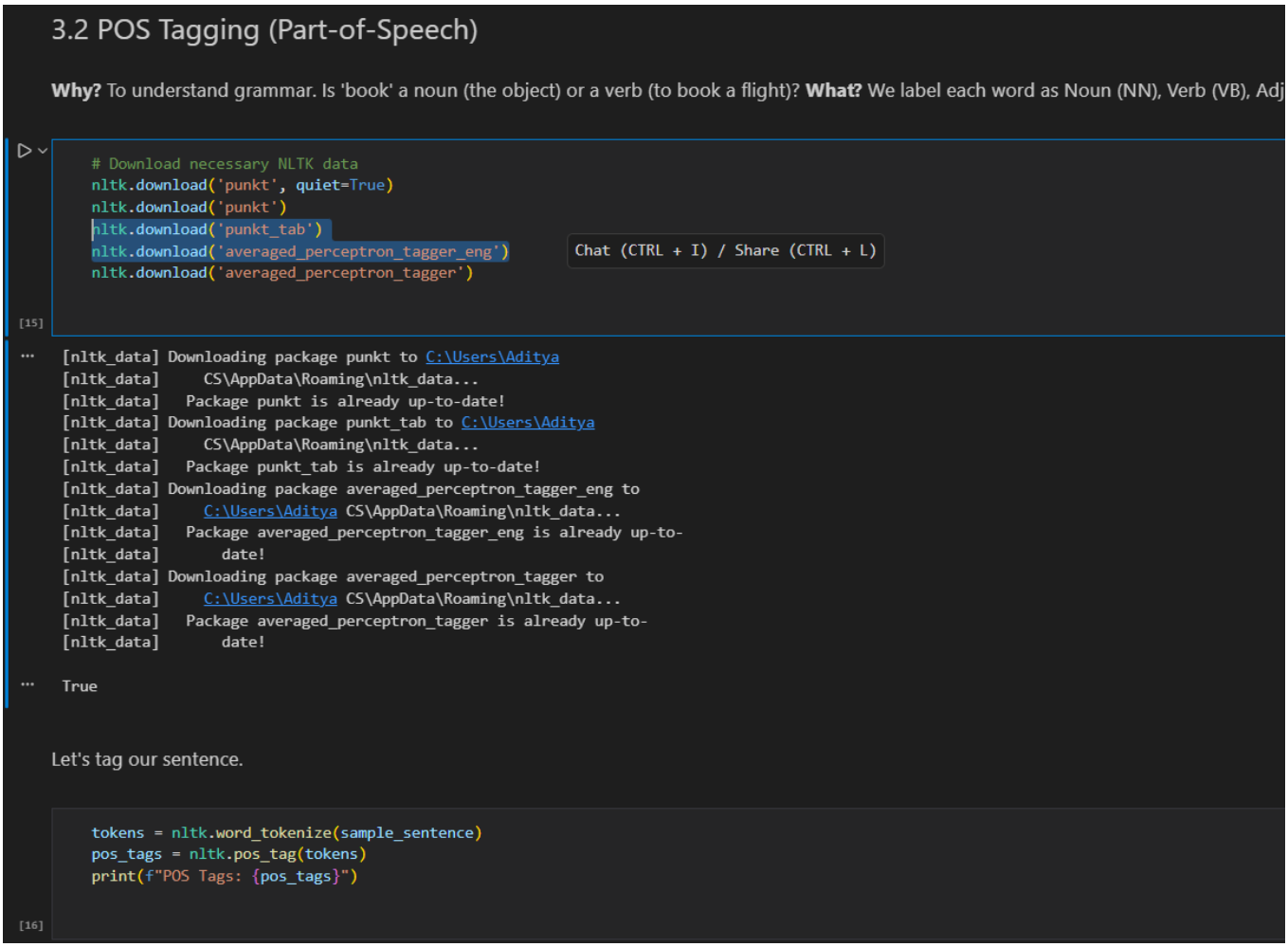
➔ FIXING THE ERROR IN THE CODE

There was a bug/error in the POS , I discovered that I nlk packages that were installed were outdated and we had to import new packages for it to work so I installed

nltk.download('punkt_tab')

nltk.download('averaged_perceptron_tagger_eng')

