DtP-Week-3 Graded Assignment -3

1) Download the Yelp review dataset "Yelp/yelp_review_full". Split each sample by calling the string method ".split()" and choose the correct statements about the dataset.	2 points
The dataset contains close to 99 million words	
here are more than 300 samples that contain a single word	
C. There are less than 300 samples that contain only a single word """. "Cheesy-melty-roasted-cauliflower-with-fresh-bread-crumbs-on top.\\nTo-die-for." is one of the single words in the dataset	
Cheesymensynostect-administrative microstructurings on top, (introduction, is one of the single words in the dataset	
The versage region of a sample is rout.	
4 NB	
9	
_ VE	
F F	
Partially Correct.	
Score: 1.2	
2) Load the "bert-base-uncased" pre-trained tokenizer and choose the correct statements about the tokenizer.	2 points
2) Load the "bert-base-uncased" pre-trained tokenizer and choose the correct statements about the tokenizer. The tokenizer is used for the BERT model with the context length of 512 The tokenizer has 5 special tokens The tokenizer has 5 special tokens	z ponto
o. The tokenizer has 5 special tokens	
C. Tokenizing a sample that contains more than 512 words would result in truncation of all tokens beyond the length 512	
D. Tokenizer inserts all the special tokens when it processes a single sample as an input	
F. Tokenizer inserts only [CLS] and [SEP] special tokens when it processes a single sample as an input F. Tokenizer inserts only [CLS] special token when it processes a single sample as an input Que Tokenizer inserts only [CLS] special token when it processes a single sample as an input	
F. Tokenizer inserts only [CLS]special token when it processes a single sample as an input	
added automatically	
added authority	
A STATE OF THE STA	
c c	
_ D	
€ E	
□ F	
3) Use "BertConfig" and "BertForMaskedLM" to construct the default (original) BERT model. Choose the correct statements	1 point
The moder has 12 Bert layers	
B. The model has 6 Bert layers GPF 2 has 12 SUL attent weaks	`
C. The model uses abso <u>lute position embeddings</u>	- ^\
D. The word embedding (token embedding) layer has about 23 million learnable parameters	~~
3) Use "BertConfig" and "BertForMaskedLM" to construct the default (original) BERT model. Choose the correct statements The model mass 12 Bert layers B. The model has 6 Bert layers C. The model uses absolute position embeddings D. The word embedding (token embedding) layer has about 23 million learnable parameters E. The total number of parameters in the model is close to 110 million	
$\sqrt{1 + 1}$	
10, 100	
111.435 Caleula is week-4	/ . a . \
S 117 (1 C V Calina is well 1	(L 2 -1
- 11 6 .453	()
- 116.433 - 1	
to be exact	
to be enact	
to be exact 4) Double the context length from 512 to 1024 (Ny can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default	
d) Double the context length firm 512 to 1024 (b) u can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration.	
to be exact 4) Double the context length fire 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 1.39 1.30	
to be exact 4) Double the context length fire 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 1.39 1.30	
to be exact 4) Double the context length fire 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 1.39 1.30	nuch
to be exact 4) Double the context length fire 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 1.39 1.30	nuch.
4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. Output Ves, the answer is correct. Score: 1 Lete no of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. Answer is correct.	nuch.
to be exact 4) Double the context length fire 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 1.39 1.30	nuch.
4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. Output Ves, the answer is correct. Score: 1 Lete no of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. Answer is correct.	nuch.
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers:	nuch.
