###Problem Statement: Task 1: Generate datasets of functions with their Taylor expansions up the fourth order. Tokenize the dataset.

Task 2: Train an LSTM model to learn the Taylor expansion of each function.

Task 3: Similarly Train a Transformer model to learn the Taylor expansion of each function.

```
import re
import random
import math
import numpy as np
import torch
from sympy import *
from tqdm import tqdm
import tokenize
from io import StringIO
from torch import nn
from torch.autograd import Variable
import pandas as pd
import torch.nn.functional as F

class MathExpression:
    operations = {
        'sin': 1, 'cos': 1, 'tan': 1, 'square': 1, 'cube': 1, 'exp':
1, 'log': 1,
        '+': 2, '-': 2, '*': 2, '/': 2, '**': 2
    }

    infix_notation = {
        'sin': lambda a: f'sin({a[0]})',
        'cos': lambda a: f'cos({a[0]})',
        'tan': lambda a: f'tan({a[0]})',
        'square': lambda a: f'({a[0]})**2',
        'cube': lambda a: f'({a[0]})**3',
        'exp': lambda a: f'exp({a[0]})',
        'log': lambda a: f'log({a[0]})',
        '+': lambda a: f'({a[0]})+({a[1]})',
        '-': lambda a: f'({a[0]})-({a[1]})',
        '*': lambda a: f'({a[0]})*({a[1]})',
        '/': lambda a: f'({a[0]})/({a[1]})',
        '**': lambda a: f'({a[0]})**({a[1]})'
    }

    unary_probabilities = {'sin': 1, 'cos': 1, 'tan': 2, 'square': 4,
'cube': 3, 'exp': 2, 'log': 1}
    binary_probabilities = {'+': 3, '-': 3, '*': 2, '/': 2, '**': 1}
    sympy_to_ops = {sin: 'sin', cos: 'cos', tan: 'tan', exp: 'exp',
log: 'log', Add: '+', Mul: '*', Pow: '**'}

    def unary_binary_dist(self, size):
```

```
        max_e = size * 2
        D = np.zeros((max_e + 2, size + 1))
        D[:, 0] = (self._num_leaves ** np.arange(max_e + 2))
        D[0, 0] = 0

        for n in range(1, size + 1):
            for e in range(1, max_e + 1):
                D[e, n] = (self._num_leaves * D[e-1, n] +
                           self._num_unary_ops * D[e, n-1] +
                           self._num_bin_ops * D[e+1, n-1])
        return D[:max_e + 1, :size + 1]

    def sample(self, e, n):
        P = np.zeros((e, 2))
        k_vals = np.arange(e)
        P[:, 0] = (self._num_leaves ** k_vals) * self._num_unary_ops *
self._unary_binary_dist[e - k_vals, n-1]
        P[:, 1] = (self._num_leaves ** k_vals) * self._num_bin_ops *
self._unary_binary_dist[e - k_vals + 1, n-1]
        P_flat = P.T.flatten()
        P_flat /= P_flat.sum()
        k = np.random.choice(2 * e, p=P_flat)
        return k % e, 1 if k < e else 2

    def choose_unary_op(self):
        ops, probs = zip(*self.unary_probabilities.items())
        probs = np.array(probs) / sum(probs)
        return np.random.choice(ops, p=probs)

    def choose_bin_op(self):
        ops, probs = zip(*self.binary_probabilities.items())
        probs = np.array(probs) / sum(probs)
        return np.random.choice(ops, p=probs)

    def choose_leaf(self):
        return 'x' if random.random() < 0.3 else random.randint(0, 9)

    def gen_from_sympy(self, expr):
        self._rep = []
        stack = [expr]
        while stack:
            curr = stack.pop()
            if isinstance(curr, (Symbol, Integer)):
                self._rep.append(str(curr))
            elif isinstance(curr, Rational):
                self._rep.extend(['/', str(curr.p), str(curr.q)])
            elif curr in [E, pi, I]:
                self._rep.append(str(curr).lower())
            else:
                op = self.sympy_to_ops.get(type(curr), None)
```

```
                if op:
                    args = curr.args
                    self._rep.extend([op] * (len(args) - 1))
                    stack.extend(reversed(args))

    def gen_random(self, num_ops):
        self._num_leaves = 1
        self._num_bin_ops = len(self.binary_probabilities)
        self._num_unary_ops = len(self.unary_probabilities)
        self._unary_binary_dist = self.unary_binary_dist(num_ops + 1)

        rep = [None]
