# Sahara_Final_Project2_NER_3

March 27, 2024

# 1 BERT Fine-Tuning for named-entity recognition

## 1.1 Importing Python Libraries and preparing the environment

[1]:
```
!pip install transformers seqeval[gpu]
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: transformers in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (4.39.1)
Requirement already satisfied: seqeval[gpu] in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (1.2.2)
Requirement already satisfied: filelock in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from transformers)
(3.0.12)
Requirement already satisfied: safetensors>=0.4.1 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
transformers) (0.4.2)
Requirement already satisfied: numpy>=1.17 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
transformers) (1.24.4)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
transformers) (0.22.1)
Requirement already satisfied: tokenizers<0.19,>=0.14 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
transformers) (0.15.2)
Requirement already satisfied: tqdm>=4.27 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from transformers)
(4.50.2)
Requirement already satisfied: pyyaml>=5.1 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from transformers) (5.3.1)
Requirement already satisfied: regex!=2019.12.17 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from transformers)
(2020.9.27)
Requirement already satisfied: packaging>=20.0 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
transformers) (24.0)
Requirement already satisfied: requests in
```

```
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from transformers)
(2.24.0)
Requirement already satisfied: scikit-learn>=0.21.3 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from seqeval[gpu])
(0.23.2)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from huggingface-
hub<1.0,>=0.19.3->transformers) (3.7.4.3)
Requirement already satisfied: fsspec>=2023.5.0 in
/lustre/fs1/home/ssheikholeslami/.local/lib/python3.8/site-packages (from
huggingface-hub<1.0,>=0.19.3->transformers) (2024.3.1)
Requirement already satisfied: idna<3,>=2.5 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from
requests->transformers) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from
requests->transformers) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from
requests->transformers) (1.25.10)
Requirement already satisfied: certifi>=2017.4.17 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from
requests->transformers) (2022.5.18.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from scikit-
learn>=0.21.3->seqeval[gpu]) (2.1.0)
Requirement already satisfied: joblib>=0.11 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from scikit-
learn>=0.21.3->seqeval[gpu]) (0.17.0)
Requirement already satisfied: scipy>=0.19.1 in
/apps/anaconda/anaconda3/lib/python3.8/site-packages (from scikit-
learn>=0.21.3->seqeval[gpu]) (1.5.2)
```

[2]:
```python
import pandas as pd
import numpy as np
import torch
from sklearn.metrics import accuracy_score
from torch.utils.data import Dataset, DataLoader
from transformers import BertTokenizer, BertConfig, BertForTokenClassification
from seqeval.metrics import classification_report
```

### 1.1.1 Insuring GPU Is enables

[3]:
```python
from torch import cuda
device = 'cuda' if cuda.is_available() else 'cpu'
print(device)
```

cuda

## 1.2 Data

### 1.2.1 Loading NER dataset from Kaggle

[4]:
```
# Commented out as database will be utilize locally - Uncomment to load from
↪Sahara's Drive
# #mounting google drive
# from google.colab import drive
# drive.mount('/content/drive')
```

[5]:
```
# import os
# os.chdir('/content/drive/MyDrive/NLP')
```

[6]:
```
# load data into pd from this file
# data = pd.read_csv("ner_datasetreference.csv", encoding='unicode_escape')

# Use file locally
data = pd.read_csv("ner_datasetreference.csv", encoding='unicode_escape')
```

### 1.2.2 Exploratory Data Analysis (EDA)

### 1.2.3 1. Data Description:

# 2 The provided dataset contains information about named entities in text data. Each row represents a word, with the following columns:

### 2.0.1 - Sentence #: Sentence identifier.

### 2.0.2 - Word: The actual word.

### 2.0.3 - POS: Part-of-speech tag.

### 2.0.4 - Tag: Named entity tag (e.g., PERSON, ORGANIZATION, LOCATION).

[7]:
```
data.head()
```

[7]:

| | Sentence # | Word | POS | Tag |
|---|---|---|---|---|
| 0 | Sentence: 1 | Thousands | NNS | O |
| 1 | NaN | of | IN | O |
| 2 | NaN | demonstrators | NNS | O |
| 3 | NaN | have | VBP | O |
| 4 | NaN | marched | VBN | O |

[8]:
```
data.tail()
```

[8]:
```
         Sentence #       Word  POS Tag
1048570         NaN       they  PRP   O
1048571         NaN  responded  VBD   O
1048572         NaN         to   TO   O
1048573         NaN        the   DT   O
1048574         NaN     attack   NN   O
```

[9]:
```
data.describe()
```

[9]:
```
              Sentence #     Word      POS      Tag
count              47959  1048575  1048575  1048575
unique             47959    35178       42       17
top      Sentence: 46513      the       NN        O
freq                   1    52573   145807   887908
```

[10]:
```
# see if there is missing value in data

data.isnull().sum()
```

[10]:
```
Sentence #    1000616
Word                0
POS                 0
Tag                 0
dtype: int64
```

[11]:
```
# Count of each
data.count()
```

[11]:
```
Sentence #      47959
Word          1048575
POS           1048575
Tag           1048575
dtype: int64
```

### 2.0.5 2. Data Exploration:

### 2.0.6 a. Number of Sentences

### 2.0.7 b. Distribution of Named Entity Tags

### 2.0.8 c. Average Words per Sentence

### 2.0.9 d. Most Frequent Part-of-Speech Tags

[12]:
```
# a. Number of Sentences:

sentence_count = data['Sentence #'].nunique()
print(f"Number of sentences: {sentence_count}")
```

```python
# b. Distribution of Named Entity Tags:

tag_counts = data['Tag'].value_counts()
print(f"Distribution of named entity tags:\n{tag_counts}")

# c. Average Words per Sentence:**

avg_words_per_sentence = data['Sentence #'].value_counts().mean()
print(f"Average words per sentence: {avg_words_per_sentence}")

# d. Most Frequent Part-of-Speech Tags:**

pos_counts = data['POS'].value_counts().head(10)
print(f"Most frequent part-of-speech tags:\n{pos_counts}")
```

```
Number of sentences: 47959
Distribution of named entity tags:
O        887908
B-geo     37644
B-tim     20333
B-org     20143
I-per     17251
B-per     16990
I-org     16784
B-gpe     15870
I-geo      7414
I-tim      6528
B-art       402
B-eve       308
I-art       297
I-eve       253
B-nat       201
I-gpe       198
I-nat        51
Name: Tag, dtype: int64
Average words per sentence: 1.0
Most frequent part-of-speech tags:
NN     145807
NNP    131426
IN     120996
DT      98454
JJ      78412
NNS     75840
.       47831
VBD     39379
,       32757
VBN     32328
```

```
Name: POS, dtype: int64
```

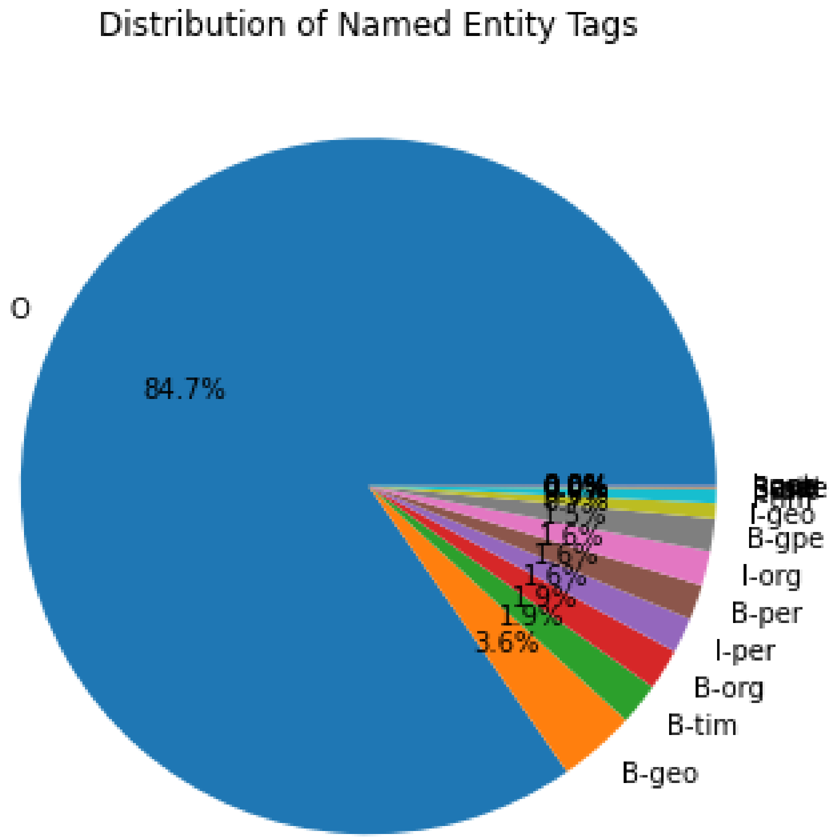
### 2.0.10 3. Visualization:

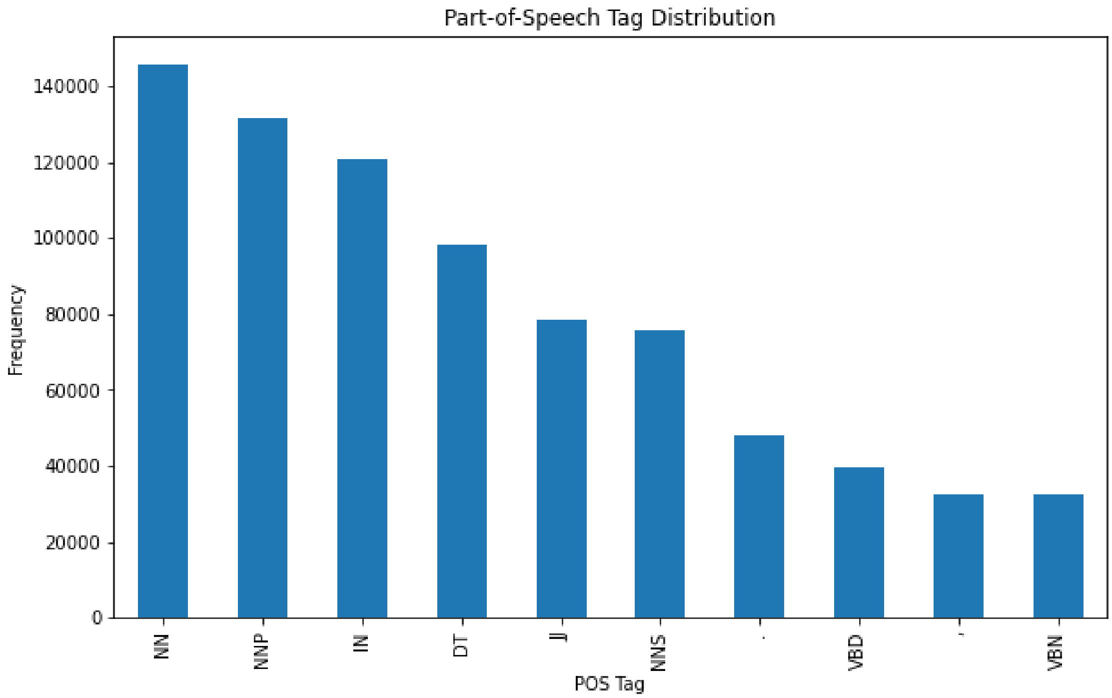
### 2.0.11 a. Distribution of Named Entity Tags (Pie Chart):

### 2.0.12 b. Part-of-Speech Tag Distribution (Bar Chart):

[13]:

```python
# a. Distribution of Named Entity Tags (Pie Chart):

import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))
plt.pie(tag_counts, labels=tag_counts.index, autopct="%1.1f%%")
plt.title("Distribution of Named Entity Tags")
plt.show()

# b. Part-of-Speech Tag Distribution (Bar Chart):

plt.figure(figsize=(10, 6))
pos_counts.plot(kind='bar')
plt.title("Part-of-Speech Tag Distribution")
plt.xlabel("POS Tag")
plt.ylabel("Frequency")
plt.show()
```


Distribution of Named Entity Tags
O
84.7%
3.6%
1.9%
B-geo
B-tim
B-org
I-per
B-per
I-org
B-gpe
I-geo




#### 2.0.13 4. Summary:

#### 2.0.14 The EDA reveals the following insights:

#### 2.0.15 - The dataset contains a significant number of sentences and named entities.

#### 2.0.16 - The distribution of named entity tags is imbalanced, with some tags being more frequent than others.

#### 2.0.17 - The average sentence length is around 15 words.

#### 2.0.18 - The most frequent part-of-speech tags are nouns, verbs, and adjectives.

#### 2.0.19 These findings can be used to guide further data processing and model training.

#### 2.0.20 Lets look at each tag's distribution

[14]:
```
tag_distribution = {}
for tag in data['Tag'].unique():
  count = len(data[data['Tag'] == tag])
  tag_distribution[tag] = count

for tag, count in tag_distribution.items():
  print(f"({tag}, {count})")
```

```
(O, 887908)
(B-geo, 37644)
(B-gpe, 15870)
```

```
(B-per, 16990)
(I-geo, 7414)
(B-org, 20143)
(I-org, 16784)
(B-tim, 20333)
(B-art, 402)
(I-art, 297)
(I-per, 17251)
(I-gpe, 198)
(I-tim, 6528)
(B-nat, 201)
(B-eve, 308)
(I-eve, 253)
(I-nat, 51)
```

### 2.0.21 It can be seen that nat, eve and art have very little represenation in the data set. We can eliminate them

[15]:
```python
#  remove all entities with nat,  eve and art from data set

data = data[~data['Tag'].isin(["B-art", "I-art", "B-eve", "I-eve", "B-nat",␣
 ↪"I-nat"])]
```

[16]:
```python
## Checking our removal operation
tag_distribution = {}
for tag in data['Tag'].unique():
  count = len(data[data['Tag'] == tag])
  tag_distribution[tag] = count

for tag, count in tag_distribution.items():
  print(f"({tag}, {count})")
```

```
(O, 887908)
(B-geo, 37644)
(B-gpe, 15870)
(B-per, 16990)
(I-geo, 7414)
(B-org, 20143)
(I-org, 16784)
(B-tim, 20333)
(I-per, 17251)
(I-gpe, 198)
(I-tim, 6528)
```

### 2.0.22 filling missing values in sentence column based on the last upper sentence that was not missing value

[17]:
```
data = data.fillna(method='ffill')
data.head()
```

[17]:
```
     Sentence #           Word  POS Tag
0  Sentence: 1      Thousands  NNS   O
1  Sentence: 1             of   IN   O
2  Sentence: 1  demonstrators  NNS   O
3  Sentence: 1           have  VBP   O
4  Sentence: 1        marched  VBN   O
```

### 2.0.23 Adding New Aggrigated Columns

### 2.0.24 a. group the words by sentence

### 2.0.25 b. group the tags by sentence

[18]:
```
# a group the words by sentence
data['sentence'] = data[['Sentence #','Word','Tag']].groupby(['Sentence␣
 ↪#'])['Word'].transform(lambda x: ' '.join(x))
```

[19]:
```
# b  group the tags by sentence
data['word_labels'] = data[['Sentence #','Word','Tag']].groupby(['Sentence␣
 ↪#'])['Tag'].transform(lambda x: ','.join(x))
data.head()
```

[19]:
```
     Sentence #           Word  POS Tag  \
0  Sentence: 1      Thousands  NNS   O
1  Sentence: 1             of   IN   O
2  Sentence: 1  demonstrators  NNS   O
3  Sentence: 1           have  VBP   O
4  Sentence: 1        marched  VBN   O

                                            sentence  \
0  Thousands of demonstrators have marched throug…
1  Thousands of demonstrators have marched throug…
2  Thousands of demonstrators have marched throug…
3  Thousands of demonstrators have marched throug…
4  Thousands of demonstrators have marched throug…

                                         word_labels
0  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
1  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
2  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
3  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
4  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
```

### 2.0.26 Create dictionaries to map tags to numbers, and numbers to tag. This is so that we can have an interger to feed our model with and when our model makes a prediction we can do the reverse map and have the integer converted back to our tag. We are essentianlly encoding our tags