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers:	
4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41	
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure	
to be exact 4) Double the context length from 512 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples	
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
4) Double the context length from \$12 to 1024 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk! if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the variable "ds_chunked". Enter the total number of samples in the new dataset.	1 point
to be exact 4) Double the context length from 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask	1 point
4) Double the context length from \$12 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the Variable 'ds_chunked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds_map()' for theanswer to match.	1 point
4) Double the context length from \$12 to 1024 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk! if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the variable "ds_chunked". Enter the total number of samples in the new dataset.	1 point
4) Double the context length from \$12 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the Variable 'ds_chunked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds_map()' for theanswer to match.	1 point
4) Double the context length from \$12 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the Variable 'ds_chunked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds_map()' for theanswer to match.	1 point
4) Double the context length from \$12 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12 5. Drop the last chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the batches in the Variable 'ds_chunked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds_map()' for theanswer to match.	1 point NO
4) Double the context length fig. 512 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of 512 5. Drop the least chunk if its length is less than 512 6. Pack all the chunks 7. Iterate over all the batches in the dataset 8. Total the country of the variable "ds_chunked". Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling "ds.map()" for theanswer to match.	1 point NO
4) Double the context length for \$12 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 4.50	1 point NO
4) Double the context length from \$12 to 1024 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize for toge input IDs and attention mask 3. Concatenate all the input IDs 4. Chunk the concatenated IDs into a size of \$12\$ 5. Drop the last chunk if its length is less than \$12\$ 6. Pack all the chunks 7. Iterate over all the batches in the dataset Store the resulting dataset in the variable "ds_chunked". Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling "ds.map()" for theanswer to match. 246695 6) Split the new dataset into training and test sets with the test_size=0.05 and seed=42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data location in the context of the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the masking probability to 0.2. Use the data location in the context of the context	1 point NO
4) Double the context length for \$12 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 4.50	1 point NO
4) Double the context length from \$12 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokerize it to pet input IDs and attention mask 3. Concatenated IDs into a size of \$12 5. Drop the list chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the bipatch size in the dataset 8. There is no concatenated IDs into a size of \$12 6. Pack all the chunks 7. Iterate over all the bipatch gataset in the variable "Sac-Dunked". Enter the total number of samples in the new dataset. Note: the batch are should be kept at 1000 while calling "ds.mag()" for theanswer to match. 4. Observed the sound of the dataset in the dataset into training and test sets with the test. size 0.05 and seed-42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data to PyTorch to load a batch of samples, and enter the token ID corresponding to the unmasked token 100 Yes, the answer is correct.	1 point NO
4) Double the context length from \$12 to 1024 (b) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 Yes, the answer is correct. Score: 1 4. Cauttext length, Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input IDs and attention mask 3. Concarenate all the input IDs 4. Churk the concarenated IDs into a size of \$12 5. Drop the lact think if its length is less than \$12 6. Pack all the churks 7. Iterate over all the batches in the dataset Store the resulting dataset in the variable "ds. churked". Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling "ds.map()" for theanswer to match. 4. Split the new dataset into training and test sets with the test_size=0.05 and seed-42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data loe PyTorch to load a batch of samples, and enter the token ID corresponding to the unmasked token	1 point NO
4) Double the context length for \$12 to 1024 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input libs and attention mask 3. Concatenated It he input lib: 4. Churk the concatenated It he input lib: 5. Pack (all the churks 7. Iterate over all the batch is the dataset. Note: the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated It has the dataset. Note: the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated It has the dataset. 5. Pack all the churks 7. Iterate over all the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated Item in the configuration in the new dataset. 5. Some Configuration 6. Split the new dataset into training and test sets with the test_size=0.05 and seed=42. Use the appropriate data collistor function for the MLM objective and set the masking probability to 0.2. Use the data longer than the configuration of the unmasked token. 1. The configuration is the properties of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration	1 point NO
4) Double the context length from \$12 to 1024 (by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (chunk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokerize it to pet input IDs and attention mask 3. Concatenated IDs into a size of \$12 5. Drop the list chunk if its length is less than \$12 6. Pack all the chunks 7. Iterate over all the bipatch size in the dataset 8. There is no concatenated IDs into a size of \$12 6. Pack all the chunks 7. Iterate over all the bipatch gataset in the variable "Sac-Dunked". Enter the total number of samples in the new dataset. Note: the batch are should be kept at 1000 while calling "ds.mag()" for theanswer to match. 4. Observed the sound of the dataset in the dataset into training and test sets with the test. size 0.05 and seed-42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data to PyTorch to load a batch of samples, and enter the token ID corresponding to the unmasked token 100 Yes, the answer is correct.	1 point NO