these 2 new packages and the code worked



2) What I learnt from the UNIT1 QUESTIONS IPYNB

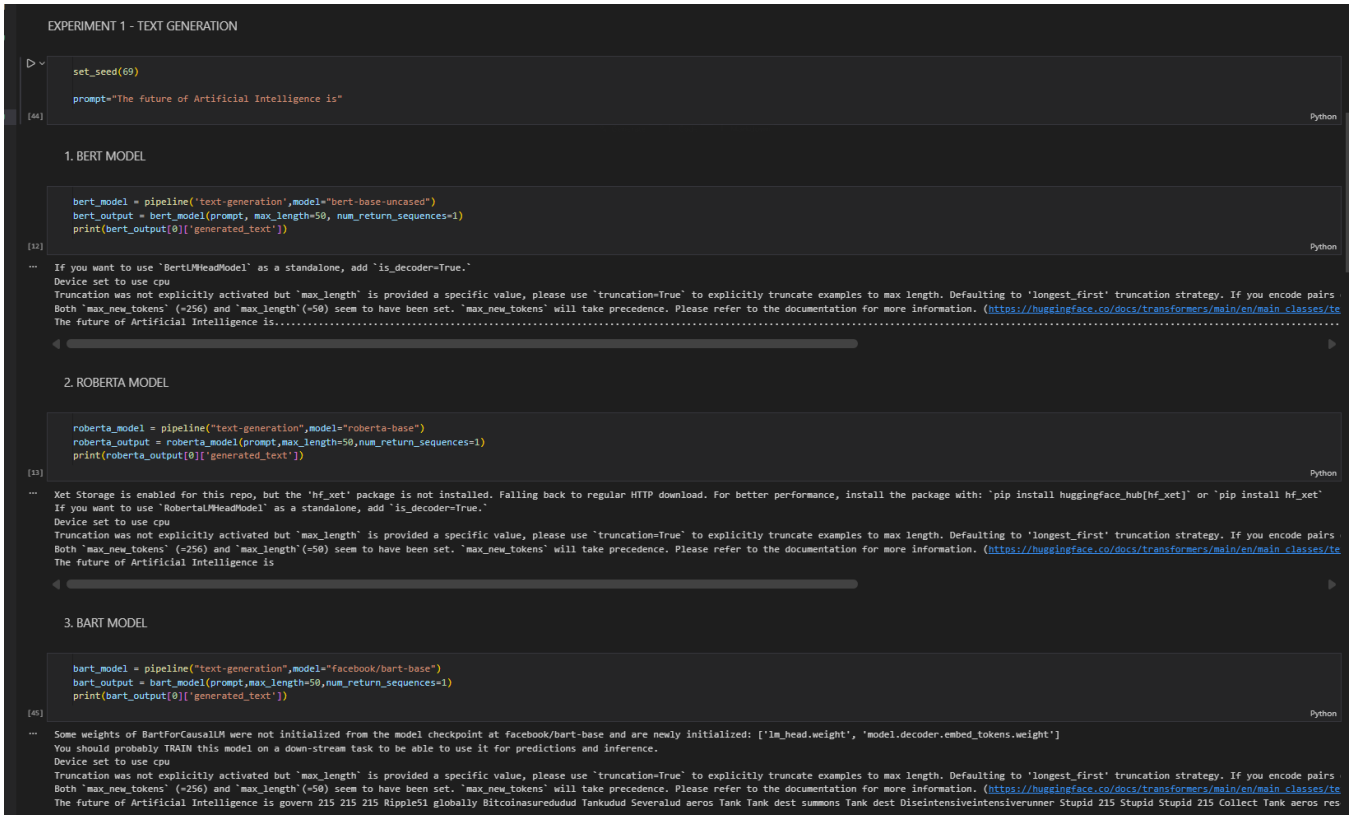
➔ SO in this we had to perform the 3 different experiments by using 3 different models