        e = 1
        skipped_total = 0
        for n in range(num_ops, 0, -1):
            k, arity = self.sample(e, n)
            skipped_total += k
            none_indices = [i for i, x in enumerate(rep) if x is None]
            pos = none_indices[skipped_total]
            op = self.choose_unary_op() if arity == 1 else
self.choose_bin_op()
            new_elems = [op] + [None] * arity
            rep[pos:pos+1] = new_elems
            e = e - k + (0 if arity == 1 else 1)
            skipped_total = 0

        for i in range(len(rep)):
            if rep[i] is None:
                rep[i] = self.choose_leaf()
        self._rep = rep

    def __init__(self, expr=None, num_ops=None):
        if expr is not None:
            self.gen_from_sympy(expr)
        else:
            self.gen_random(num_ops)

    def to_infix(self):
        stack = []
        for token in reversed(self._rep):
            if token in self.operations:
                args = [stack.pop() for _ in
range(self.operations[token])]
                stack.append(self.infix_notation[token](args[::-1]))
            else:
                stack.append(str(token))
        return stack[0]

    def get_rep(self):
        return self._rep
```

```
def taylor_series(f_str, a, order):
    x = symbols('x')
    f = parse_expr(f_str)
    taylor = f.subs(x, a)
    for i in range(1, order + 1):
        f = diff(f, x)
        term = (f * (x - a)**i) / factorial(i)
        taylor += term
    return taylor.expand()

def gen_pair(ops=3):
    expr = MathExpression(num_ops=ops)
    tay = taylor_series(expr.to_infix(), Symbol('a'), 4)
    return expr, MathExpression(expr=tay)

class FunctionDataset(torch.utils.data.Dataset):
    def __init__(self, ops=3, max_seq_length=32, num_items=100):
        self.max_seq = max_seq_length
        self.data = []
        while len(self.data) < num_items:
            src, tgt = gen_pair(ops)
            src_rep = ['<SOS>'] + src.get_rep() + ['<EOS>']
            tgt_rep = ['<SOS>'] + tgt.get_rep() + ['<EOS>']
            if len(src_rep) <= max_seq_length and len(tgt_rep) <=
max_seq_length:
                self.data.append((src_rep, tgt_rep))

        all_tokens = set(token for pair in self.data for expr in pair
for token in expr)
        self.vocab = {'<PAD>': 0}
        self.vocab.update({token: i+1 for i, token in
enumerate(all_tokens)})

        self.inputs = []
        self.targets = []
        for src, tgt in self.data:
            src_ids = [self.vocab.get(t, 0) for t in src] + [0] *
(max_seq_length - len(src))
            tgt_ids = [self.vocab.get(t, 0) for t in tgt] + [0] *
(max_seq_length - len(tgt))
            self.inputs.append(torch.LongTensor(src_ids))
            self.targets.append(torch.LongTensor(tgt_ids))

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

    def __getitem__(self, idx):
        return self.inputs[idx], self.targets[idx]

    def get_vocab(self):
```

```
        return self.vocab
    def get_alphabet(self):
        return self.vocab

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
d = FunctionDataset(num_items=500)
train_size = int(0.9 * len(d))
train_dataset, test_dataset = torch.utils.data.random_split(d,
[train_size, len(d)-train_size])

class Encoder(nn.Module):
  def __init__(self, vocab_size, embedding_dim=512, num_layers=2,
hidden_size=512, dropout=0.2):
    super(Encoder, self).__init__()
    self.embedding_dim = embedding_dim
    self.num_layers = num_layers
    self.hidden_size = hidden_size

    self.embedding = nn.Embedding(
        num_embeddings=vocab_size,
        embedding_dim=self.embedding_dim
    )
    self.lstm = nn.LSTM(
        input_size=self.embedding_dim,
        hidden_size=self.hidden_size,
        num_layers=self.num_layers,
        dropout=dropout,
    )

  def forward(self, x):
    embed = self.embedding(x)
    output, (h,c) = self.lstm(embed)
    return h, c

class Decoder(nn.Module):
  def __init__(self, vocab_size, embedding_dim=512, num_layers=2,