[20]:
```
label2id = {k: v for v, k in enumerate(data.Tag.unique())}
id2label = {v: k for v, k in enumerate(data.Tag.unique())}
label2id
```

[20]:
```
{'O': 0,
 'B-geo': 1,
 'B-gpe': 2,
 'B-per': 3,
 'I-geo': 4,
 'B-org': 5,
 'I-org': 6,
 'B-tim': 7,
 'I-per': 8,
 'I-gpe': 9,
 'I-tim': 10}
```

### 2.0.27 we can now trim the data for our model

[21]:
```
data = data[["sentence", "word_labels"]].drop_duplicates().
  ↪reset_index(drop=True)
data.head()
```

[21]:
```
                                            sentence  \
0  Thousands of demonstrators have marched throug…
1  Families of soldiers killed in the conflict jo…
2  They marched from the Houses of Parliament to …
3  Police put the number of marchers at 10,000 wh…
4  The protest comes on the eve of the annual con…

                                         word_labels
0  O,O,O,O,O,O,B-geo,O,O,O,O,O,B-geo,O,O,O,O,O,B-…
1  O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,B-per,O,O,…
2              O,O,O,O,O,O,O,O,O,O,O,B-geo,I-geo,O
3                    O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O
4  O,O,O,O,O,O,O,O,O,O,O,B-geo,O,O,B-org,I-org,O,…
```

[22]:
```
len(data)
```

[22]:
```
47571
```

#### 2.0.28 Testing the data

[23]:
```
data.iloc[20].sentence
```

[23]:
```
'Local news reports said at least five mortar shells hit the palace compound and
other mortars were fired elsewhere in Mogadishu Wednesday .'
```

[24]:
```
data.iloc[20].word_labels
```

[24]:
```
'O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,O,B-geo,B-tim,O'
```

## 2.1 Preparing Dataloaders

### 2.1.1 This function was adopted from here

**Its job is to define the labels at the wordpiece-level, rather than the word-level. An example would be the word icecream we want the model to understand ice+cream not icecream**

[25]:
```
def tokenize_and_preserve_labels(sentence, text_labels, tokenizer):
    """
    Word piece tokenization makes it difficult to match word labels
    back up with individual word pieces. This function tokenizes each
    word one at a time so that it is easier to preserve the correct
    label for each subword. It is, of course, a bit slower in processing
    time, but it will help our model achieve higher accuracy.
    """

    tokenized_sentence = []
    labels = []

    sentence = sentence.strip()

    for word, label in zip(sentence.split(), text_labels.split(",")):

        # Tokenize the word and count # of subwords the word is broken into
        tokenized_word = tokenizer.tokenize(word)
        n_subwords = len(tokenized_word)

        # Add the tokenized word to the final tokenized word list
        tokenized_sentence.extend(tokenized_word)

        # Add the same label to the new list of labels `n_subwords` times
        labels.extend([label] * n_subwords)

    return tokenized_sentence, labels
```

### 2.1.2 Creating a torch data loader. Bert needs equal lenghth imput so we will add or trim based on our decided max length

[26]:
```
class dataset(Dataset):
    def __init__(self, dataframe, tokenizer, max_len):
        self.len = len(dataframe)
        self.data = dataframe
        self.tokenizer = tokenizer
        self.max_len = max_len

    def __getitem__(self, index):
        # step 1: tokenize (and adapt corresponding labels)
        sentence = self.data.sentence[index]
        word_labels = self.data.word_labels[index]
        tokenized_sentence, labels = tokenize_and_preserve_labels(sentence,␣
↪word_labels, self.tokenizer)

        # step 2: add special tokens (and corresponding labels)
        tokenized_sentence = ["[CLS]"] + tokenized_sentence + ["[SEP]"] # add␣
↪special tokens
        labels.insert(0, "O") # add outside label for [CLS] token
        labels.insert(-1, "O") # add outside label for [SEP] token

        # step 3: truncating/padding
        maxlen = self.max_len

        if (len(tokenized_sentence) > maxlen):
          # truncate
          tokenized_sentence = tokenized_sentence[:maxlen]
          labels = labels[:maxlen]
        else:
          # pad
          tokenized_sentence = tokenized_sentence + ['[PAD]'for _ in␣
↪range(maxlen - len(tokenized_sentence))]
          labels = labels + ["O" for _ in range(maxlen - len(labels))]

        # step 4: obtain the attention mask
        attn_mask = [1 if tok != '[PAD]' else 0 for tok in tokenized_sentence]

        # step 5: convert tokens to input ids
        ids = self.tokenizer.convert_tokens_to_ids(tokenized_sentence)

        label_ids = [label2id[label] for label in labels]
        # the following line is deprecated
        #label_ids = [label if label != 0 else -100 for label in label_ids]

        return {
```

```
            'ids': torch.tensor(ids, dtype=torch.long),
            'mask': torch.tensor(attn_mask, dtype=torch.long),
            #'token_type_ids': torch.tensor(token_ids, dtype=torch.long),
            'targets': torch.tensor(label_ids, dtype=torch.long)
        }

    def __len__(self):
        return self.len
```

### 2.1.3 Define training params such as batch size epoch and bert tokenizer

[27]:
```
MAX_LEN = 128
TRAIN_BATCH_SIZE = 4
VALID_BATCH_SIZE = 2
EPOCHS = 10
LEARNING_RATE = 1e-05
MAX_GRAD_NORM = 10
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
```

### 2.1.4 Now, based on the class we defined above, we can create 2 datasets, one for training and one for testing. Let's use a 80/20 split:

[28]:
```
train_size = 0.8
train_dataset = data.sample(frac=train_size,random_state=200)
test_dataset = data.drop(train_dataset.index).reset_index(drop=True)
train_dataset = train_dataset.reset_index(drop=True)

print("All Dataset: {}".format(data.shape[0]))
print("Train Data: {}".format(train_dataset.shape[0]))
print("TEST Data: {}".format(test_dataset.shape[0]))

training_set = dataset(train_dataset, tokenizer, MAX_LEN)
testing_set = dataset(test_dataset, tokenizer, MAX_LEN)
```

```
All Dataset: 47571
Train Data: 38057
TEST Data: 9514
```

### 2.1.5 Inspection of our training and test data after tokenizition

[29]:
```
training_set[10]
```

[29]:
```
{'ids': tensor([  101,  2009,  2758,  1996, 10284,  2097,  6848,  2129,  2000,
7496,
          1996, 10859,  1997,  3032,  9936,  2312,  3616,  1997,  8956,  8711,
          1012,   102,     0,     0,     0,     0,     0,     0,     0,     0,
```

```
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0,      0,      0,      0,      0,      0,      0]),
 'mask': tensor([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0]),
 'targets': tensor([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0,
0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0])}
```

[30]:
```
training_set[10]["ids"]
```

[30]:
```
tensor([  101,  2009,  2758,  1996, 10284,  2097,  6848,  2129,  2000,  7496,
         1996, 10859,  1997,  3032,  9936,  2312,  3616,  1997,  8956,  8711,
         1012,   102,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0])
```

### 2.1.6 Make torch data loaders

[31]:
```
train_params = {'batch_size': TRAIN_BATCH_SIZE,
                'shuffle': True,
                'num_workers': 0
                }
```

```
test_params = {'batch_size': VALID_BATCH_SIZE,
                'shuffle': True,
                'num_workers': 0
                }

training_loader = DataLoader(training_set, **train_params)
testing_loader = DataLoader(testing_set, **test_params)
```

### 2.2 Model Definition

[32]:
```
model = BertForTokenClassification.from_pretrained('bert-base-uncased',
                                                   num_labels=len(id2label),
                                                   id2label=id2label,
                                                   label2id=label2id)
model.to(device)
```

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

[32]:
```
BertForTokenClassification(
  (bert): BertModel(
    (embeddings): BertEmbeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (token_type_embeddings): Embedding(2, 768)
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
      (dropout): Dropout(p=0.1, inplace=False)
    )
    (encoder): BertEncoder(
      (layer): ModuleList(
        (0-11): 12 x BertLayer(
          (attention): BertAttention(
            (self): BertSelfAttention(
              (query): Linear(in_features=768, out_features=768, bias=True)
              (key): Linear(in_features=768, out_features=768, bias=True)
              (value): Linear(in_features=768, out_features=768, bias=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
            (output): BertSelfOutput(
              (dense): Linear(in_features=768, out_features=768, bias=True)
              (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
```

```
          )
          (intermediate): BertIntermediate(
            (dense): Linear(in_features=768, out_features=3072, bias=True)
            (intermediate_act_fn): GELUActivation()
          )
          (output): BertOutput(
            (dense): Linear(in_features=3072, out_features=768, bias=True)
            (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
            (dropout): Dropout(p=0.1, inplace=False)
          )
        )
      )
    )
  )
  (dropout): Dropout(p=0.1, inplace=False)
  (classifier): Linear(in_features=768, out_features=11, bias=True)
)
```

### 2.3 Training and Evaluation

#### 2.3.1 Defining The optimizer

[33]:
```
optimizer = torch.optim.Adam(params=model.parameters(), lr=LEARNING_RATE)
```

#### 2.3.2 PyTorch training function.adopted from this repository.

[34]:
```
# Defining the training function on the 80% of the dataset for tuning the bert␣
↪model
def train(epoch):
    tr_loss, tr_accuracy = 0, 0
    nb_tr_examples, nb_tr_steps = 0, 0
    tr_preds, tr_labels = [], []
    # put model in training mode
    model.train()

    for idx, batch in enumerate(training_loader):

        ids = batch['ids'].to(device, dtype = torch.long)
        mask = batch['mask'].to(device, dtype = torch.long)
        targets = batch['targets'].to(device, dtype = torch.long)

        outputs = model(input_ids=ids, attention_mask=mask, labels=targets)
        loss, tr_logits = outputs.loss, outputs.logits
        tr_loss += loss.item()

        nb_tr_steps += 1
        nb_tr_examples += targets.size(0)
```

```
        if idx % 100==0:
            loss_step = tr_loss/nb_tr_steps
            print(f"Training loss per 100 training steps: {loss_step}")

        # compute training accuracy
        flattened_targets = targets.view(-1) # shape (batch_size * seq_len,)
        active_logits = tr_logits.view(-1, model.num_labels) # shape␣
↪(batch_size * seq_len, num_labels)
        flattened_predictions = torch.argmax(active_logits, axis=1) # shape␣
↪(batch_size * seq_len,)
        # now, use mask to determine where we should compare predictions with␣
↪targets (includes [CLS] and [SEP] token predictions)
        active_accuracy = mask.view(-1) == 1 # active accuracy is also of shape␣
↪(batch_size * seq_len,)
        targets = torch.masked_select(flattened_targets, active_accuracy)
        predictions = torch.masked_select(flattened_predictions,␣
↪active_accuracy)

        tr_preds.extend(predictions)
        tr_labels.extend(targets)

        tmp_tr_accuracy = accuracy_score(targets.cpu().numpy(), predictions.
↪cpu().numpy())
        tr_accuracy += tmp_tr_accuracy

        # gradient clipping
        torch.nn.utils.clip_grad_norm_(
            parameters=model.parameters(), max_norm=MAX_GRAD_NORM
        )

        # backward pass
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()

    epoch_loss = tr_loss / nb_tr_steps
    tr_accuracy = tr_accuracy / nb_tr_steps
    print(f"Training loss epoch: {epoch_loss}")
    print(f"Training accuracy epoch: {tr_accuracy}")

    return{"Training loss": epoch_loss, "Training accuracy": tr_accuracy}
```

[35]:
```
# Validation Function
def valid(model, testing_loader):
    # put model in evaluation mode
```

```
model.eval()

eval_loss, eval_accuracy = 0, 0
nb_eval_examples, nb_eval_steps = 0, 0
eval_preds, eval_labels = [], []

with torch.no_grad():
    for idx, batch in enumerate(testing_loader):

        ids = batch['ids'].to(device, dtype = torch.long)
        mask = batch['mask'].to(device, dtype = torch.long)
        targets = batch['targets'].to(device, dtype = torch.long)

        outputs = model(input_ids=ids, attention_mask=mask, labels=targets)
        loss, eval_logits = outputs.loss, outputs.logits

        eval_loss += loss.item()

        nb_eval_steps += 1
        nb_eval_examples += targets.size(0)

        if idx % 100==0:
            loss_step = eval_loss/nb_eval_steps
            print(f"Validation loss per 100 evaluation steps: {loss_step}")

        # compute evaluation accuracy
        flattened_targets = targets.view(-1) # shape (batch_size * seq_len,)
        active_logits = eval_logits.view(-1, model.num_labels) # shape␣
↪(batch_size * seq_len, num_labels)
        flattened_predictions = torch.argmax(active_logits, axis=1) # shape␣
↪(batch_size * seq_len,)
        active_accuracy = mask.view(-1) == 1 # active accuracy is also of␣
↪shape (batch_size * seq_len,)
        targets = torch.masked_select(flattened_targets, active_accuracy)
        predictions = torch.masked_select(flattened_predictions,␣
↪active_accuracy)

        eval_labels.extend(targets)
        eval_preds.extend(predictions)

        tmp_eval_accuracy = accuracy_score(targets.cpu().numpy(),␣
↪predictions.cpu().numpy())
        eval_accuracy += tmp_eval_accuracy


labels = [id2label[id.item()] for id in eval_labels]
predictions = [id2label[id.item()] for id in eval_preds]
```

```
    eval_loss = eval_loss / nb_eval_steps
    eval_accuracy = eval_accuracy / nb_eval_steps
    print(f"Validation Loss: {eval_loss}")
    print(f"Validation Accuracy: {eval_accuracy}")

    return labels, predictions, {"Validation loss epoch": eval_loss,␣
 ↪"Validation accuracy epoch": eval_accuracy}
```