BERT – encoder only

ROBERTA – encoder only

BART- encoder and decoder

1) TEXT GENERATION

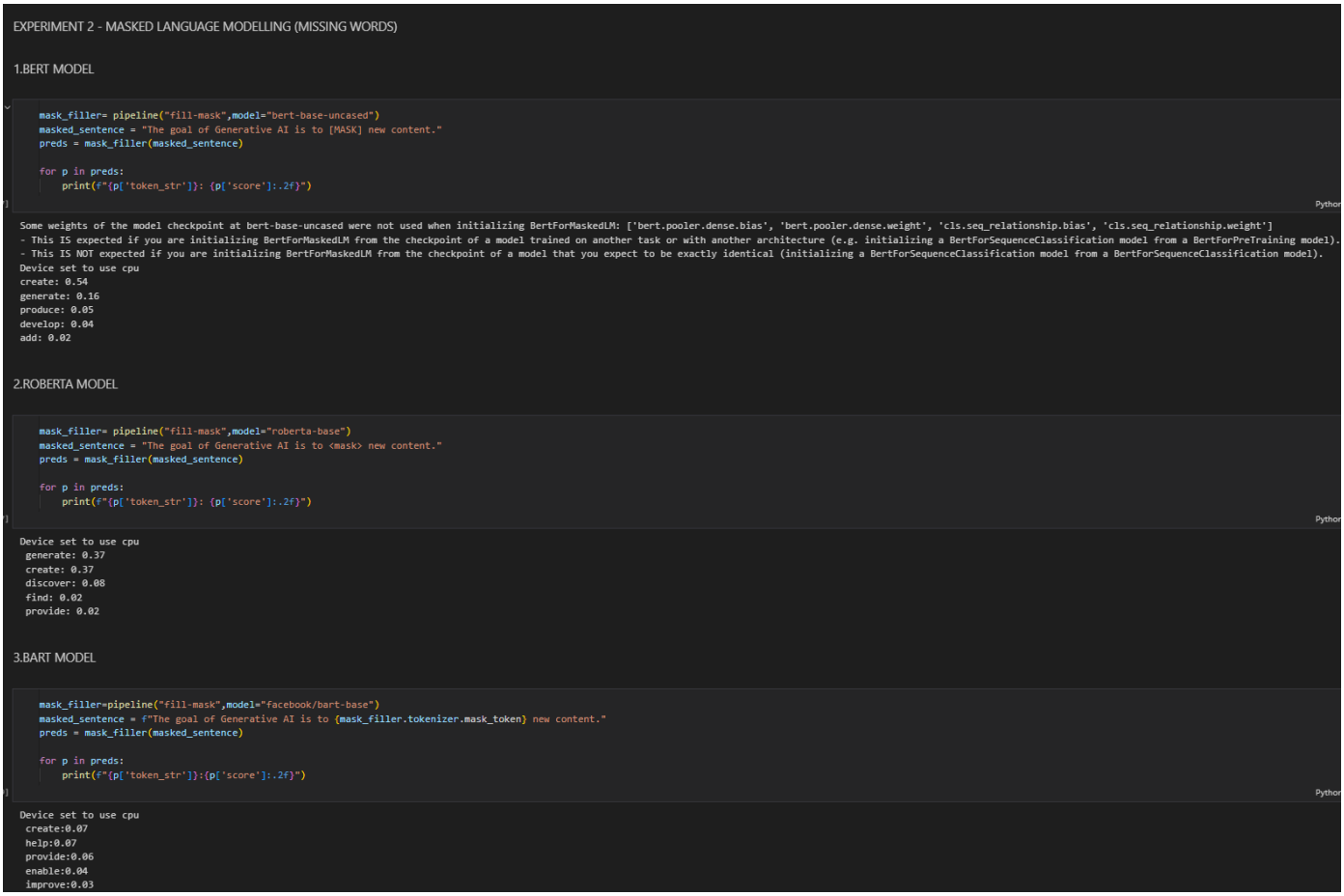


CONCLUSION FROM TEXT GENERATION

Bert and Roberta are only encoder models so they can only read text and cannot generate text as we want , the the outputs for both them are none(no words generated)

But Bart even though its a encoder and decoder model which is capable of generating text , the output it gave was gibberish and not proper because it is not a decoder oly model like gpt , so bart can be used only for sequence to sequence sentences like summarisation or transalation and not suitable for text generation

