

2025 USA-NA-AIO Round 2, Problem 3, Part 6

[Topics](#)[My posts](#)[More](#)[CATEGORIES](#)[General](#)[Site Feedback](#)[All categories](#)[Skip to main content](#)**USAAIO** **May 2025**

Part 6 (5 points, coding task)

In this part, we preprocess text data `text_list`.

1. Do tokenization with

```
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
```

2. Call

```
token_id_list = tokenizer(text_list)['input_ids']
```

3. Print `token_id_list`.

4. Print the type of `token_id_list`.

5. Print the length of `token_id_list`.

6. Print `token_id_list[5]`.

7. Print the type of `token_id_list[5]`.

8. Print the type of `token_id_list[5][0]`.

9. For each `idx`, convert `token_id_list[idx]` from the above type to a 1-dim tensor. That is, after this step, `token_id_list` is a list that consists of all 1-dim tensors.

10. Print `token_id_list[5:7]`.

11. Print the data type of `token_id_list[5][0]`.**USAAIO** **May 2025***### WRITE YOUR SOLUTION HERE ###*

```
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')

token_id_list = tokenizer(text_list)[‘input_ids’]

print(token_id_list)
print(type(token_id_list))
print(len(token_id_list))

print(token_id_list[5])
print(type(token_id_list[5]))
print(type(token_id_list[5][0]))

token_id_list = [torch.tensor(token_id_list[idx]) for idx in range(len(token_id_l
print(token_id_list[5:7])
print(token_id_list[5][0].dtype)

""” END OF THIS PART ”””
```

**◆ Related topics**[Skip to main content](#)

Topic

Replies Activity

Topic	Replies	Activity
2025 USA-NA-AIO Round 2, Problem 3, Part 7	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 9	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 4	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 8	1	May 2025
2025 USA-NA-AIO Round 2, Problem 3, Part 10	1	May 2025

 Powered by Discourse