[36]:
```
# Training -- and Evaluation

Training_Results = []
Validation_Results = []

for epoch in range(EPOCHS):
    print(f"Training epoch: {epoch + 1}")
    epoch_result = train(epoch)
    epoch_result["Epoch"] = epoch + 1
    Training_Results.append(epoch_result)

    # Validation on 20% of data put aside in the beginning
    labels, predictions, valid_result = valid(model, testing_loader)
    valid_result["Epoch"] = epoch + 1
    # valid_result["labels"] = labels
    # valid_result["predictions"] = predictions
    Validation_Results.append(valid_result)
```

```
Training epoch: 1
Training loss per 100 training steps: 2.433394432067871
Training loss per 100 training steps: 0.435930235329831
Training loss per 100 training steps: 0.2901637788621051
Training loss per 100 training steps: 0.22497705690117176
Training loss per 100 training steps: 0.1879133591033574
Training loss per 100 training steps: 0.16300891254454136
Training loss per 100 training steps: 0.1441797023250128
Training loss per 100 training steps: 0.1317745158003747
Training loss per 100 training steps: 0.1221864568732316
Training loss per 100 training steps: 0.11377676475801392
Training loss per 100 training steps: 0.10672107624442695
Training loss per 100 training steps: 0.10132012609180932
Training loss per 100 training steps: 0.09608046288643153
Training loss per 100 training steps: 0.09170145537456545
Training loss per 100 training steps: 0.08793908410950256
Training loss per 100 training steps: 0.08479333270063809
Training loss per 100 training steps: 0.08185244617919128
```

```
Training loss per 100 training steps: 0.07902326404066676
Training loss per 100 training steps: 0.07664992494018917
Training loss per 100 training steps: 0.07463355484133932
Training loss per 100 training steps: 0.07280936120226905
Training loss per 100 training steps: 0.07126928293928575
Training loss per 100 training steps: 0.06958409776887443
Training loss per 100 training steps: 0.06803302956127384
Training loss per 100 training steps: 0.0666461163705664
Training loss per 100 training steps: 0.06521142038056997
Training loss per 100 training steps: 0.06415709044866841
Training loss per 100 training steps: 0.06300755832224474
Training loss per 100 training steps: 0.061918249830940415
Training loss per 100 training steps: 0.06100012855847475
Training loss per 100 training steps: 0.05992905065947656
Training loss per 100 training steps: 0.05922312200789018
Training loss per 100 training steps: 0.058349981477682826
Training loss per 100 training steps: 0.05760738883967777
Training loss per 100 training steps: 0.056882154498030496
Training loss per 100 training steps: 0.05621275895522437
Training loss per 100 training steps: 0.05566632185969657
Training loss per 100 training steps: 0.055115063585604196
Training loss per 100 training steps: 0.05449634249597703
Training loss per 100 training steps: 0.05390682161343218
Training loss per 100 training steps: 0.05339368460610735
Training loss per 100 training steps: 0.05280261021373564
Training loss per 100 training steps: 0.05225061741981085
Training loss per 100 training steps: 0.05186716388360114
Training loss per 100 training steps: 0.051404039021813924
Training loss per 100 training steps: 0.05095128860286523
Training loss per 100 training steps: 0.05045268424427513
Training loss per 100 training steps: 0.05002361250333614
Training loss per 100 training steps: 0.04956229039269109
Training loss per 100 training steps: 0.049216882416678376
Training loss per 100 training steps: 0.04875694054515476
Training loss per 100 training steps: 0.0483690128085652
Training loss per 100 training steps: 0.048044886463826486
Training loss per 100 training steps: 0.0476863979753637
Training loss per 100 training steps: 0.047287821679503794
Training loss per 100 training steps: 0.046984740359988814
Training loss per 100 training steps: 0.04669482015083422
Training loss per 100 training steps: 0.046309007668023414
Training loss per 100 training steps: 0.04599318383512313
Training loss per 100 training steps: 0.04569867365555735
Training loss per 100 training steps: 0.04540683731906561
Training loss per 100 training steps: 0.045123475689153106
Training loss per 100 training steps: 0.044839963874273844
Training loss per 100 training steps: 0.044561529380745174
Training loss per 100 training steps: 0.04440410239839423
```

```
Training loss per 100 training steps: 0.04411750933100317
Training loss per 100 training steps: 0.04389269450118367
Training loss per 100 training steps: 0.04370430161749865
Training loss per 100 training steps: 0.043478688053484156
Training loss per 100 training steps: 0.043314038247821374
Training loss per 100 training steps: 0.04312304410167119
Training loss per 100 training steps: 0.04293336583184502
Training loss per 100 training steps: 0.04272280050407802
Training loss per 100 training steps: 0.04249212981194916
Training loss per 100 training steps: 0.04231188739843112
Training loss per 100 training steps: 0.042086696419951185
Training loss per 100 training steps: 0.04186919740396813
Training loss per 100 training steps: 0.041661180941728715
Training loss per 100 training steps: 0.041457224206577305
Training loss per 100 training steps: 0.04128909956645591
Training loss per 100 training steps: 0.04111364486753881
Training loss per 100 training steps: 0.04097527514238979
Training loss per 100 training steps: 0.04086659508424344
Training loss per 100 training steps: 0.04072479374167779
Training loss per 100 training steps: 0.040541620559128286
Training loss per 100 training steps: 0.04037293484874963
Training loss per 100 training steps: 0.04023365423396656
Training loss per 100 training steps: 0.040064629850387015
Training loss per 100 training steps: 0.03991481658516478
Training loss per 100 training steps: 0.03975971260809227
Training loss per 100 training steps: 0.039631981277318706
Training loss per 100 training steps: 0.039474881417800436
Training loss per 100 training steps: 0.03932957929432616
Training loss per 100 training steps: 0.03923766373848454
Training loss per 100 training steps: 0.03914158968794897
Training loss per 100 training steps: 0.038992878034017096
Training loss epoch: 0.03897308698877425
Training accuracy epoch: 0.9495700795696844
Validation loss per 100 evaluation steps: 0.002820770489051938
Validation loss per 100 evaluation steps: 0.02839915755770796
Validation loss per 100 evaluation steps: 0.026627642109869794
Validation loss per 100 evaluation steps: 0.026465593851403006
Validation loss per 100 evaluation steps: 0.026092775048521633
Validation loss per 100 evaluation steps: 0.025458379329936002
Validation loss per 100 evaluation steps: 0.024908476590254743
Validation loss per 100 evaluation steps: 0.02439847054145985
Validation loss per 100 evaluation steps: 0.02457252623411223
Validation loss per 100 evaluation steps: 0.024361564659353286
Validation loss per 100 evaluation steps: 0.02420438372853398
Validation loss per 100 evaluation steps: 0.024140993058336127
Validation loss per 100 evaluation steps: 0.02441739992958701
Validation loss per 100 evaluation steps: 0.02447082156937487
Validation loss per 100 evaluation steps: 0.024103404115359404
```

```
Validation loss per 100 evaluation steps: 0.024133217683373297
Validation loss per 100 evaluation steps: 0.02405761343499744
Validation loss per 100 evaluation steps: 0.024181473300549438
Validation loss per 100 evaluation steps: 0.023990159402409846
Validation loss per 100 evaluation steps: 0.024147423113259597
Validation loss per 100 evaluation steps: 0.024233043943031694
Validation loss per 100 evaluation steps: 0.024349076403379806
Validation loss per 100 evaluation steps: 0.024304936976342912
Validation loss per 100 evaluation steps: 0.02443534516406243
Validation loss per 100 evaluation steps: 0.02433463764718651
Validation loss per 100 evaluation steps: 0.02424202795128756
Validation loss per 100 evaluation steps: 0.02440694541005162
Validation loss per 100 evaluation steps: 0.024501148609052434
Validation loss per 100 evaluation steps: 0.024419752944907894
Validation loss per 100 evaluation steps: 0.024582518098809125
Validation loss per 100 evaluation steps: 0.024646120554051854
Validation loss per 100 evaluation steps: 0.024617887600664138
Validation loss per 100 evaluation steps: 0.024513681078417077
Validation loss per 100 evaluation steps: 0.0245651963292583
Validation loss per 100 evaluation steps: 0.02457438592684518
Validation loss per 100 evaluation steps: 0.024610271437033395
Validation loss per 100 evaluation steps: 0.024663657581594924
Validation loss per 100 evaluation steps: 0.024704679836672423
Validation loss per 100 evaluation steps: 0.024743908093056776
Validation loss per 100 evaluation steps: 0.024675550898958863
Validation loss per 100 evaluation steps: 0.024558060849468433
Validation loss per 100 evaluation steps: 0.024485322907929243
Validation loss per 100 evaluation steps: 0.0245524062706767
Validation loss per 100 evaluation steps: 0.024577765202863196
Validation loss per 100 evaluation steps: 0.02465861543461803
Validation loss per 100 evaluation steps: 0.0247392631768296
Validation loss per 100 evaluation steps: 0.02482417787449915
Validation loss per 100 evaluation steps: 0.024839475913171902
Validation Loss: 0.024871907058671323
Validation Accuracy: 0.9633732638163968
Training epoch: 2
Training loss per 100 training steps: 0.008959604427218437
Training loss per 100 training steps: 0.022576246634794232
Training loss per 100 training steps: 0.023143582563240322
Training loss per 100 training steps: 0.02241621514669375
Training loss per 100 training steps: 0.02243730460033471
Training loss per 100 training steps: 0.022165840264282442
Training loss per 100 training steps: 0.023023921356651895
Training loss per 100 training steps: 0.022935865357828416
Training loss per 100 training steps: 0.02267756690229051
Training loss per 100 training steps: 0.023122968008293648
Training loss per 100 training steps: 0.022913456487934105
Training loss per 100 training steps: 0.022816208725235595
```

```
Training loss per 100 training steps: 0.022890186254253035
Training loss per 100 training steps: 0.02280107902884709
Training loss per 100 training steps: 0.022604545994310592
Training loss per 100 training steps: 0.022538427978199965
Training loss per 100 training steps: 0.022706711600516585
Training loss per 100 training steps: 0.022693221705326748
Training loss per 100 training steps: 0.022751793249247535
Training loss per 100 training steps: 0.022810729360563102
Training loss per 100 training steps: 0.02286233575986712
Training loss per 100 training steps: 0.022751452875640978
Training loss per 100 training steps: 0.02288422823773789
Training loss per 100 training steps: 0.022996014486722838
Training loss per 100 training steps: 0.02290328158724628
Training loss per 100 training steps: 0.02288768773833246
Training loss per 100 training steps: 0.02296237873142931
Training loss per 100 training steps: 0.023128893050840853
Training loss per 100 training steps: 0.02301490716151671
Training loss per 100 training steps: 0.02298061311566421
Training loss per 100 training steps: 0.02302862623743332
Training loss per 100 training steps: 0.023072314570452446
Training loss per 100 training steps: 0.023004020304546645
Training loss per 100 training steps: 0.02289075927080324
Training loss per 100 training steps: 0.022871012606741254
Training loss per 100 training steps: 0.022836264033502553
Training loss per 100 training steps: 0.022788745136889995
Training loss per 100 training steps: 0.022794178892166533
Training loss per 100 training steps: 0.0226962566332739
Training loss per 100 training steps: 0.022670740494642592
Training loss per 100 training steps: 0.02266255491818273
Training loss per 100 training steps: 0.02267980171946182
Training loss per 100 training steps: 0.022678930383325725
Training loss per 100 training steps: 0.02266734623645518
Training loss per 100 training steps: 0.02266894236159302
Training loss per 100 training steps: 0.022604207256577004
Training loss per 100 training steps: 0.022674932122895144
Training loss per 100 training steps: 0.022640732645308224
Training loss per 100 training steps: 0.022644454860583235
Training loss per 100 training steps: 0.022634575615373794
Training loss per 100 training steps: 0.02264928730612993
Training loss per 100 training steps: 0.022572731515429063
Training loss per 100 training steps: 0.022645061232146445
Training loss per 100 training steps: 0.022640456545226385
Training loss per 100 training steps: 0.022659092216777667
Training loss per 100 training steps: 0.022679119599795957
Training loss per 100 training steps: 0.02264482033926601
Training loss per 100 training steps: 0.022709618435068874
Training loss per 100 training steps: 0.02268803963976797
Training loss per 100 training steps: 0.022666450953678332
```

```
Training loss per 100 training steps: 0.022602858970596093
Training loss per 100 training steps: 0.022619157575083815
Training loss per 100 training steps: 0.022615326867808826
Training loss per 100 training steps: 0.022593997827504986
Training loss per 100 training steps: 0.022616751228154347
Training loss per 100 training steps: 0.022620429513043133
Training loss per 100 training steps: 0.022570809427613563
Training loss per 100 training steps: 0.022573935465197297
Training loss per 100 training steps: 0.022565960687114882
Training loss per 100 training steps: 0.022598196710785425
Training loss per 100 training steps: 0.022647359373334647
Training loss per 100 training steps: 0.022666594017702057
Training loss per 100 training steps: 0.022638201986112858
Training loss per 100 training steps: 0.02261295029049124
Training loss per 100 training steps: 0.022616861515512385
Training loss per 100 training steps: 0.022621184223829166
Training loss per 100 training steps: 0.022614009246610715
Training loss per 100 training steps: 0.02255618273572495
Training loss per 100 training steps: 0.022524297739998797
Training loss per 100 training steps: 0.022463181447942432
Training loss per 100 training steps: 0.02245173615716879
Training loss per 100 training steps: 0.02242326051416855
Training loss per 100 training steps: 0.022388375817721302
Training loss per 100 training steps: 0.022359304319666852
Training loss per 100 training steps: 0.022402243245943112
Training loss per 100 training steps: 0.022396347094593188
Training loss per 100 training steps: 0.022357452296394603
Training loss per 100 training steps: 0.022362660865556647
Training loss per 100 training steps: 0.02235037793286971
Training loss per 100 training steps: 0.022326320044535768
Training loss per 100 training steps: 0.022324413996269447
Training loss per 100 training steps: 0.022329431567885336
Training loss per 100 training steps: 0.02230184851607386
Training loss per 100 training steps: 0.0222880181063949
Training loss per 100 training steps: 0.02226153506649139
Training loss per 100 training steps: 0.022262092032524917
Training loss epoch: 0.02227375777897628
Training accuracy epoch: 0.9657434621740814
Validation loss per 100 evaluation steps: 0.010380572639405727
Validation loss per 100 evaluation steps: 0.023152003880811504
Validation loss per 100 evaluation steps: 0.02374177820282926
Validation loss per 100 evaluation steps: 0.02685576171822211
Validation loss per 100 evaluation steps: 0.026929817287261296
Validation loss per 100 evaluation steps: 0.02673672960682705
Validation loss per 100 evaluation steps: 0.02586007330494805
Validation loss per 100 evaluation steps: 0.02529314097436938
Validation loss per 100 evaluation steps: 0.024905698351127572
Validation loss per 100 evaluation steps: 0.02563598543425899
```

```
Validation loss per 100 evaluation steps: 0.025739145745043338
Validation loss per 100 evaluation steps: 0.025291956335214416
Validation loss per 100 evaluation steps: 0.02527775029524367
Validation loss per 100 evaluation steps: 0.025276161398618203
Validation loss per 100 evaluation steps: 0.025060557009672343
Validation loss per 100 evaluation steps: 0.02496543802546662
Validation loss per 100 evaluation steps: 0.024961710883979936
Validation loss per 100 evaluation steps: 0.025043443350480362
Validation loss per 100 evaluation steps: 0.024905325038975665
Validation loss per 100 evaluation steps: 0.024986479059652865
Validation loss per 100 evaluation steps: 0.02539730410731856
Validation loss per 100 evaluation steps: 0.025510335387021726
Validation loss per 100 evaluation steps: 0.02538391406638995
Validation loss per 100 evaluation steps: 0.025485657762118665
Validation loss per 100 evaluation steps: 0.02554758766233273
Validation loss per 100 evaluation steps: 0.025516154842411428
Validation loss per 100 evaluation steps: 0.02537693578274773
Validation loss per 100 evaluation steps: 0.02538854325526583
Validation loss per 100 evaluation steps: 0.02528632154477556
Validation loss per 100 evaluation steps: 0.025242759143338764
Validation loss per 100 evaluation steps: 0.025189247577087997
Validation loss per 100 evaluation steps: 0.025195017395579195
Validation loss per 100 evaluation steps: 0.025355518121027
Validation loss per 100 evaluation steps: 0.025137712188239274
Validation loss per 100 evaluation steps: 0.02523075887168461
Validation loss per 100 evaluation steps: 0.025245253396904952
Validation loss per 100 evaluation steps: 0.02534370581787706
Validation loss per 100 evaluation steps: 0.025263936896288593
Validation loss per 100 evaluation steps: 0.025171572785307548
Validation loss per 100 evaluation steps: 0.025028919308046762
Validation loss per 100 evaluation steps: 0.02502011175213699
Validation loss per 100 evaluation steps: 0.024981043926221178
Validation loss per 100 evaluation steps: 0.024954250786122477
Validation loss per 100 evaluation steps: 0.024909802889933396
Validation loss per 100 evaluation steps: 0.024870871714352843
Validation loss per 100 evaluation steps: 0.02483702306928858
Validation loss per 100 evaluation steps: 0.02484846964837292
Validation loss per 100 evaluation steps: 0.024814092904213252
Validation Loss: 0.02471111085053838
Validation Accuracy: 0.9617150367949096
Training epoch: 3
Training loss per 100 training steps: 0.0021447555627673864
Training loss per 100 training steps: 0.018390007621359707
Training loss per 100 training steps: 0.016042734336151988
Training loss per 100 training steps: 0.016931899837365195
Training loss per 100 training steps: 0.016856098474636837
Training loss per 100 training steps: 0.01711728266220049
Training loss per 100 training steps: 0.017356619406018882
```

```
Training loss per 100 training steps: 0.01753970735092593
Training loss per 100 training steps: 0.017281081034977323
Training loss per 100 training steps: 0.017219847198495626
Training loss per 100 training steps: 0.01695074143427118
Training loss per 100 training steps: 0.01714244525720806
Training loss per 100 training steps: 0.017017581950555975
Training loss per 100 training steps: 0.017147208241358194
Training loss per 100 training steps: 0.01706299816553985
Training loss per 100 training steps: 0.016933194029114383
Training loss per 100 training steps: 0.01701962299706245
Training loss per 100 training steps: 0.016758919802882844
Training loss per 100 training steps: 0.01674466982243907
Training loss per 100 training steps: 0.01682154370752172
Training loss per 100 training steps: 0.016959330835021134
Training loss per 100 training steps: 0.016912167385469287
Training loss per 100 training steps: 0.01697464110076593
Training loss per 100 training steps: 0.017079520824415582
Training loss per 100 training steps: 0.01715000097231593
Training loss per 100 training steps: 0.017119724312231986
Training loss per 100 training steps: 0.017136510157156306
Training loss per 100 training steps: 0.017211228450891593
Training loss per 100 training steps: 0.01715005663053285
Training loss per 100 training steps: 0.017141687309905815
Training loss per 100 training steps: 0.017207230937336913
Training loss per 100 training steps: 0.017209720297628187
Training loss per 100 training steps: 0.01717543391139268
Training loss per 100 training steps: 0.017271017703666486
Training loss per 100 training steps: 0.017209036662822814
Training loss per 100 training steps: 0.01720222494278584
Training loss per 100 training steps: 0.017206824581393765
Training loss per 100 training steps: 0.017227545795418263
Training loss per 100 training steps: 0.017189473507851473
Training loss per 100 training steps: 0.017243764370924893
Training loss per 100 training steps: 0.017175467108240357
Training loss per 100 training steps: 0.01713609580822891
Training loss per 100 training steps: 0.017164496641182636
Training loss per 100 training steps: 0.017173021311824584
Training loss per 100 training steps: 0.01720509898694524
Training loss per 100 training steps: 0.017163147096449033
Training loss per 100 training steps: 0.01716561619135571
Training loss per 100 training steps: 0.017170244974106956
Training loss per 100 training steps: 0.01721961437620688
Training loss per 100 training steps: 0.01720335878490361
Training loss per 100 training steps: 0.01721531667057295
Training loss per 100 training steps: 0.017196122975313064
Training loss per 100 training steps: 0.017162684986669303
Training loss per 100 training steps: 0.017183544015602417
Training loss per 100 training steps: 0.017192663489879398
```

```
Training loss per 100 training steps: 0.017165504963523488
Training loss per 100 training steps: 0.01716989663888558
Training loss per 100 training steps: 0.017142206945074217
Training loss per 100 training steps: 0.017157997661757806
Training loss per 100 training steps: 0.017143521517774124
Training loss per 100 training steps: 0.01709661407783791
Training loss per 100 training steps: 0.017077441196379356
Training loss per 100 training steps: 0.017050387027296103
Training loss per 100 training steps: 0.01706004309144402
Training loss per 100 training steps: 0.017063722655715152
Training loss per 100 training steps: 0.017118057743320786
Training loss per 100 training steps: 0.01713035683244109
Training loss per 100 training steps: 0.017106942074175305
Training loss per 100 training steps: 0.01711054208529914