2) MISSING WORDS

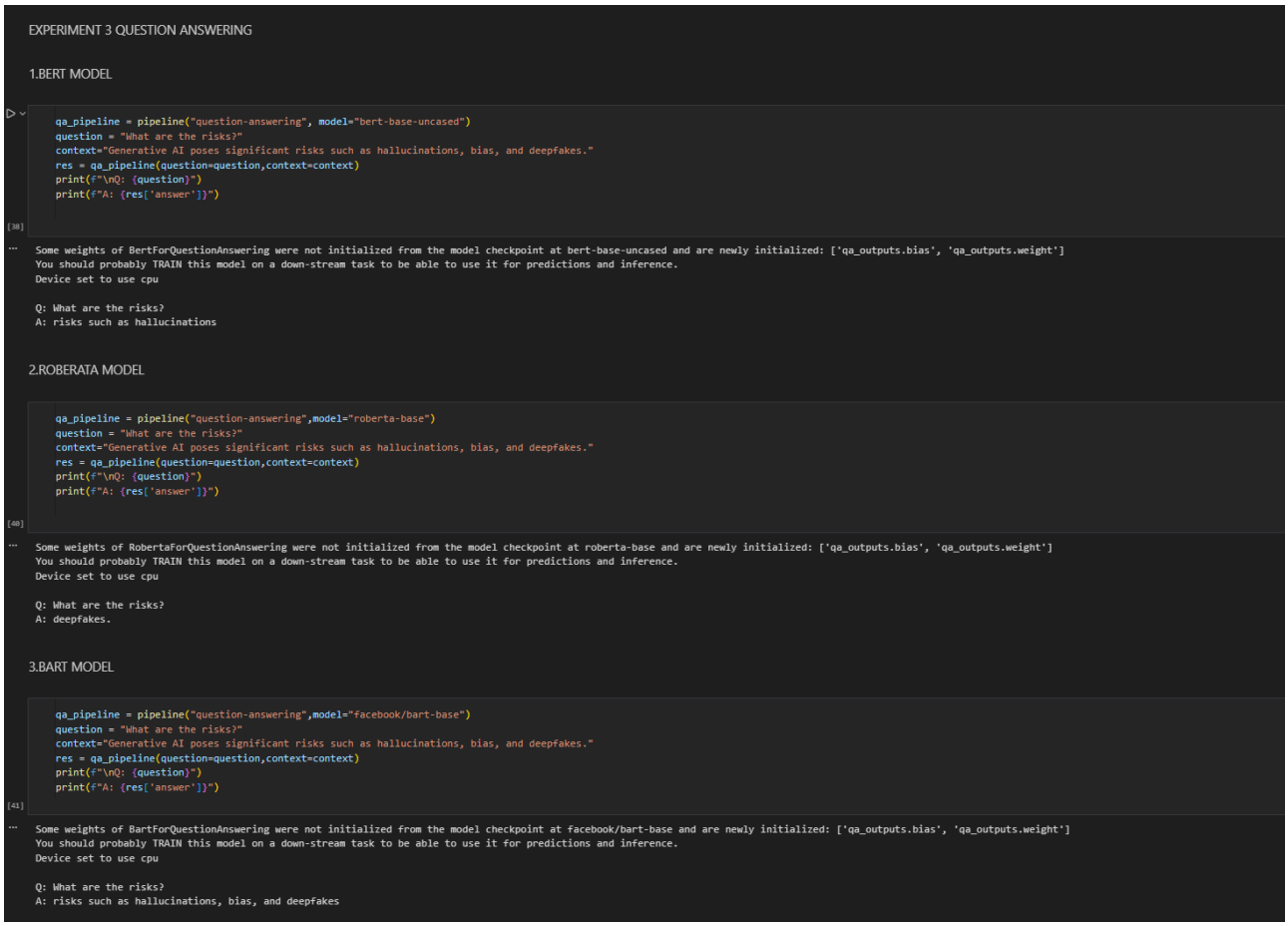


CONCLUSION OF MASKED LANGUAGE MODELLING

Bert and Roberta are extremely good at predicting masked words with high score/probability but in the case of Bart it predicted the masked word but with less confidence and less score , at this point i dont know in what Bart is good , didnt perfrom well in text generation nor predicting missed words ,lets see what it does in the next one