hidden_size=512, dropout=0.2):
    super(Decoder, self).__init__()
    self.embedding_dim = embedding_dim
    self.num_layers = num_layers
    self.output_size = vocab_size
    self.hidden_size = hidden_size

    self.embedding = nn.Embedding(
        num_embeddings=vocab_size,
        embedding_dim=self.embedding_dim
    )
    self.lstm = nn.LSTM(
        input_size=self.embedding_dim,
        hidden_size=self.hidden_size,
```

```
        num_layers=self.num_layers,
        dropout=0.2,
    )
    self.out = nn.Linear(self.hidden_size, self.output_size)
    self.softmax = nn.LogSoftmax(dim=2)
    self.to(device)

  def forward(self, input, h_0, c_0):
    embedded = self.embedding(input.unsqueeze(0))
    output, (h,c) = self.lstm(embedded, (h_0, c_0))
    output = self.out(output)
    output = self.softmax(output)
    return output.squeeze(0), h , c

class Model(nn.Module):
  def __init__(self, encoder, decoder):
    super(Model, self).__init__()
    self.encoder = encoder
    self.decoder = decoder
    self.to(device)

  def forward(self, input, tgt=None):
    if len(input.shape) < 2:
        input = input.unsqueeze(1)
    batch_size = input.shape[1]
    h, c = enc(input)
    target = torch.zeros(batch_size, dtype=torch.long).to(device)
    if tgt is None:
      max_seq_length = input.shape[0]
      target[:] = d.token_to_idx['<SOS>']
    else:
      max_seq_length = tgt.shape[1]
      target[:] = tgt[:,0]
    outputs = torch.zeros(max_seq_length, batch_size, dec.output_size,
dtype=torch.float).to(device)
    for i in range(max_seq_length):
        prediction, h, c = dec(target, h, c)
        outputs[i] = prediction
        if tgt is None:
          target = prediction.argmax(dim=1)
        else:
          target = tgt[:,i]
    return outputs

enc = Encoder(len(d.get_alphabet()) + 1)
dec = Decoder(len(d.get_alphabet()) + 1)
m = Model(enc,dec).to(device)

def test_epoch_LSTM(model, test_loader, criterion, batch_size=4):
  model.eval()
```

```
  total_loss = 0
  total_items = 0
  num_correct = 0
  for src, tgt in tqdm(test_loader):
    src = src.to(device)
    tgt = tgt.to(device)

    pred = model(src.squeeze().T, tgt=tgt[:,:-1])
    pred = pred.permute((1,2,0))
    tgt_out = tgt[:,1:]

    loss = criterion(pred, tgt_out)

    total_loss += loss.item()
    total_items += (tgt_out != 0).sum(dim=(0,1))

    num_correct += (torch.logical_and((logits.argmax(dim=2) ==
tgt_out), (tgt_out != 0))).sum(dim=(0,1))
  return total_loss, num_correct / total_items

def train_epoch_LSTM(model, train_loader, optimizer, criterion,
batch_size=4):
  model.train()
  total_loss = 0
  total_items = 0
  num_correct = 0
  for src, tgt in tqdm(train_loader):
    src = src.to(device)
    tgt = tgt.to(device)

    pred = model(src.squeeze().T,tgt=tgt[:,:-1])

    pred = pred.permute((1,2,0))
    tgt_out = tgt[:,1:]
    loss = criterion(pred, tgt_out)

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

    total_loss += loss.item()
    total_items += (tgt_out != 0).sum(dim=(0,1))

    num_correct += (torch.logical_and((pred.argmax(dim=1) == tgt_out),
(tgt_out != 0))).sum(dim=(0,1))
  return total_loss, num_correct / total_items

def train_LSTM(model, train_dataset,  test_dataset, batch_size=32,
epochs=100):
  train_loader = torch.utils.data.DataLoader(train_dataset,
```

```
batch_size=batch_size, shuffle=True)
  test_loader = torch.utils.data.DataLoader(test_dataset,
batch_size=batch_size, shuffle=True)
  criterion = nn.CrossEntropyLoss()


  optim = torch.optim.Adam(model.parameters(), lr=1e-3)
  for e in range(epochs):
    train_loss, train_acc = train_epoch_LSTM(model, train_loader,
optim, criterion, batch_size=batch_size)
    test_loss, test_acc = train_epoch_LSTM(model, test_loader, optim,
criterion, batch_size=batch_size)
    print(f'Epoch: {e + 1} Training Loss: {train_loss} Training
Accuracy: {train_acc} Test Loss: {test_loss} Test Accuracy:
{test_acc}')