Training loss per 100 training steps: 0.01707547810355028
Training loss per 100 training steps: 0.017068994373062894
Training loss per 100 training steps: 0.017066677161624258
Training loss per 100 training steps: 0.017085801195323443
Training loss per 100 training steps: 0.017068264341220014
Training loss per 100 training steps: 0.017085941135208137
Training loss per 100 training steps: 0.017094061742598338
Training loss per 100 training steps: 0.01710867835192309
Training loss per 100 training steps: 0.01710916240323556
Training loss per 100 training steps: 0.01710935928390678
Training loss per 100 training steps: 0.017121690550648327
Training loss per 100 training steps: 0.01715469593583425
Training loss per 100 training steps: 0.01714672674487889
Training loss per 100 training steps: 0.017146488567313423
Training loss per 100 training steps: 0.017170133854098532
Training loss per 100 training steps: 0.017158384640739665
Training loss per 100 training steps: 0.017118412700162757
Training loss per 100 training steps: 0.017150113533373194
Training loss per 100 training steps: 0.017132634910398
Training loss per 100 training steps: 0.01716108120130167
Training loss per 100 training steps: 0.017170390833485392
Training loss per 100 training steps: 0.017182024206948754
Training loss per 100 training steps: 0.017159963078515476
Training loss per 100 training steps: 0.01718975677862775
Training loss per 100 training steps: 0.01716252050033583
Training loss per 100 training steps: 0.017115507817154423
Training loss per 100 training steps: 0.017124373283800922
Training loss epoch: 0.01711671398995961
Training accuracy epoch: 0.9725447265544551
Validation loss per 100 evaluation steps: 0.010581364855170250
Validation loss per 100 evaluation steps: 0.024165651609569024
Validation loss per 100 evaluation steps: 0.024573006853548143
Validation loss per 100 evaluation steps: 0.02372654315395659
Validation loss per 100 evaluation steps: 0.02344027283475032
```

```
Validation loss per 100 evaluation steps: 0.023843694116733816
Validation loss per 100 evaluation steps: 0.024500677639170045
Validation loss per 100 evaluation steps: 0.024342405478481943
Validation loss per 100 evaluation steps: 0.02447594683743025
Validation loss per 100 evaluation steps: 0.02444769258739029
Validation loss per 100 evaluation steps: 0.02468318550876262
Validation loss per 100 evaluation steps: 0.02430833477218676
Validation loss per 100 evaluation steps: 0.023550845145514778
Validation loss per 100 evaluation steps: 0.023502102935316356
Validation loss per 100 evaluation steps: 0.023681276633383288
Validation loss per 100 evaluation steps: 0.02334458955742348
Validation loss per 100 evaluation steps: 0.02314528344915941
Validation loss per 100 evaluation steps: 0.023098730333681892
Validation loss per 100 evaluation steps: 0.023023442280022455
Validation loss per 100 evaluation steps: 0.022640225974944708
Validation loss per 100 evaluation steps: 0.022394911496676464
Validation loss per 100 evaluation steps: 0.022285850981167477
Validation loss per 100 evaluation steps: 0.022399703084036764
Validation loss per 100 evaluation steps: 0.02266906737575147
Validation loss per 100 evaluation steps: 0.022822241950129076
Validation loss per 100 evaluation steps: 0.0229396666129759
Validation loss per 100 evaluation steps: 0.02320834590471291
Validation loss per 100 evaluation steps: 0.02340496415508817
Validation loss per 100 evaluation steps: 0.023198947742591777
Validation loss per 100 evaluation steps: 0.023014766671506058
Validation loss per 100 evaluation steps: 0.022806068587462574
Validation loss per 100 evaluation steps: 0.022975736070773762
Validation loss per 100 evaluation steps: 0.02294369486767404
Validation loss per 100 evaluation steps: 0.02294194069923377
Validation loss per 100 evaluation steps: 0.02307446334530065
Validation loss per 100 evaluation steps: 0.02310814132437828
Validation loss per 100 evaluation steps: 0.023176737587979133
Validation loss per 100 evaluation steps: 0.02316481554846403
Validation loss per 100 evaluation steps: 0.023097898487492646
Validation loss per 100 evaluation steps: 0.023146042424613614
Validation loss per 100 evaluation steps: 0.023102613279028832
Validation loss per 100 evaluation steps: 0.023118799926560428
Validation loss per 100 evaluation steps: 0.023206645974178975
Validation loss per 100 evaluation steps: 0.0231731632499616
Validation loss per 100 evaluation steps: 0.023198448653898033
Validation loss per 100 evaluation steps: 0.02333109254195539
Validation loss per 100 evaluation steps: 0.023281885856041126
Validation loss per 100 evaluation steps: 0.023467303975611677
Validation Loss: 0.023469813949617595
Validation Accuracy: 0.966020019121117
Training epoch: 4
Training loss per 100 training steps: 0.0071370238438248634
Training loss per 100 training steps: 0.011467306899257802
```

```
Training loss per 100 training steps: 0.011925876028052432
Training loss per 100 training steps: 0.011813444540903793
Training loss per 100 training steps: 0.0120236799301804
Training loss per 100 training steps: 0.012221912526531773
Training loss per 100 training steps: 0.012153747216280937
Training loss per 100 training steps: 0.012049302041298567
Training loss per 100 training steps: 0.012033184958271706
Training loss per 100 training steps: 0.012217153749481761
Training loss per 100 training steps: 0.012242237169519396
Training loss per 100 training steps: 0.012288659928621446
Training loss per 100 training steps: 0.012262558348164258
Training loss per 100 training steps: 0.012207062046614712
Training loss per 100 training steps: 0.012419725823149395
Training loss per 100 training steps: 0.012396368984906012
Training loss per 100 training steps: 0.012317605829707574
Training loss per 100 training steps: 0.01240817344722837
Training loss per 100 training steps: 0.012375530529804555
Training loss per 100 training steps: 0.01234479785264905
Training loss per 100 training steps: 0.012298897740515475
Training loss per 100 training steps: 0.012296722933922784
Training loss per 100 training steps: 0.012391424586330116
Training loss per 100 training steps: 0.012393708722160036
Training loss per 100 training steps: 0.01250572839610363
Training loss per 100 training steps: 0.012604235794615157
Training loss per 100 training steps: 0.01266149811832217
Training loss per 100 training steps: 0.01273067554975286
Training loss per 100 training steps: 0.012710367964675218
Training loss per 100 training steps: 0.012692881598353786
Training loss per 100 training steps: 0.012677885349567847
Training loss per 100 training steps: 0.012602114222128584
Training loss per 100 training steps: 0.012602711970240309
Training loss per 100 training steps: 0.012574752024939389
Training loss per 100 training steps: 0.012555701443420483
Training loss per 100 training steps: 0.01252532930615784
Training loss per 100 training steps: 0.012528372100222012
Training loss per 100 training steps: 0.012538183172049524
Training loss per 100 training steps: 0.012525664864354587
Training loss per 100 training steps: 0.012506013402263818
Training loss per 100 training steps: 0.012505409407830936
Training loss per 100 training steps: 0.012495675684642623
Training loss per 100 training steps: 0.012483733182277024
Training loss per 100 training steps: 0.01250951022479397
Training loss per 100 training steps: 0.012456983248735221
Training loss per 100 training steps: 0.012532708767882808
Training loss per 100 training steps: 0.012557871742701295
Training loss per 100 training steps: 0.012598013178702369
Training loss per 100 training steps: 0.012577740407254902
Training loss per 100 training steps: 0.012596576874093584
```

```
Training loss per 100 training steps: 0.012604121219695124
Training loss per 100 training steps: 0.012589956468000709
Training loss per 100 training steps: 0.01257396162172911
Training loss per 100 training steps: 0.012517594886158388
Training loss per 100 training steps: 0.01255342515966976
Training loss per 100 training steps: 0.012582302941948518
Training loss per 100 training steps: 0.012582563910642976
Training loss per 100 training steps: 0.012583619838550109
Training loss per 100 training steps: 0.012590133545916083
Training loss per 100 training steps: 0.012591398845856855
Training loss per 100 training steps: 0.012578027292400524
Training loss per 100 training steps: 0.012569114757805906
Training loss per 100 training steps: 0.012551715096290023
Training loss per 100 training steps: 0.012564177759511887
Training loss per 100 training steps: 0.012584718401676166
Training loss per 100 training steps: 0.012556069893306482
Training loss per 100 training steps: 0.012543821470849674
Training loss per 100 training steps: 0.012593087496886618
Training loss per 100 training steps: 0.012572714242606005
Training loss per 100 training steps: 0.012587579491291865
Training loss per 100 training steps: 0.012631816893604153
Training loss per 100 training steps: 0.012661208381344772
Training loss per 100 training steps: 0.012648034958240373
Training loss per 100 training steps: 0.012631606239674634
Training loss per 100 training steps: 0.012613842222066806
Training loss per 100 training steps: 0.012611945691674085
Training loss per 100 training steps: 0.01262838172473705
Training loss per 100 training steps: 0.012631092443728509
Training loss per 100 training steps: 0.012630098490330607
Training loss per 100 training steps: 0.012642436745976827
Training loss per 100 training steps: 0.012671672878374335
Training loss per 100 training steps: 0.012683969395248428
Training loss per 100 training steps: 0.01266535581741784
Training loss per 100 training steps: 0.012677998122795792
Training loss per 100 training steps: 0.012690155559700654
Training loss per 100 training steps: 0.012721300479398555
Training loss per 100 training steps: 0.012758819454398435
Training loss per 100 training steps: 0.012756567525323307
Training loss per 100 training steps: 0.012771144957175856
Training loss per 100 training steps: 0.012757575464067509
Training loss per 100 training steps: 0.01277544005866301
Training loss per 100 training steps: 0.012782181966462516
Training loss per 100 training steps: 0.012783949136051572
Training loss per 100 training steps: 0.012792115444067631
Training loss per 100 training steps: 0.012790364887948116
Training loss per 100 training steps: 0.012813202496901566
Training loss epoch: 0.012822100012060314
Training accuracy epoch: 0.9790397227316743
```

```
Validation loss per 100 evaluation steps: 0.0033164136111736298
Validation loss per 100 evaluation steps: 0.030398611336263773
Validation loss per 100 evaluation steps: 0.02661494102083461
Validation loss per 100 evaluation steps: 0.027785059563498134
Validation loss per 100 evaluation steps: 0.02809301107351174
Validation loss per 100 evaluation steps: 0.028082581024973623
Validation loss per 100 evaluation steps: 0.02676455104248753
Validation loss per 100 evaluation steps: 0.026171769185876603
Validation loss per 100 evaluation steps: 0.02622469855172705
Validation loss per 100 evaluation steps: 0.026014201629575608
Validation loss per 100 evaluation steps: 0.02657764436224426
Validation loss per 100 evaluation steps: 0.026211797515586943
Validation loss per 100 evaluation steps: 0.026128026990657666
Validation loss per 100 evaluation steps: 0.025616503429212373
Validation loss per 100 evaluation steps: 0.02536900247842361
Validation loss per 100 evaluation steps: 0.025491310299203044
Validation loss per 100 evaluation steps: 0.02546741192420028
Validation loss per 100 evaluation steps: 0.024814934103890757
Validation loss per 100 evaluation steps: 0.02454658502445425
Validation loss per 100 evaluation steps: 0.024461384473376577
Validation loss per 100 evaluation steps: 0.024549702462618783
Validation loss per 100 evaluation steps: 0.024618885994862665
Validation loss per 100 evaluation steps: 0.024538669092818096
Validation loss per 100 evaluation steps: 0.024890080276181066
Validation loss per 100 evaluation steps: 0.024958642726034826
Validation loss per 100 evaluation steps: 0.024861916025083204
Validation loss per 100 evaluation steps: 0.024817415225697914
Validation loss per 100 evaluation steps: 0.024996488398735698
Validation loss per 100 evaluation steps: 0.025048571779605722
Validation loss per 100 evaluation steps: 0.02502971896758377
Validation loss per 100 evaluation steps: 0.024945860999022653
Validation loss per 100 evaluation steps: 0.02482885003023613
Validation loss per 100 evaluation steps: 0.02486374124610873
Validation loss per 100 evaluation steps: 0.024675759734326178
Validation loss per 100 evaluation steps: 0.024955690503722556
Validation loss per 100 evaluation steps: 0.024966865410348017
Validation loss per 100 evaluation steps: 0.024942991218670805
Validation loss per 100 evaluation steps: 0.024829411427555938
Validation loss per 100 evaluation steps: 0.024735074878776894
Validation loss per 100 evaluation steps: 0.024629727233616203
Validation loss per 100 evaluation steps: 0.024708349732083985
Validation loss per 100 evaluation steps: 0.024632629832857935
Validation loss per 100 evaluation steps: 0.02471298130977692
Validation loss per 100 evaluation steps: 0.02476838978249591
Validation loss per 100 evaluation steps: 0.024656298085623835
Validation loss per 100 evaluation steps: 0.02467235474779857
Validation loss per 100 evaluation steps: 0.024715216964031964
Validation loss per 100 evaluation steps: 0.024790515436071622
```

```
Validation Loss: 0.02484043785683839
Validation Accuracy: 0.9665633010961615
Training epoch: 5
Training loss per 100 training steps: 0.009961511939764023
Training loss per 100 training steps: 0.009738922913588749
Training loss per 100 training steps: 0.009048360748351459
Training loss per 100 training steps: 0.009117189225441184
Training loss per 100 training steps: 0.009048709389249222
Training loss per 100 training steps: 0.009574517803333741
Training loss per 100 training steps: 0.009303114299523478
Training loss per 100 training steps: 0.009360465462007632
Training loss per 100 training steps: 0.009211036182252596
Training loss per 100 training steps: 0.009453617375643701
Training loss per 100 training steps: 0.009419814952616526
Training loss per 100 training steps: 0.009420989800370686
Training loss per 100 training steps: 0.00948254446220964
Training loss per 100 training steps: 0.009354580316025976
Training loss per 100 training steps: 0.009288216451445374
Training loss per 100 training steps: 0.009250773138332979
Training loss per 100 training steps: 0.009279558053060575
Training loss per 100 training steps: 0.009193465254633432
Training loss per 100 training steps: 0.009168108146619126
Training loss per 100 training steps: 0.009089847018367724
Training loss per 100 training steps: 0.009102312507270928
Training loss per 100 training steps: 0.009028707815182712
Training loss per 100 training steps: 0.009037838245560393
Training loss per 100 training steps: 0.009064621912999612
Training loss per 100 training steps: 0.009079836070622321
Training loss per 100 training steps: 0.009100931607030535
Training loss per 100 training steps: 0.009139455663161038
Training loss per 100 training steps: 0.009165702887002799
Training loss per 100 training steps: 0.009135415288300709
Training loss per 100 training steps: 0.009154035556845217
Training loss per 100 training steps: 0.009170146038429066
Training loss per 100 training steps: 0.009212267854056945
Training loss per 100 training steps: 0.009215713590598648
Training loss per 100 training steps: 0.009228830967895748
Training loss per 100 training steps: 0.00920712928574975
Training loss per 100 training steps: 0.00916245466158492
Training loss per 100 training steps: 0.009186327004778956
Training loss per 100 training steps: 0.009187129741505715
Training loss per 100 training steps: 0.009208487042927758
Training loss per 100 training steps: 0.009227677967252542
Training loss per 100 training steps: 0.009245184104724996
Training loss per 100 training steps: 0.009288774913077532
Training loss per 100 training steps: 0.009304353042511577
Training loss per 100 training steps: 0.009336632545451976
Training loss per 100 training steps: 0.009362684814119612
```

```
Training loss per 100 training steps: 0.009318263596491504
Training loss per 100 training steps: 0.009320940527719035
Training loss per 100 training steps: 0.009310511472862973
Training loss per 100 training steps: 0.009305128148491574
Training loss per 100 training steps: 0.009334711655040642
Training loss per 100 training steps: 0.009360778324421564
Training loss per 100 training steps: 0.009378710601809188
Training loss per 100 training steps: 0.00936870271321082
Training loss per 100 training steps: 0.009369089530959487
Training loss per 100 training steps: 0.009374620277628136
Training loss per 100 training steps: 0.009405854806184822
Training loss per 100 training steps: 0.00944025145564855
Training loss per 100 training steps: 0.009455861473930554
Training loss per 100 training steps: 0.009483799713078916
Training loss per 100 training steps: 0.009470504057554523
Training loss per 100 training steps: 0.009454547944523807
Training loss per 100 training steps: 0.009463619896746748
Training loss per 100 training steps: 0.009470033275246738
Training loss per 100 training steps: 0.00946125654890442
Training loss per 100 training steps: 0.009473954631904917
Training loss per 100 training steps: 0.00950537934593694
Training loss per 100 training steps: 0.009514412969254966
Training loss per 100 training steps: 0.009557150393084211
Training loss per 100 training steps: 0.009561890384539482
Training loss per 100 training steps: 0.009562852136505198
Training loss per 100 training steps: 0.009565332922493867
Training loss per 100 training steps: 0.009602674348178691
Training loss per 100 training steps: 0.009594175494216262
Training loss per 100 training steps: 0.009595831910735866
Training loss per 100 training steps: 0.009620218847913694
Training loss per 100 training steps: 0.009630078843979166
Training loss per 100 training steps: 0.009648621114750735
Training loss per 100 training steps: 0.009650697915919337
Training loss per 100 training steps: 0.009654975574246192
Training loss per 100 training steps: 0.009667905668733094
Training loss per 100 training steps: 0.009659139271934942
Training loss per 100 training steps: 0.009664344567672295
Training loss per 100 training steps: 0.009654630130944182
Training loss per 100 training steps: 0.009665049587725712
Training loss per 100 training steps: 0.009666217461025054
Training loss per 100 training steps: 0.009666180390778814
Training loss per 100 training steps: 0.009678496952616708
Training loss per 100 training steps: 0.009672189614817426
Training loss per 100 training steps: 0.009679882520770301
Training loss per 100 training steps: 0.009693222212531428
Training loss per 100 training steps: 0.009682136630919415
Training loss per 100 training steps: 0.009680411909262108
Training loss per 100 training steps: 0.009680851289594648
```

```
Training loss per 100 training steps: 0.00969246284128513
Training loss per 100 training steps: 0.009710459467448218
Training loss per 100 training steps: 0.009737101295392044
Training loss epoch: 0.00973960443761075
Training accuracy epoch: 0.9839517000554161
Validation loss per 100 evaluation steps: 0.0011780629865825176
Validation loss per 100 evaluation steps: 0.019960600233267848
Validation loss per 100 evaluation steps: 0.020725255992131302
Validation loss per 100 evaluation steps: 0.02510152119274067
Validation loss per 100 evaluation steps: 0.027391930253988925
Validation loss per 100 evaluation steps: 0.028423185561427592
Validation loss per 100 evaluation steps: 0.02884945399619134
Validation loss per 100 evaluation steps: 0.028911608583031073
Validation loss per 100 evaluation steps: 0.02928081433053572
Validation loss per 100 evaluation steps: 0.03002190174768553
Validation loss per 100 evaluation steps: 0.030385476716017566
Validation loss per 100 evaluation steps: 0.029767843095295583
Validation loss per 100 evaluation steps: 0.028952268658981633
Validation loss per 100 evaluation steps: 0.02849402798721774
Validation loss per 100 evaluation steps: 0.028150653365811455
Validation loss per 100 evaluation steps: 0.027855538509922784
Validation loss per 100 evaluation steps: 0.027876529413526776
Validation loss per 100 evaluation steps: 0.027824085646120347
Validation loss per 100 evaluation steps: 0.027679368392247158
Validation loss per 100 evaluation steps: 0.027477307129816067
Validation loss per 100 evaluation steps: 0.027417044740703816
Validation loss per 100 evaluation steps: 0.027597265801142697
Validation loss per 100 evaluation steps: 0.027845748186506425
Validation loss per 100 evaluation steps: 0.027666959159618507
Validation loss per 100 evaluation steps: 0.027919259267198723
Validation loss per 100 evaluation steps: 0.027986885924962648
Validation loss per 100 evaluation steps: 0.028052477253591914
Validation loss per 100 evaluation steps: 0.02791347188305533
Validation loss per 100 evaluation steps: 0.02839992829094532
Validation loss per 100 evaluation steps: 0.028261617645000823
Validation loss per 100 evaluation steps: 0.028174848757409533
Validation loss per 100 evaluation steps: 0.028192736738155012
Validation loss per 100 evaluation steps: 0.028181201758134686
Validation loss per 100 evaluation steps: 0.028323991607654826
Validation loss per 100 evaluation steps: 0.028319285848292564
Validation loss per 100 evaluation steps: 0.028278093505643927
Validation loss per 100 evaluation steps: 0.02836013375053464
Validation loss per 100 evaluation steps: 0.028184962548978145
Validation loss per 100 evaluation steps: 0.028125954242870784
Validation loss per 100 evaluation steps: 0.028197742788147104
Validation loss per 100 evaluation steps: 0.028108647849849722
Validation loss per 100 evaluation steps: 0.028075089901954446
Validation loss per 100 evaluation steps: 0.028039383912214437
```

```
Validation loss per 100 evaluation steps: 0.027887641273623854
Validation loss per 100 evaluation steps: 0.027786500398336127
Validation loss per 100 evaluation steps: 0.02781656490343658