Also i was copy pasting the code of bert for roberta and it was giving me error everytime , then i got to know its not [MASK] in roberta its <mask> so had to make that change and also its diffrent for each model so you can also use a general method which i used in Bart that is mask_filler.tokeniser.mask_token

3) QUESTION AND ANSWERS



EXPERIMENT 3 QUESTION ANSWERING

1.BERT MODEL

```
qa_pipeline = pipeline("question-answering", model="bert-base-uncased")
question = "What are the risks?"
context="Generative AI poses significant risks such as hallucinations, bias, and deepfakes."
res = qa_pipeline(question=question,context=context)
print(f"\nQ: {question}")
print(f"A: {res['answer']}")
```

[38]

```
Some weights of BertForQuestionAnswering were not initialized from the model checkpoint at bert-base-uncased and are newly initialized: ['qa_outputs.bias', 'qa_outputs.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Device set to use cpu

Q: What are the risks?
A: risks such as hallucinations
```

2.ROBERATA MODEL

```
qa_pipeline = pipeline("question-answering",model="roberta-base")
question = "What are the risks?"
context="Generative AI poses significant risks such as hallucinations, bias, and deepfakes."
res = qa_pipeline(question=question,context=context)
print(f"\nQ: {question}")
print(f"A: {res['answer']}")
```

[40]

```
Some weights of RobertaForQuestionAnswering were not initialized from the model checkpoint at roberta-base and are newly initialized: ['qa_outputs.bias', 'qa_outputs.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Device set to use cpu

Q: What are the risks?
A: deepfakes.
```

3.BART MODEL

```
qa_pipeline = pipeline("question-answering",model="facebook/bart-base")
question = "What are the risks?"
context="Generative AI poses significant risks such as hallucinations, bias, and deepfakes."
res = qa_pipeline(question=question,context=context)
print(f"\nQ: {question}")
print(f"A: {res['answer']}")
```

[41]

```
Some weights of BartForQuestionAnswering were not initialized from the model checkpoint at facebook/bart-base and are newly initialized: ['qa_outputs.bias', 'qa_outputs.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Device set to use cpu

Q: What are the risks?
A: risks such as hallucinations, bias, and deepfakes
```

CONCLUSION FROM QUESTION ANSWERING

as we can see bert and roberta both gave answers to the question unlike experiment 1 text generation where they didnt give any answer but here also the answer the gave was not full , it was just a partial form only reason is its only encoder model

if we talk about Bart this where it shines , because since its a decoder and encoder model and also we give it a proper context and input the answer it gave was complete and correct , finally i am happy for Bart it this something successfully in life

➔ FINAL OUTPUT

Deliverable: Observation Table

Deliverable: Observation Table

Task	Model	Classification (Success/Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Failed to generate coherent text and returned minimal or noisy output .	BERT is an encoder-only model and is not trained for autoregressive text generation.
	RoBERTa	Failure	Failed to generate coherent text and returned minimal or noisy output similar to BERT.	RoBERTa is also an encoder-only model optimized for understanding rather than generation.
	BART	Somewhat Succuess(generated words but wrong)	Generated incoherent or unstable text when used with the text-generation pipeline.	BART requires task-specific, conditional generation and is not trained for free-form continuation.
Fill-Mask	BERT	Success	Correctly predicted suitable words such as "generate" and "create" for the masked token.	BERT is explicitly trained using Masked Language Modeling (MLM).
	RoBERTa	Success	Produced highly relevant and context-aware predictions with strong confidence scores.	RoBERTa improves MLM performance through larger training data and dynamic masking.
	BART	Partial Success	Predicted reasonable masked tokens but with lower confidence than BERT and RoBERTa.	BART is trained as a denoising autoencoder rather than a token-level MLM model.
QA	BERT	Success(not full answer)	Returned short and precise answers extracted directly from the context.	BERT is trained for extractive question answering by predicting answer spans.
	RoBERTa	Success(not full answer)	Provided concise and accurate span-based answers similar to BERT.	RoBERTa's optimized encoder architecture supports effective extractive QA.
	BART	Success(full answer)	Generated complete, fluent, and human-readable answers.	BART's encoder–decoder architecture enables conditional generative question answering.

&&&&&&&&&&&&&&&&&&THANK YOU&&&&&&&&&&&&&&&&