4) Double the context length for \$12 to 1024 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 0.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38,0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenize it to get input libs and attention mask 3. Concatenated It he input lib: 4. Churk the concatenated It he input lib: 5. Pack (all the churks 7. Iterate over all the batch is the dataset. Note: the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated It has the dataset. Note: the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated It has the dataset. 5. Pack all the churks 7. Iterate over all the batch size should be kept at 1000 while calling 'ds.map()' for theanswer to match. 4. Some Configuration 4. Churk the concatenated Item in the configuration in the new dataset. 5. Some Configuration 6. Split the new dataset into training and test sets with the test_size=0.05 and seed=42. Use the appropriate data collistor function for the MLM objective and set the masking probability to 0.2. Use the data longer than the configuration of the unmasked token. 1. The configuration is the properties of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration is the configuration of the unmasked token. 1. The configuration	1 point NO
4) Double the context length for \$12 to 1024 (c) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenice it to get input IDs and attention mask 3. Ocnocatemest life ineput IDs 4. Churk the concatemated IDs into a size of \$12\$ 5. Pack all the churks 7. Interate over all the batches in the dataset Store the resulting dataset in the dataset Store the resulting dataset in the variable 'dis. churked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds. msp()' for theanswer to match. 246:695 4. Split the new dataset into training and test sets with the test_size=0.05 and seed=42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data to Pyrifoch to load a batch of samples, and enter the token ID corresponding to the sumsafted token Yes, the answer is correct. Score: 1 Accepted Answers:	1 point NO
4) Double the context length for \$12 to 1024 (c) can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 Yes, the answer is correct. Score: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Take a batch of 1000 samples 2. Tokenice it to get input IDs and attention mask 3. Ocnocatemest life ineput IDs 4. Churk the concatemated IDs into a size of \$12\$ 5. Pack all the churks 7. Interate over all the batches in the dataset Store the resulting dataset in the dataset Store the resulting dataset in the variable 'dis. churked'. Enter the total number of samples in the new dataset. Note: the batch size should be kept at 1000 while calling 'ds. msp()' for theanswer to match. 246:695 4. Split the new dataset into training and test sets with the test_size=0.05 and seed=42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data to Pyrifoch to load a batch of samples, and enter the token ID corresponding to the sumsafted token Yes, the answer is correct. Score: 1 Accepted Answers:	JG ader from
4) Double the context length from \$12 to 1934 by can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 2.39 2.40 2.50 parameters in correct. 3.50 pack (churk) the samples such that the length of all the samples in the dataset is \$12 (for efficient training). Define a mapping function that implements the following procedure 1. Takes a batch of 1000 samples 2. Tokenize it to get input (ib and attention mask as one of \$12 to the configuration). 3.50 more that are the insput Disc. 3.50 more based returned if its length to the samples in the dataset is \$12.50 to the resulting dataset in the validaset. 3.50 more the resulting dataset in the dataset. 3.50 more the resulting dataset in the dataset. 4.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer to match. 3.60 pack at the church if its length to 1000 while calling 'ds map()' for the answer to match. 4.6055 4.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer to match. 4.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer to match. 4.6055 4.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer to match. 4.6055 4.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 5.60 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 6.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 5.60 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 6.50 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 5.60 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 5.60 pack at the church if its length to 1000 while calling 'ds map()' for the answer is correct. 6.50 pack at the	JG ader from
4) Double the context length in \$12 to 1024 or can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 Yes, the answer is correct. Scorre: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a black in 16 to 1000 samples 2. Tolerize it to get input ID and attention mask 3. Concatenated in the input IDs 4. Churk the concatenated (ib into a size of 512 5. Drop the last chark if its length is less than 512 6. Pack all the churks 7. Hearts over all the bicarbes in the dataset Store the textual fit all the samples, and enter the token ID corresponding to the unmasked token -100 Ves, the answer is correct. Score: 1 Accepted Answers: (Type: Nameric)-100 7. Create a small ERET model by changing the following lypser-parameters and keeping the other hypser-parameters as is	JG ader from
4) Double the context length for \$12 to 1001 () can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.38 4. Change of the answer is correct. 5. Core: 1 4. Change of 1000 samples 2. Tokerize it to get input 05 and attention mask 3. Conceitenate of the input 10s 4. Chank the conceitenated for inou 5 and attention mask 3. Conceitenate of the input 10s 4. Chank the conceitenated for inou 5 and attention mask 5. Pack of the child in less than 512 7. Inserior over all the batches in the distance. 8. Pack of the child inserts in the setting is less than 512 8. Pack of the child inserts in the distance. 9. Sold the new dataset into training and test sets with the test slave-0.05 and seed-42. Use the appropriate data collator function for the MLM objective and set the masking probability to 0.2. Use the data to PyTorch to load a batch of a samples, and enter the token ID corresponding to the unmasked token 100 Yee, the answer is correct. 5. Core: 1 Accepted Answers: (Type: Nameric) - 100 7) Cheate a small BERT mode by changing the following hyper-parameters and keeping the other hyper-parameters as is	JG ader from
4) Double the context length in \$12 to 1024 or can change it in the configuration). Count the number of parameters and enter the change in the number of parameters (in millions) compared to the default configuration. 3.39 Yes, the answer is correct. Scorre: 1 Accepted Answers: (Type: Range) 0.38.0.41 5) Pack (churk) the samples such that the length of all the samples in the dataset is 512 (for efficient training). Define a mapping function that implements the following procedure 1. Take a black in 16 to 1000 samples 2. Tolerize it to get input ID and attention mask 3. Concatenated in the input IDs 4. Churk the concatenated (ib into a size of 512 5. Drop the last chark if its length is less than 512 6. Pack all the churks 7. Hearts over all the bicarbes in the dataset Store the textual fit all the samples, and enter the token ID corresponding to the unmasked token -100 Ves, the answer is correct. Score: 1 Accepted Answers: (Type: Nameric)-100 7. Create a small ERET model by changing the following lypser-parameters and keeping the other hypser-parameters as is	JG ader from