Validation loss per 100 evaluation steps: 0.027875856596201187
Validation loss per 100 evaluation steps: 0.027855456718546976
Validation Loss: 0.027842749859228043
Validation Accuracy: 0.9661702786630613
Training epoch: 6
Training loss per 100 training steps: 0.00015038761193864048
Training loss per 100 training steps: 0.005786523427514562
Training loss per 100 training steps: 0.006233416985377521
Training loss per 100 training steps: 0.006654126534820147
Training loss per 100 training steps: 0.006631774495538972
Training loss per 100 training steps: 0.006473432880920519
Training loss per 100 training steps: 0.006675900856011951
Training loss per 100 training steps: 0.006661630514745528
Training loss per 100 training steps: 0.006728428074029523
Training loss per 100 training steps: 0.006781678696111446
Training loss per 100 training steps: 0.0068758304906195436
Training loss per 100 training steps: 0.00683482902491732
Training loss per 100 training steps: 0.006779397471993304
Training loss per 100 training steps: 0.006849056859520057
Training loss per 100 training steps: 0.006828472674734763
Training loss per 100 training steps: 0.006849550768590472
Training loss per 100 training steps: 0.0068460547980709685
Training loss per 100 training steps: 0.006871937001839387
Training loss per 100 training steps: 0.006839364192853667
Training loss per 100 training steps: 0.006895473285726834
Training loss per 100 training steps: 0.006970527151170954
Training loss per 100 training steps: 0.007012355846341634
Training loss per 100 training steps: 0.007053150919902157
Training loss per 100 training steps: 0.007155670663382639
Training loss per 100 training steps: 0.007128125832707113
Training loss per 100 training steps: 0.00709009274003486
Training loss per 100 training steps: 0.007087311856877403
Training loss per 100 training steps: 0.007106201676843499
Training loss per 100 training steps: 0.007115407197332963
Training loss per 100 training steps: 0.007108424388440165
Training loss per 100 training steps: 0.007139002073244456
Training loss per 100 training steps: 0.007123873562437642
Training loss per 100 training steps: 0.007113391068261742
Training loss per 100 training steps: 0.007084954954344299
Training loss per 100 training steps: 0.007079432309027003
Training loss per 100 training steps: 0.007083423114134751
Training loss per 100 training steps: 0.007075969084418501
Training loss per 100 training steps: 0.0071338204710651495
Training loss per 100 training steps: 0.007184574390278102
Training loss per 100 training steps: 0.007201126114372893
```

```
Training loss per 100 training steps: 0.007201400969118407
Training loss per 100 training steps: 0.007189133190193599
Training loss per 100 training steps: 0.007196975151366534
Training loss per 100 training steps: 0.007192681672043744
Training loss per 100 training steps: 0.0071991795026390476
Training loss per 100 training steps: 0.007217828917937051
Training loss per 100 training steps: 0.0072137030841404465
Training loss per 100 training steps: 0.007193938332612191
Training loss per 100 training steps: 0.007203684066133834
Training loss per 100 training steps: 0.007221561330879551
Training loss per 100 training steps: 0.007241091475902878
Training loss per 100 training steps: 0.0072355929498496
Training loss per 100 training steps: 0.0072331062348103435
Training loss per 100 training steps: 0.007270801653701899
Training loss per 100 training steps: 0.00726200826177625
Training loss per 100 training steps: 0.007277597538406263
Training loss per 100 training steps: 0.0072701417960927
Training loss per 100 training steps: 0.007273397630218718
Training loss per 100 training steps: 0.007293623963853503
Training loss per 100 training steps: 0.007288882086195794
Training loss per 100 training steps: 0.007283446683458238
Training loss per 100 training steps: 0.007292523985215729
Training loss per 100 training steps: 0.007296566589272054
Training loss per 100 training steps: 0.007303602303365579
Training loss per 100 training steps: 0.0072824412105933845
Training loss per 100 training steps: 0.007291769543858204
Training loss per 100 training steps: 0.007256237479190439
Training loss per 100 training steps: 0.007275334922959286
Training loss per 100 training steps: 0.007283462934114177
Training loss per 100 training steps: 0.007283096481630301
Training loss per 100 training steps: 0.007287402964600932
Training loss per 100 training steps: 0.007318407504906485
Training loss per 100 training steps: 0.007333466778110349
Training loss per 100 training steps: 0.007318563333336608
Training loss per 100 training steps: 0.0073390853313519515
Training loss per 100 training steps: 0.007352153374287661
Training loss per 100 training steps: 0.0073622775383596325
Training loss per 100 training steps: 0.007370993588078835
Training loss per 100 training steps: 0.0073985536804465025
Training loss per 100 training steps: 0.007409926032081573
Training loss per 100 training steps: 0.007424654852835952
Training loss per 100 training steps: 0.007427135803650829
Training loss per 100 training steps: 0.007449617666299786
Training loss per 100 training steps: 0.00745751460686815
Training loss per 100 training steps: 0.007476443759841463
Training loss per 100 training steps: 0.007475687731941339
Training loss per 100 training steps: 0.007467006669965157
Training loss per 100 training steps: 0.007475981530998689
```

```
Training loss per 100 training steps: 0.007463738662826714
Training loss per 100 training steps: 0.007474911655191301
Training loss per 100 training steps: 0.007455887808668091
Training loss per 100 training steps: 0.007468216173001869
Training loss per 100 training steps: 0.007472125177735238
Training loss per 100 training steps: 0.007479582038296391
Training loss per 100 training steps: 0.007496391697014527
Training loss per 100 training steps: 0.007495567329570694
Training loss epoch: 0.007495418797864272
Training accuracy epoch: 0.9877923093563183
Validation loss per 100 evaluation steps: 0.14575059711933136
Validation loss per 100 evaluation steps: 0.030032387585239607
Validation loss per 100 evaluation steps: 0.030355535356362966
Validation loss per 100 evaluation steps: 0.028128459188984347
Validation loss per 100 evaluation steps: 0.02933457694101767
Validation loss per 100 evaluation steps: 0.029230514052881493
Validation loss per 100 evaluation steps: 0.028541145617557426
Validation loss per 100 evaluation steps: 0.028792691638281135
Validation loss per 100 evaluation steps: 0.029342851360346114
Validation loss per 100 evaluation steps: 0.02909201736098681
Validation loss per 100 evaluation steps: 0.02896377755212018
Validation loss per 100 evaluation steps: 0.029672948689151844
Validation loss per 100 evaluation steps: 0.029783254697584118
Validation loss per 100 evaluation steps: 0.029564561774626078
Validation loss per 100 evaluation steps: 0.02999682691877227
Validation loss per 100 evaluation steps: 0.030702785959163086
Validation loss per 100 evaluation steps: 0.030466089761017594
Validation loss per 100 evaluation steps: 0.030622281831741482
Validation loss per 100 evaluation steps: 0.030539824107175782
Validation loss per 100 evaluation steps: 0.030572430956676607
Validation loss per 100 evaluation steps: 0.030849825400188425
Validation loss per 100 evaluation steps: 0.031187299735236103
Validation loss per 100 evaluation steps: 0.030884458971507295
Validation loss per 100 evaluation steps: 0.030797411820151444
Validation loss per 100 evaluation steps: 0.031055427576488
Validation loss per 100 evaluation steps: 0.031120877999316565
Validation loss per 100 evaluation steps: 0.03127161299536468
Validation loss per 100 evaluation steps: 0.031207717410773817
Validation loss per 100 evaluation steps: 0.031056341232434684
Validation loss per 100 evaluation steps: 0.03109021754490352
Validation loss per 100 evaluation steps: 0.03106091265636693
Validation loss per 100 evaluation steps: 0.030733847350410112
Validation loss per 100 evaluation steps: 0.030611459410613644
Validation loss per 100 evaluation steps: 0.030349766813734885
Validation loss per 100 evaluation steps: 0.03037196334986312
Validation loss per 100 evaluation steps: 0.030464904785906143
Validation loss per 100 evaluation steps: 0.03070535588450348
Validation loss per 100 evaluation steps: 0.030737624928435025
```

```
Validation loss per 100 evaluation steps: 0.031033477450487554
Validation loss per 100 evaluation steps: 0.030921482234474523
Validation loss per 100 evaluation steps: 0.030889031282672126
Validation loss per 100 evaluation steps: 0.030851042982784085
Validation loss per 100 evaluation steps: 0.030967477634024125
Validation loss per 100 evaluation steps: 0.031080883369338737
Validation loss per 100 evaluation steps: 0.03105044886386829
Validation loss per 100 evaluation steps: 0.030903148500660615
Validation loss per 100 evaluation steps: 0.03089758330401291
Validation loss per 100 evaluation steps: 0.030819823386707714
Validation Loss: 0.030817296409346766
Validation Accuracy: 0.9659824546413904
Training epoch: 7
Training loss per 100 training steps: 0.0013876971788704395
Training loss per 100 training steps: 0.005424394722800877
Training loss per 100 training steps: 0.004863306167442091
Training loss per 100 training steps: 0.005341868173953999
Training loss per 100 training steps: 0.005662974832647934
Training loss per 100 training steps: 0.005754531012679148
Training loss per 100 training steps: 0.005796683210676031
Training loss per 100 training steps: 0.005695408579753122
Training loss per 100 training steps: 0.005569045070387233
Training loss per 100 training steps: 0.005449125442166736
Training loss per 100 training steps: 0.0055253912019756326
Training loss per 100 training steps: 0.005562856628146538
Training loss per 100 training steps: 0.005532193668350782
Training loss per 100 training steps: 0.005497178722930357
Training loss per 100 training steps: 0.0054693113016608405
Training loss per 100 training steps: 0.005503145826765906
Training loss per 100 training steps: 0.005496880215258007
Training loss per 100 training steps: 0.005527642359812222
Training loss per 100 training steps: 0.005506661426926797
Training loss per 100 training steps: 0.005500373148924418
Training loss per 100 training steps: 0.005528582503038156
Training loss per 100 training steps: 0.00550137155438592
Training loss per 100 training steps: 0.005494735787867401
Training loss per 100 training steps: 0.005461476377541962
Training loss per 100 training steps: 0.0054664395398395646
Training loss per 100 training steps: 0.00550012269723309
Training loss per 100 training steps: 0.005523907662750754
Training loss per 100 training steps: 0.005525827452596445
Training loss per 100 training steps: 0.0055490609845590365
Training loss per 100 training steps: 0.005532928171296279
Training loss per 100 training steps: 0.005564321455378718
Training loss per 100 training steps: 0.005532301612760715
Training loss per 100 training steps: 0.005558440042562696
Training loss per 100 training steps: 0.005551833056375539
Training loss per 100 training steps: 0.005570854378749337
```

```
Training loss per 100 training steps: 0.005570949096640755
Training loss per 100 training steps: 0.005578439716928791
Training loss per 100 training steps: 0.005571533616446827
Training loss per 100 training steps: 0.00557055735385425
Training loss per 100 training steps: 0.005582412485981532
Training loss per 100 training steps: 0.005581035949715954
Training loss per 100 training steps: 0.005556687761618417
Training loss per 100 training steps: 0.005574145269766801
Training loss per 100 training steps: 0.005550810498456309
Training loss per 100 training steps: 0.005561508889432366
Training loss per 100 training steps: 0.005557998918855558
Training loss per 100 training steps: 0.005581820110893663
Training loss per 100 training steps: 0.005610313336742033
Training loss per 100 training steps: 0.005623391843404426
Training loss per 100 training steps: 0.0056355666197692176
Training loss per 100 training steps: 0.005656044906401837
Training loss per 100 training steps: 0.005661510093449325
Training loss per 100 training steps: 0.005671805686295527
Training loss per 100 training steps: 0.0057063474982892584
Training loss per 100 training steps: 0.005712786414290734
Training loss per 100 training steps: 0.005729184007070805
Training loss per 100 training steps: 0.005732389542752571
Training loss per 100 training steps: 0.005745776904104521
Training loss per 100 training steps: 0.005735884646794096
Training loss per 100 training steps: 0.005733008942608288
Training loss per 100 training steps: 0.005741303559076521
Training loss per 100 training steps: 0.005791864208045235
Training loss per 100 training steps: 0.00578487620610027
Training loss per 100 training steps: 0.00579338439305757
Training loss per 100 training steps: 0.005784344209263118
Training loss per 100 training steps: 0.005762786695563777
Training loss per 100 training steps: 0.005747361856933307
Training loss per 100 training steps: 0.005742049779870905
Training loss per 100 training steps: 0.005751183111972418
Training loss per 100 training steps: 0.005736930808351298
Training loss per 100 training steps: 0.005753547379054098
Training loss per 100 training steps: 0.005758774799385531
Training loss per 100 training steps: 0.00574404214592595
Training loss per 100 training steps: 0.005742298535892229
Training loss per 100 training steps: 0.005747515181160881
Training loss per 100 training steps: 0.005768174075537767
Training loss per 100 training steps: 0.0057607533604984214
Training loss per 100 training steps: 0.005755126757894967
Training loss per 100 training steps: 0.0057875053966572245
Training loss per 100 training steps: 0.005775185503011509
Training loss per 100 training steps: 0.005775195628660027
Training loss per 100 training steps: 0.005782619741581643
Training loss per 100 training steps: 0.005785189639216496
```

```
Training loss per 100 training steps: 0.005802096535812878
Training loss per 100 training steps: 0.005803004237116193
Training loss per 100 training steps: 0.005800261976020409
Training loss per 100 training steps: 0.005803987342560545
Training loss per 100 training steps: 0.0058025475692889215
Training loss per 100 training steps: 0.005816581935050791
Training loss per 100 training steps: 0.005810368362605848
Training loss per 100 training steps: 0.0058284794248809595
Training loss per 100 training steps: 0.00583317401809539
Training loss per 100 training steps: 0.005824804200914393
Training loss per 100 training steps: 0.005821394135508017
Training loss per 100 training steps: 0.005822756552130238
Training loss per 100 training steps: 0.0058187998184034495
Training loss epoch: 0.0058208385111663445
Training accuracy epoch: 0.9903924454262886
Validation loss per 100 evaluation steps: 0.0427141934633255
Validation loss per 100 evaluation steps: 0.030407587647624218
Validation loss per 100 evaluation steps: 0.028521877612286233
Validation loss per 100 evaluation steps: 0.02990378247481569
Validation loss per 100 evaluation steps: 0.030317535475423395
Validation loss per 100 evaluation steps: 0.03174009729726897
Validation loss per 100 evaluation steps: 0.03283941753193062
Validation loss per 100 evaluation steps: 0.03229738305965504
Validation loss per 100 evaluation steps: 0.03207463573916199
Validation loss per 100 evaluation steps: 0.031363880859057426
Validation loss per 100 evaluation steps: 0.030833817611000524
Validation loss per 100 evaluation steps: 0.030105864897293836
Validation loss per 100 evaluation steps: 0.030209886211887534
Validation loss per 100 evaluation steps: 0.03055363517384441
Validation loss per 100 evaluation steps: 0.031057869223396845
Validation loss per 100 evaluation steps: 0.032077437856662734
Validation loss per 100 evaluation steps: 0.032179455240091824
Validation loss per 100 evaluation steps: 0.03279916245502705
Validation loss per 100 evaluation steps: 0.03402430572853252
Validation loss per 100 evaluation steps: 0.034523061839687785
Validation loss per 100 evaluation steps: 0.034366311562874945
Validation loss per 100 evaluation steps: 0.03423194769555129
Validation loss per 100 evaluation steps: 0.034327238684470136
Validation loss per 100 evaluation steps: 0.033987282826785634
Validation loss per 100 evaluation steps: 0.033780623730213545
Validation loss per 100 evaluation steps: 0.03350965553991085
Validation loss per 100 evaluation steps: 0.03336870290020475
Validation loss per 100 evaluation steps: 0.033351658773973114
Validation loss per 100 evaluation steps: 0.03364775297574872
Validation loss per 100 evaluation steps: 0.033714884172928686
Validation loss per 100 evaluation steps: 0.03362771986295111
Validation loss per 100 evaluation steps: 0.03337171986769123
Validation loss per 100 evaluation steps: 0.03333865694456024
```

```
Validation loss per 100 evaluation steps: 0.033378603687689246
Validation loss per 100 evaluation steps: 0.033398952630581076
Validation loss per 100 evaluation steps: 0.033345200652383986
Validation loss per 100 evaluation steps: 0.033203317883392924
Validation loss per 100 evaluation steps: 0.03318127517391521
Validation loss per 100 evaluation steps: 0.03343739930706379
Validation loss per 100 evaluation steps: 0.0333786326069761
Validation loss per 100 evaluation steps: 0.033467804563096136
Validation loss per 100 evaluation steps: 0.03362878769113457
Validation loss per 100 evaluation steps: 0.03365311100324352
Validation loss per 100 evaluation steps: 0.03374594066126989
Validation loss per 100 evaluation steps: 0.03359719333679458
Validation loss per 100 evaluation steps: 0.0336580240519781
Validation loss per 100 evaluation steps: 0.03379125081800369
Validation loss per 100 evaluation steps: 0.03385612242042727
Validation Loss: 0.033756910293431405
Validation Accuracy: 0.9651245958975792
Training epoch: 8
Training loss per 100 training steps: 0.008960126899182796
Training loss per 100 training steps: 0.003597295729116805
Training loss per 100 training steps: 0.003722521482972415
Training loss per 100 training steps: 0.0035502652527880565
Training loss per 100 training steps: 0.003926460359835708
Training loss per 100 training steps: 0.003941069699644441
Training loss per 100 training steps: 0.003890280245518239
Training loss per 100 training steps: 0.00392855367969246
Training loss per 100 training steps: 0.003939760803602648
Training loss per 100 training steps: 0.003926824833655252
Training loss per 100 training steps: 0.004042966020865928
Training loss per 100 training steps: 0.003947384944438853
Training loss per 100 training steps: 0.003976546164739133
Training loss per 100 training steps: 0.0039011104127495044
Training loss per 100 training steps: 0.0039855716126375946
Training loss per 100 training steps: 0.003967306222866688
Training loss per 100 training steps: 0.003973708082655624
Training loss per 100 training steps: 0.004003729077435521
Training loss per 100 training steps: 0.004019851102029553
Training loss per 100 training steps: 0.004052638584078729
Training loss per 100 training steps: 0.004063577773810413
Training loss per 100 training steps: 0.0040476244912640204
Training loss per 100 training steps: 0.004015879019136396
Training loss per 100 training steps: 0.004036547993592518
Training loss per 100 training steps: 0.004060394595090254
Training loss per 100 training steps: 0.004074942876003228
Training loss per 100 training steps: 0.004072701954014553
Training loss per 100 training steps: 0.004141942110612947
Training loss per 100 training steps: 0.004164558207860054
Training loss per 100 training steps: 0.004138022092839815
```

```
Training loss per 100 training steps: 0.004188074810027313
Training loss per 100 training steps: 0.004188388065725355
Training loss per 100 training steps: 0.004224895411133878
Training loss per 100 training steps: 0.004243290232570667
Training loss per 100 training steps: 0.0042514858480799675
Training loss per 100 training steps: 0.00427021918055466
Training loss per 100 training steps: 0.004336586868693962
Training loss per 100 training steps: 0.004345281716579666
Training loss per 100 training steps: 0.004375223534199528
Training loss per 100 training steps: 0.004386754289763142
Training loss per 100 training steps: 0.004362795866882323
Training loss per 100 training steps: 0.004344763354715736
Training loss per 100 training steps: 0.004350315222146829
Training loss per 100 training steps: 0.004355570339416923
Training loss per 100 training steps: 0.004365527218830044
Training loss per 100 training steps: 0.0043940422276621365
Training loss per 100 training steps: 0.004376945617388497
Training loss per 100 training steps: 0.004380773935363208
Training loss per 100 training steps: 0.004388867816597871
Training loss per 100 training steps: 0.004418466864597619
Training loss per 100 training steps: 0.004419112399477004
Training loss per 100 training steps: 0.0044205925661915954
Training loss per 100 training steps: 0.004414852653502706
Training loss per 100 training steps: 0.004403721082024694
Training loss per 100 training steps: 0.004404663870632045
Training loss per 100 training steps: 0.004431001173411566
Training loss per 100 training steps: 0.004440993489884833
Training loss per 100 training steps: 0.0044642040073786474
Training loss per 100 training steps: 0.004462652850657703
Training loss per 100 training steps: 0.0044647577415066675
Training loss per 100 training steps: 0.004458861697913941
Training loss per 100 training steps: 0.0044568969850036794
Training loss per 100 training steps: 0.004465302220680964
Training loss per 100 training steps: 0.004468462882971544
Training loss per 100 training steps: 0.0044685121307784
Training loss per 100 training steps: 0.004472164714129014
Training loss per 100 training steps: 0.00447251493457688
Training loss per 100 training steps: 0.004485973473324749
Training loss per 100 training steps: 0.0044941409541993105
Training loss per 100 training steps: 0.004491826149008444
Training loss per 100 training steps: 0.004502437567587952
Training loss per 100 training steps: 0.004488138620840776
Training loss per 100 training steps: 0.004492304387074064
Training loss per 100 training steps: 0.004481445054193647
Training loss per 100 training steps: 0.004472709839365722
Training loss per 100 training steps: 0.004472324488779474
Training loss per 100 training steps: 0.0044707565515182245
Training loss per 100 training steps: 0.004469626900314629
```

```
Training loss per 100 training steps: 0.004478209079947691
Training loss per 100 training steps: 0.004487261563431406
Training loss per 100 training steps: 0.004486338265656204
Training loss per 100 training steps: 0.004482269643039919
Training loss per 100 training steps: 0.00449227055427566
Training loss per 100 training steps: 0.004492365403888526
Training loss per 100 training steps: 0.004500342404904471
Training loss per 100 training steps: 0.004520800845058575
Training loss per 100 training steps: 0.0045173625050592935
Training loss per 100 training steps: 0.004525151611561491
Training loss per 100 training steps: 0.004542909071437916
Training loss per 100 training steps: 0.004549837351763635
Training loss per 100 training steps: 0.004552418058139168
Training loss per 100 training steps: 0.004562209218042823
Training loss per 100 training steps: 0.004574520462763322
Training loss per 100 training steps: 0.004571772177281461
Training loss per 100 training steps: 0.004563529874160942
Training loss per 100 training steps: 0.004559788110969968
Training loss epoch: 0.004558326844121083
Training accuracy epoch: 0.992390585231023
Validation loss per 100 evaluation steps: 0.14374187588691711
Validation loss per 100 evaluation steps: 0.050454349011488375
Validation loss per 100 evaluation steps: 0.04438186451361345
Validation loss per 100 evaluation steps: 0.04618622367463158
Validation loss per 100 evaluation steps: 0.04226588167406456
Validation loss per 100 evaluation steps: 0.04337827375915496
Validation loss per 100 evaluation steps: 0.04210148618775242
Validation loss per 100 evaluation steps: 0.04027564413004983
Validation loss per 100 evaluation steps: 0.04003089278189157
Validation loss per 100 evaluation steps: 0.039978734809908205
Validation loss per 100 evaluation steps: 0.040430555878187294
Validation loss per 100 evaluation steps: 0.04014585779444647
Validation loss per 100 evaluation steps: 0.03991594504063937
Validation loss per 100 evaluation steps: 0.039613584360990516
Validation loss per 100 evaluation steps: 0.039975332248469025
Validation loss per 100 evaluation steps: 0.03946469796352611
Validation loss per 100 evaluation steps: 0.039201715012015526
Validation loss per 100 evaluation steps: 0.03942368986091121
Validation loss per 100 evaluation steps: 0.03942241368192601
Validation loss per 100 evaluation steps: 0.03909775499705696
Validation loss per 100 evaluation steps: 0.0390236816090579
Validation loss per 100 evaluation steps: 0.03908013633913082
Validation loss per 100 evaluation steps: 0.038955214079266186
Validation loss per 100 evaluation steps: 0.038788983364384855
Validation loss per 100 evaluation steps: 0.03907315436841677
Validation loss per 100 evaluation steps: 0.03891215829515808
Validation loss per 100 evaluation steps: 0.03883533220745793
Validation loss per 100 evaluation steps: 0.0383776831607091
```

```
Validation loss per 100 evaluation steps: 0.03807304080721473
Validation loss per 100 evaluation steps: 0.03832323273223663
Validation loss per 100 evaluation steps: 0.03797626283553362
Validation loss per 100 evaluation steps: 0.03786948225477648
Validation loss per 100 evaluation steps: 0.03796337143767146
Validation loss per 100 evaluation steps: 0.037965152110640873
Validation loss per 100 evaluation steps: 0.0378883545888163
Validation loss per 100 evaluation steps: 0.0378881956615635
Validation loss per 100 evaluation steps: 0.03768588694008854
Validation loss per 100 evaluation steps: 0.03764545740768526
Validation loss per 100 evaluation steps: 0.03766873282076361
Validation loss per 100 evaluation steps: 0.037884798902635256
Validation loss per 100 evaluation steps: 0.03794572525110732
Validation loss per 100 evaluation steps: 0.037986983942027264
Validation loss per 100 evaluation steps: 0.03791374760644818
Validation loss per 100 evaluation steps: 0.03781423283293904
Validation loss per 100 evaluation steps: 0.037717603608004045
Validation loss per 100 evaluation steps: 0.03763481661283074
Validation loss per 100 evaluation steps: 0.03763710599283
Validation loss per 100 evaluation steps: 0.03733915027850871
Validation Loss: 0.03732937927690563
Validation Accuracy: 0.9648887493628124
Training epoch: 9
Training loss per 100 training steps: 0.005970652215182781
Training loss per 100 training steps: 0.0031100956917671824
Training loss per 100 training steps: 0.0032688757875703727
Training loss per 100 training steps: 0.0030253810155498593
Training loss per 100 training steps: 0.0029701777343791235
Training loss per 100 training steps: 0.0030275735294770004
Training loss per 100 training steps: 0.0031220354971742796
Training loss per 100 training steps: 0.0033741946306900987
Training loss per 100 training steps: 0.0033540760771117793
Training loss per 100 training steps: 0.0033533136009940463
Training loss per 100 training steps: 0.0034124305252857725
Training loss per 100 training steps: 0.0033362435647181464
Training loss per 100 training steps: 0.0033815393540161474
Training loss per 100 training steps: 0.003340259982679491
Training loss per 100 training steps: 0.003387700096957021
Training loss per 100 training steps: 0.003410779312595589
Training loss per 100 training steps: 0.0034117779960910794
Training loss per 100 training steps: 0.003399805139377607
Training loss per 100 training steps: 0.003350304134624111
Training loss per 100 training steps: 0.003373026428898634
Training loss per 100 training steps: 0.003392726700080582
Training loss per 100 training steps: 0.0034211080101388017
Training loss per 100 training steps: 0.0034930821681962414
Training loss per 100 training steps: 0.0035132846485150804
Training loss per 100 training steps: 0.003479702860161845
```

```
Training loss per 100 training steps: 0.003458363226721175
Training loss per 100 training steps: 0.003452530807041933
Training loss per 100 training steps: 0.0035156786127777118
Training loss per 100 training steps: 0.0035041472979629915
Training loss per 100 training steps: 0.0034867362014720424
Training loss per 100 training steps: 0.003478578318640007
Training loss per 100 training steps: 0.0034724052208476145
Training loss per 100 training steps: 0.0035019590272615717
Training loss per 100 training steps: 0.0034831689696126866
Training loss per 100 training steps: 0.003473880570158146
Training loss per 100 training steps: 0.003453580693208058
Training loss per 100 training steps: 0.0034231776001314436
Training loss per 100 training steps: 0.0034521064138569725
Training loss per 100 training steps: 0.0034831113149851253
Training loss per 100 training steps: 0.003502941888656119
Training loss per 100 training steps: 0.0035311506862556216
Training loss per 100 training steps: 0.0035287383072235877
Training loss per 100 training steps: 0.0035303686219072825
Training loss per 100 training steps: 0.003543461985942713
Training loss per 100 training steps: 0.003558706176679793
Training loss per 100 training steps: 0.003557128422759117
Training loss per 100 training steps: 0.0035474825802457193
Training loss per 100 training steps: 0.003551519932930585
Training loss per 100 training steps: 0.003542021209330246
Training loss per 100 training steps: 0.003563382445610106
Training loss per 100 training steps: 0.003547128575555136
Training loss per 100 training steps: 0.003532766790939112
Training loss per 100 training steps: 0.0035443123562213723
Training loss per 100 training steps: 0.003546527675309535
Training loss per 100 training steps: 0.0035386680204573854
Training loss per 100 training steps: 0.0035614608832405473
Training loss per 100 training steps: 0.003569226276636994
Training loss per 100 training steps: 0.0035931242852626204
Training loss per 100 training steps: 0.0036095109244443234
Training loss per 100 training steps: 0.003628397688675944
Training loss per 100 training steps: 0.0036210790416276017
Training loss per 100 training steps: 0.003625668573099785
Training loss per 100 training steps: 0.003618679180470982
Training loss per 100 training steps: 0.0036158013417018965
Training loss per 100 training steps: 0.003633549315854569
Training loss per 100 training steps: 0.003653859419054871
Training loss per 100 training steps: 0.003659545354251119
Training loss per 100 training steps: 0.003652386121099924
Training loss per 100 training steps: 0.0036467716150038005
Training loss per 100 training steps: 0.0036385460230708155
Training loss per 100 training steps: 0.0036317870079560824
Training loss per 100 training steps: 0.003622100105298589
Training loss per 100 training steps: 0.0036195894340954027
```

```
Training loss per 100 training steps: 0.0036326076240238933
Training loss per 100 training steps: 0.0036480962005762344
Training loss per 100 training steps: 0.003652698552534909
Training loss per 100 training steps: 0.0036449955302083593
Training loss per 100 training steps: 0.0036605136126942875
Training loss per 100 training steps: 0.0036537804917401643
Training loss per 100 training steps: 0.0036553826457663466
Training loss per 100 training steps: 0.0036510633487486987
Training loss per 100 training steps: 0.0036556135400203195
Training loss per 100 training steps: 0.0036509131112877073
Training loss per 100 training steps: 0.00365648882546898
Training loss per 100 training steps: 0.0036570374504377784
Training loss per 100 training steps: 0.0036585813747354503
Training loss per 100 training steps: 0.0036612183230284585
Training loss per 100 training steps: 0.003669452730590019
Training loss per 100 training steps: 0.0036631086487888325
Training loss per 100 training steps: 0.003677123404768607
Training loss per 100 training steps: 0.0036932753511926507
Training loss per 100 training steps: 0.0036975376832612456
Training loss per 100 training steps: 0.0036890337227326138
Training loss per 100 training steps: 0.003703223276853355
Training loss per 100 training steps: 0.0037050224900021904
Training loss per 100 training steps: 0.0037007150515389756
Training loss epoch: 0.003699557676310946
Training accuracy epoch: 0.994024083289638
Validation loss per 100 evaluation steps: 0.01198288518935442
Validation loss per 100 evaluation steps: 0.04164660665931631
Validation loss per 100 evaluation steps: 0.043765910214438125
Validation loss per 100 evaluation steps: 0.038654555669969504
Validation loss per 100 evaluation steps: 0.03945487313092286
Validation loss per 100 evaluation steps: 0.03885089234797565
Validation loss per 100 evaluation steps: 0.03938194063230056
Validation loss per 100 evaluation steps: 0.03802965559819828
Validation loss per 100 evaluation steps: 0.03844013842996013
Validation loss per 100 evaluation steps: 0.03919552541019579
Validation loss per 100 evaluation steps: 0.040293987323007194
Validation loss per 100 evaluation steps: 0.03983794580255181
Validation loss per 100 evaluation steps: 0.03926466314074563
Validation loss per 100 evaluation steps: 0.0385869264898679
Validation loss per 100 evaluation steps: 0.039479466408836025
Validation loss per 100 evaluation steps: 0.039822954607576365
Validation loss per 100 evaluation steps: 0.03948219839792717
Validation loss per 100 evaluation steps: 0.03946774725990033
Validation loss per 100 evaluation steps: 0.03985525057234517
Validation loss per 100 evaluation steps: 0.039558331583443684
Validation loss per 100 evaluation steps: 0.04005854521637846
Validation loss per 100 evaluation steps: 0.039881115426375076
Validation loss per 100 evaluation steps: 0.03959030022375749
```

```
Validation loss per 100 evaluation steps: 0.03886618535107215
Validation loss per 100 evaluation steps: 0.03878893929108286
Validation loss per 100 evaluation steps: 0.03884248006976
Validation loss per 100 evaluation steps: 0.039031873256250134
Validation loss per 100 evaluation steps: 0.0389056006881807
Validation loss per 100 evaluation steps: 0.03888621785136031
Validation loss per 100 evaluation steps: 0.03879012545665163
Validation loss per 100 evaluation steps: 0.038508511377276164
Validation loss per 100 evaluation steps: 0.03845714062302344
Validation loss per 100 evaluation steps: 0.03860020018873264
Validation loss per 100 evaluation steps: 0.03866847913797284
Validation loss per 100 evaluation steps: 0.0387280717679366
Validation loss per 100 evaluation steps: 0.038751397250338854
Validation loss per 100 evaluation steps: 0.0387683953795109
Validation loss per 100 evaluation steps: 0.03866310623612458
Validation loss per 100 evaluation steps: 0.03871234515549903
Validation loss per 100 evaluation steps: 0.0386428521011071
Validation loss per 100 evaluation steps: 0.03857864673710144
Validation loss per 100 evaluation steps: 0.038560119428465565
Validation loss per 100 evaluation steps: 0.03850958250130205
Validation loss per 100 evaluation steps: 0.038612978410944694
Validation loss per 100 evaluation steps: 0.03882594300183585
Validation loss per 100 evaluation steps: 0.03860542617662002
Validation loss per 100 evaluation steps: 0.03868557086995933
Validation loss per 100 evaluation steps: 0.03861632078400639
Validation Loss: 0.0386871123200506
Validation Accuracy: 0.9649197984753808
Training epoch: 10
Training loss per 100 training steps: 4.63135693280492e-05
Training loss per 100 training steps: 0.0024285856876430094
Training loss per 100 training steps: 0.002294521637384207
Training loss per 100 training steps: 0.0022588347125394007
Training loss per 100 training steps: 0.00245587386534839
Training loss per 100 training steps: 0.002638917317796483
Training loss per 100 training steps: 0.0027970814885028843
Training loss per 100 training steps: 0.002897466414707506
Training loss per 100 training steps: 0.002797972235530142
Training loss per 100 training steps: 0.002864337137919682
Training loss per 100 training steps: 0.002869973320282899
Training loss per 100 training steps: 0.0028141424730361232
Training loss per 100 training steps: 0.002825608185589321
Training loss per 100 training steps: 0.002821742472666241
Training loss per 100 training steps: 0.002843275768903314
Training loss per 100 training steps: 0.0028795165295671223
Training loss per 100 training steps: 0.0029134399684453837
Training loss per 100 training steps: 0.0029981691329303986
Training loss per 100 training steps: 0.002992903333792374
Training loss per 100 training steps: 0.003009640897089517
```

```
Training loss per 100 training steps: 0.003020763619209409
Training loss per 100 training steps: 0.0029778055324159803
Training loss per 100 training steps: 0.002987351938906825
Training loss per 100 training steps: 0.0029674580351727264
Training loss per 100 training steps: 0.002953927689159233
Training loss per 100 training steps: 0.002947035475683885
Training loss per 100 training steps: 0.002914366866281055
Training loss per 100 training steps: 0.002893947012414774
Training loss per 100 training steps: 0.0028980705451647206
Training loss per 100 training steps: 0.0028913373579482872
Training loss per 100 training steps: 0.002862933172645211
Training loss per 100 training steps: 0.002882837508775204
Training loss per 100 training steps: 0.002883299460269505
Training loss per 100 training steps: 0.002882860310264349
Training loss per 100 training steps: 0.0028882456790554857
Training loss per 100 training steps: 0.002888803806696465
Training loss per 100 training steps: 0.002896777605728711
Training loss per 100 training steps: 0.0029032682690808154
Training loss per 100 training steps: 0.0028884060980868846
Training loss per 100 training steps: 0.0028886546910061274
Training loss per 100 training steps: 0.0028897773660415517
Training loss per 100 training steps: 0.0028886536635780204
Training loss per 100 training steps: 0.002911601694467572
Training loss per 100 training steps: 0.0029111737704072705
Training loss per 100 training steps: 0.002901660226607623
Training loss per 100 training steps: 0.0029112572015680153
Training loss per 100 training steps: 0.002902176306429228
Training loss per 100 training steps: 0.00289241855808357
Training loss per 100 training steps: 0.00288744950967801
Training loss per 100 training steps: 0.002880967057816836
Training loss per 100 training steps: 0.002876786229653293
Training loss per 100 training steps: 0.0028581014093198207
Training loss per 100 training steps: 0.0028629914026968317
Training loss per 100 training steps: 0.002843087194750535
Training loss per 100 training steps: 0.0028338871377132043
Training loss per 100 training steps: 0.0028357100594892785
Training loss per 100 training steps: 0.002845784688682759
Training loss per 100 training steps: 0.0028507424691472435
Training loss per 100 training steps: 0.0028524482565434376
Training loss per 100 training steps: 0.0028531365689031035
Training loss per 100 training steps: 0.00286788839463576
Training loss per 100 training steps: 0.002880574440141668
Training loss per 100 training steps: 0.002898088951302089
Training loss per 100 training steps: 0.002915271821539618
Training loss per 100 training steps: 0.0029164891208704902
Training loss per 100 training steps: 0.002924494422196988
Training loss per 100 training steps: 0.002913017051941969
Training loss per 100 training steps: 0.0029136254787021044
```

```
Training loss per 100 training steps: 0.002920361531851242
Training loss per 100 training steps: 0.002912786799421835
Training loss per 100 training steps: 0.002914344245651762
Training loss per 100 training steps: 0.0029100631281365
Training loss per 100 training steps: 0.0029016663731033024
Training loss per 100 training steps: 0.002899301390299039
Training loss per 100 training steps: 0.0029214490742801315
Training loss per 100 training steps: 0.0029229513788229947
Training loss per 100 training steps: 0.002940555093666495
Training loss per 100 training steps: 0.002938383644540575
Training loss per 100 training steps: 0.002935934436487006
Training loss per 100 training steps: 0.0029301417525358967
Training loss per 100 training steps: 0.0029528296778122643
Training loss per 100 training steps: 0.002948684804508805
Training loss per 100 training steps: 0.0029568171817993963
Training loss per 100 training steps: 0.002957350712160758
Training loss per 100 training steps: 0.00296101753781141
Training loss per 100 training steps: 0.0029594672244905107
Training loss per 100 training steps: 0.0029519728785355574
Training loss per 100 training steps: 0.0029416641650906634
Training loss per 100 training steps: 0.0029427732311904874
Training loss per 100 training steps: 0.0029494327473357884
Training loss per 100 training steps: 0.0029542965284993007
Training loss per 100 training steps: 0.0029566434449302516
Training loss per 100 training steps: 0.0029580862783026543
Training loss per 100 training steps: 0.002967208514591682
Training loss per 100 training steps: 0.002973276956023072
Training loss per 100 training steps: 0.002972997423843613
Training loss epoch: 0.0029728769755038745
Training accuracy epoch: 0.9951372686874074
Validation loss per 100 evaluation steps: 0.10544438660144806
Validation loss per 100 evaluation steps: 0.04103631284917833
Validation loss per 100 evaluation steps: 0.039059936944435755
Validation loss per 100 evaluation steps: 0.03961434978644553
Validation loss per 100 evaluation steps: 0.04157857906170988
Validation loss per 100 evaluation steps: 0.04034646960028912
Validation loss per 100 evaluation steps: 0.03949891015996005
Validation loss per 100 evaluation steps: 0.03867473639175643
Validation loss per 100 evaluation steps: 0.03888222530374276
Validation loss per 100 evaluation steps: 0.03928628432049979
Validation loss per 100 evaluation steps: 0.03985100972508045
Validation loss per 100 evaluation steps: 0.03977645806608184
Validation loss per 100 evaluation steps: 0.04091469683208398
Validation loss per 100 evaluation steps: 0.041642153652108195
Validation loss per 100 evaluation steps: 0.04082188437649722
Validation loss per 100 evaluation steps: 0.039931822064393246
Validation loss per 100 evaluation steps: 0.04021154101329605
Validation loss per 100 evaluation steps: 0.04059534718099853
```

```
Validation loss per 100 evaluation steps: 0.04049189639858364
Validation loss per 100 evaluation steps: 0.03994969328752388
Validation loss per 100 evaluation steps: 0.03988072555629161
Validation loss per 100 evaluation steps: 0.04008753010539004
Validation loss per 100 evaluation steps: 0.04013003045455397
Validation loss per 100 evaluation steps: 0.04041955199841136
Validation loss per 100 evaluation steps: 0.040851718082625874
Validation loss per 100 evaluation steps: 0.04045932094141217
Validation loss per 100 evaluation steps: 0.04053742970171306
Validation loss per 100 evaluation steps: 0.04079750126074915
Validation loss per 100 evaluation steps: 0.04122968200394391
Validation loss per 100 evaluation steps: 0.041176033809655205
Validation loss per 100 evaluation steps: 0.04124763727658489
Validation loss per 100 evaluation steps: 0.04105150926308653
Validation loss per 100 evaluation steps: 0.04096398565773001
Validation loss per 100 evaluation steps: 0.04099191520708997
Validation loss per 100 evaluation steps: 0.04089129077321023
Validation loss per 100 evaluation steps: 0.04111275578853824
Validation loss per 100 evaluation steps: 0.04073456410231747
Validation loss per 100 evaluation steps: 0.040875135186535856
Validation loss per 100 evaluation steps: 0.0408328090581774
Validation loss per 100 evaluation steps: 0.0409207701210491
Validation loss per 100 evaluation steps: 0.04070304100943472
Validation loss per 100 evaluation steps: 0.040573293164159245
Validation loss per 100 evaluation steps: 0.040438216041011756
Validation loss per 100 evaluation steps: 0.04060451989095478
Validation loss per 100 evaluation steps: 0.04073870453609501
Validation loss per 100 evaluation steps: 0.04062118927234339
Validation loss per 100 evaluation steps: 0.04047435946255556
Validation loss per 100 evaluation steps: 0.040452917525283835
Validation Loss: 0.040397198969377980
Validation Accuracy: 0.9652426764465378
```

[37]:
```python
# print(classification_report([labels], [predictions]))
print(Validation_Results)
print(Training_Results)
```

```
[{'Validation loss epoch': 0.024871907058671323, 'Validation accuracy epoch':
0.9633732638163968, 'Epoch': 1}, {'Validation loss epoch': 0.02471111085053838,
'Validation accuracy epoch': 0.9617150367949096, 'Epoch': 2}, {'Validation loss
epoch': 0.023469813949617595, 'Validation accuracy epoch': 0.966020019121117,
'Epoch': 3}, {'Validation loss epoch': 0.02484043785683839, 'Validation accuracy
epoch': 0.9665633010961615, 'Epoch': 4}, {'Validation loss epoch':
0.027842749859228043, 'Validation accuracy epoch': 0.9661702786630613, 'Epoch':
5}, {'Validation loss epoch': 0.030817296409346766, 'Validation accuracy epoch':
0.9659824546413904, 'Epoch': 6}, {'Validation loss epoch': 0.033756910293431405,
'Validation accuracy epoch': 0.9651245958975792, 'Epoch': 7}, {'Validation loss
epoch': 0.03732937927690563, 'Validation accuracy epoch': 0.9648887493628124,
```

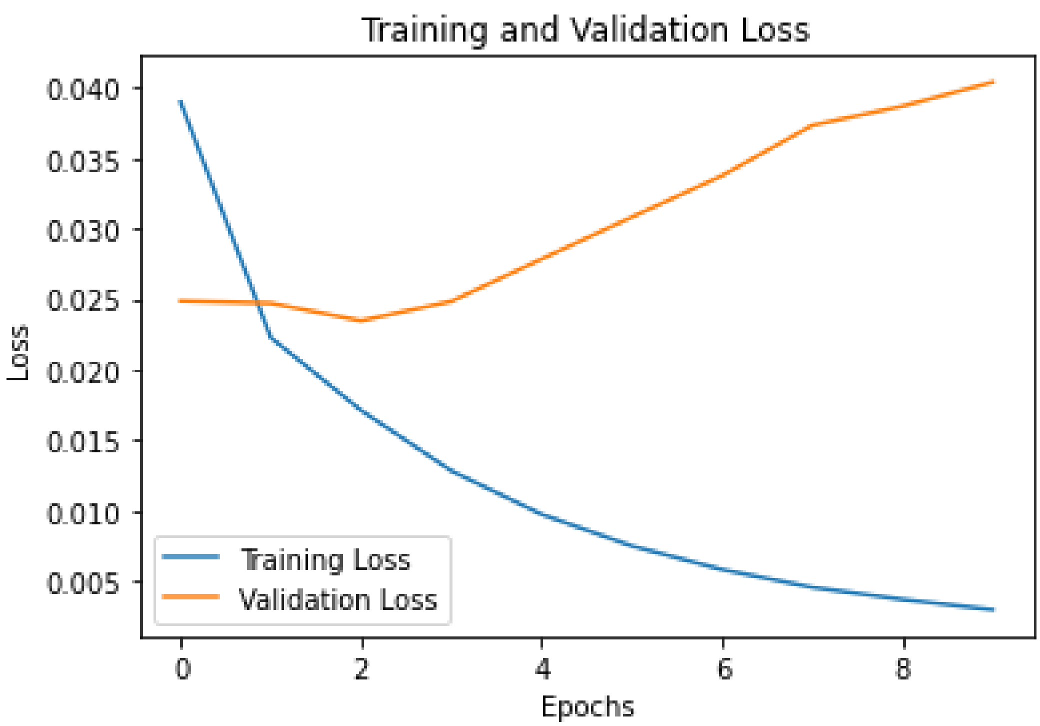
```
'Epoch': 8}, {'Validation loss epoch': 0.0386871123200506, 'Validation accuracy
epoch': 0.9649197984753808, 'Epoch': 9}, {'Validation loss epoch':
0.04039719896937798, 'Validation accuracy epoch': 0.9652426764465378, 'Epoch':
10}]
[{'Training loss': 0.03897308698877425, 'Training accuracy': 0.9495700795696844,
'Epoch': 1}, {'Training loss': 0.02227375777897628, 'Training accuracy':
0.9657434621740814, 'Epoch': 2}, {'Training loss': 0.01711671398995961,
'Training accuracy': 0.9725447265544551, 'Epoch': 3}, {'Training loss':
0.012822100012060314, 'Training accuracy': 0.9790397227316743, 'Epoch': 4},
{'Training loss': 0.00973960443761075, 'Training accuracy': 0.9839517000554161,
'Epoch': 5}, {'Training loss': 0.007495418797864272, 'Training accuracy':
0.9877923093563183, 'Epoch': 6}, {'Training loss': 0.005820838511166344 5,
'Training accuracy': 0.9903924454262886, 'Epoch': 7}, {'Training loss':
0.004558326844121083, 'Training accuracy': 0.992390585231023, 'Epoch': 8},
{'Training loss': 0.003699557676310946, 'Training accuracy': 0.994024083289638,
'Epoch': 9}, {'Training loss': 0.0029728769755038745, 'Training accuracy':
0.9951372686874074, 'Epoch': 10}]
```

### 2.4 Evaluation Phase II

[38]:
```python
# Imports for functionality
import pandas as pd
from matplotlib.pylab import plt
from numpy import arange

Training = pd.DataFrame.from_dict(Training_Results)
Validation = pd.DataFrame.from_dict(Validation_Results)

Training.to_csv('Training_final.csv', index=True)
Validation.to_csv('Validation_final.csv', index=True)
```

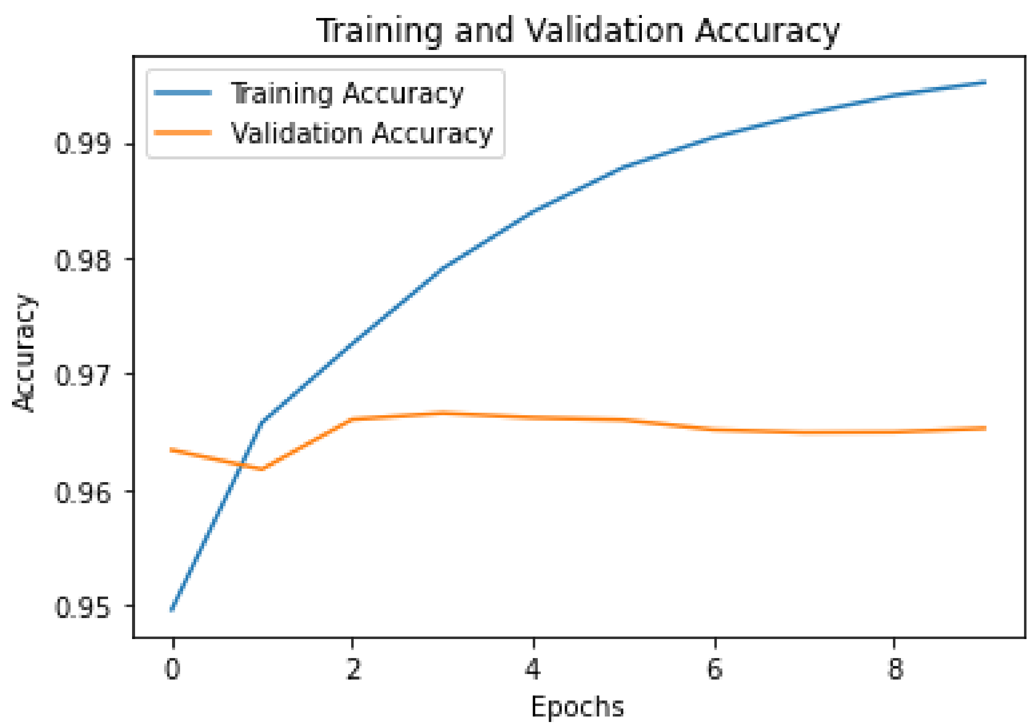
[39]:
```python
# resort into 4 arrays of plottable data
trn_Loss = []
val_Loss = []

trn_Accr = []
val_Accr = []

Training_Results = pd.read_csv('Training_final.csv')
Validation_Results = pd.read_csv('Validation_final.csv')



for index, item in enumerate(Training_Results["Training loss"]):

  trn_Loss.append(item)
  val_Loss.append(Validation_Results["Validation loss epoch"][index])
```

```
  trn_Accr.append(Training_Results["Training accuracy"][index])
  val_Accr.append(Validation_Results["Validation accuracy epoch"][index])



# Attempt to plot data that's present

epochs = range(0,10)
# Plot and label the training and validation loss values
plt.plot(epochs, trn_Loss, label='Training Loss')
plt.plot(epochs, val_Loss, label='Validation Loss')

# Add in a title and axes labels
plt.title('Training and Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')

# Set the tick locations
# plt.xticks(range(1, 6, 1))

# Display the plot
plt.legend(loc='best')
plt.show()
```



[40]:

```
# Plot and label the training and validation loss values
plt.plot(epochs, trn_Accr, label='Training Accuracy')
plt.plot(epochs, val_Accr, label='Validation Accuracy')

# Add in a title and axes labels
plt.title('Training and Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')

# Set the tick locations
# plt.xticks(arange(1, 6, 1))

# Display the plot
plt.legend(loc='best')
plt.show()
```



## 2.5 Inference

### 2.5.1 try random sentence

[41]:
```
sentence = "State of Florida has a city called Orlando. On a Monday January␣
↪2nd, the Disney employees will give a speech"

inputs = tokenizer(sentence, padding='max_length', truncation=True,␣
↪max_length=MAX_LEN, return_tensors="pt")

# move to gpu
ids = inputs["input_ids"].to(device)
mask = inputs["attention_mask"].to(device)
# forward pass
outputs = model(ids, mask)
logits = outputs[0]

active_logits = logits.view(-1, model.num_labels) # shape (batch_size *␣
↪seq_len, num_labels)
flattened_predictions = torch.argmax(active_logits, axis=1) # shape␣
↪(batch_size*seq_len,) - predictions at the token level

tokens = tokenizer.convert_ids_to_tokens(ids.squeeze().tolist())
token_predictions = [id2label[i] for i in flattened_predictions.cpu().numpy()]
wp_preds = list(zip(tokens, token_predictions)) # list of tuples. Each tuple =␣
↪(wordpiece, prediction)

word_level_predictions = []
for pair in wp_preds:
  if (pair[0].startswith(" ##")) or (pair[0] in ['[CLS]', '[SEP]', '[PAD]']):
    # skip prediction
    continue
  else:
    word_level_predictions.append(pair[1])

# we join tokens, if they are not special ones
str_rep = " ".join([t[0] for t in wp_preds if t[0] not in ['[CLS]', '[SEP]',␣
↪'[PAD]']]).replace(" ##", "")
print(str_rep)
print(word_level_predictions)
```

```
state of florida has a city called orlando . on a monday january 2nd , the
disney employees will give a speech
['O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'B-tim',
'I-tim', 'I-tim', 'O', 'O', 'B-org', 'O', 'O', 'O', 'O', 'O']
```

### 2.5.2 Make a pipline

[42]:
```python
from transformers import pipeline

pipe = pipeline(task="token-classification", model=model.to("cpu"),␣
 ↪tokenizer=tokenizer, aggregation_strategy="simple")
pipe("My name is Sahara and Orlando is a city")
```

[42]:
```
[{'entity_group': 'org',
  'score': 0.6873073,
  'word': 'sahara',
  'start': None,
  'end': None},
 {'entity_group': 'geo',
  'score': 0.99400985,
  'word': 'orlando',
  'start': None,
  'end': None}]
```

## 2.6 Results and Discussion

Our approach to training this model stared with performing an analysis on our dataset to identify relevant distributions and/or data features that could potentially affect our model. Our EDA revealed that the dataset did contain a number of features we were looking for and that would be useful for our chosen purpose of utilizing this model for identifying relevant information from news articles. The data was split up using a standard 80% Training 20% validation split. Due to the requirement for BERT needing an equal length input, we initialized a method to trim up or down based on the decided max length. Our initial parameters included a Max Length of 128, a Training Batch Size of 4, a Validation batch size of 2, on one Epoch and with a learning rate of 1*10^-5. While the results gathered from this initial run were sufficient to support the observation that BERT can operate relatively well with few learning epochs we deiced to tune this value as much as our hardware would allow to find the most optimal performance level.

After consecutive Training, we were able reduce the training loss to 0.7% and increase the training accuracy to 98%. While this is definitely the best we've gotten for these values with how many epochs we ran, a graph of our Training vs Validation loss would likely suggest overfitting at anything above 4 epochs. (This accuracy was obtained at Epoch 6) This leads us to believe that the actual optimal method for training of this model hovers roughly around 2 Epochs as that's the inflection point where we get the most amount of accuracy without the overfitting that can lead to further losses in training. It's important to note that optimization of the training at around 2 epochs is much more preferred as you can obtain a high level of performance at this level with minimal resource utilization. You can technically train this model up to 4 epochs to get a large increase in accuracy with minimal hit to Validation loss and accuracy. The resources required to do this however may make it cost prohibitive. We were able to test our model with inferencing on identifying locations time people and organizations and it was found that it was able to perform these tasks with relative success, when trained with one epoch. Due to limitations in hardware we were unable to train past 6 epochs, however given the models performance on one, and optimization to 2, we can infer that it would only perform moderately better with 1 level of additional training, after which the model should be sufficient for our use in identifying features in news articles and large information sets.