train_LSTM(m, train_dataset, test_dataset, batch_size=32)

100%|██████████| 15/15 [00:01<00:00, 12.48it/s]
100%|██████████| 2/2 [00:00<00:00, 16.12it/s]

Epoch: 1 Training Loss: 23.17210751771927 Training Accuracy:
0.05978068709373474 Test Loss: 1.4430557489395142 Test Accuracy:
0.0917721539735794

100%|██████████| 15/15 [00:00<00:00, 16.87it/s]
100%|██████████| 2/2 [00:00<00:00, 17.24it/s]

Epoch: 2 Training Loss: 9.684578090906143 Training Accuracy:
0.16413158178329468 Test Loss: 1.2323780059814453 Test Accuracy:
0.2278480976819992

100%|██████████| 15/15 [00:00<00:00, 16.94it/s]
100%|██████████| 2/2 [00:00<00:00, 17.55it/s]

Epoch: 3 Training Loss: 8.879249095916748 Training Accuracy:
0.33569154143333435 Test Loss: 1.2321085631847382 Test Accuracy:
0.3639240562915802

100%|██████████| 15/15 [00:00<00:00, 16.85it/s]
100%|██████████| 2/2 [00:00<00:00, 17.25it/s]

Epoch: 4 Training Loss: 7.683849483728409 Training Accuracy:
0.4163424074649811 Test Loss: 1.0423645377159119 Test Accuracy:
0.42721518874168396

100%|██████████| 15/15 [00:00<00:00, 16.77it/s]
100%|██████████| 2/2 [00:00<00:00, 16.90it/s]

Epoch: 5 Training Loss: 7.06634846329689 Training Accuracy:
0.46338874101638794 Test Loss: 1.0176019966602325 Test Accuracy:
0.43670886754989624
```

```
100%|██████████| 15/15 [00:00<00:00, 16.94it/s]
100%|██████████| 2/2 [00:00<00:00, 17.23it/s]

Epoch: 6 Training Loss: 6.639508962631226 Training Accuracy:
0.485320121049881 Test Loss: 0.9513146281242371 Test Accuracy:
0.452531635761261

100%|██████████| 15/15 [00:01<00:00, 11.90it/s]
100%|██████████| 2/2 [00:00<00:00, 11.79it/s]

Epoch: 7 Training Loss: 6.3657273054122925 Training Accuracy:
0.5132649540901184 Test Loss: 0.8951953053474426 Test Accuracy:
0.474683552980423

100%|██████████| 15/15 [00:01<00:00, 13.09it/s]
100%|██████████| 2/2 [00:00<00:00, 17.91it/s]

Epoch: 8 Training Loss: 6.160875976085663 Training Accuracy:
0.5274142026901245 Test Loss: 0.838809609413147 Test Accuracy:
0.4905063211917877

100%|██████████| 15/15 [00:00<00:00, 16.86it/s]
100%|██████████| 2/2 [00:00<00:00, 17.85it/s]

Epoch: 9 Training Loss: 6.166071176528931 Training Accuracy:
0.5447470545768738 Test Loss: 0.8958266079425812 Test Accuracy: 0.5

100%|██████████| 15/15 [00:00<00:00, 16.44it/s]
100%|██████████| 2/2 [00:00<00:00, 16.61it/s]

Epoch: 10 Training Loss: 6.4832653403282166 Training Accuracy:
0.5458082556724548 Test Loss: 0.8859668672084808 Test Accuracy:
0.503164529800415

100%|██████████| 15/15 [00:00<00:00, 17.06it/s]
100%|██████████| 2/2 [00:00<00:00, 16.96it/s]

Epoch: 11 Training Loss: 5.892133116722107 Training Accuracy:
0.5461620092391968 Test Loss: 0.8361568748950958 Test Accuracy:
0.5158227682113647

100%|██████████| 15/15 [00:00<00:00, 17.19it/s]
100%|██████████| 2/2 [00:00<00:00, 16.29it/s]

Epoch: 12 Training Loss: 5.592693477869034 Training Accuracy:
0.5642023086547852 Test Loss: 0.7784326374530792 Test Accuracy:
0.5348101258277893

100%|██████████| 15/15 [00:00<00:00, 17.01it/s]
100%|██████████| 2/2 [00:00<00:00, 16.08it/s]
```

```
Epoch: 13 Training Loss: 5.547985315322876 Training Accuracy:
0.5794128179550171 Test Loss: 0.8149541914463043 Test Accuracy:
0.5316455960273743

100%|██████████| 15/15 [00:00<00:00, 16.81it/s]
100%|██████████| 2/2 [00:00<00:00, 16.28it/s]

Epoch: 14 Training Loss: 5.271307989954948 Training Accuracy:
0.5953307151794434 Test Loss: 0.7655770480632782 Test Accuracy:
0.5537974834442139

100%|██████████| 15/15 [00:00<00:00, 16.84it/s]
100%|██████████| 2/2 [00:00<00:00, 16.37it/s]

Epoch: 15 Training Loss: 5.0963268876075745 Training Accuracy:
0.6084188222885132 Test Loss: 0.7226317822933197 Test Accuracy:
0.5664557218551636

100%|██████████| 15/15 [00:00<00:00, 16.80it/s]
100%|██████████| 2/2 [00:00<00:00, 16.09it/s]

Epoch: 16 Training Loss: 5.02226784825325 Training Accuracy:
0.6059426665306091 Test Loss: 0.6910622715950012 Test Accuracy:
0.5696202516555786

100%|██████████| 15/15 [00:00<00:00, 16.65it/s]
100%|██████████| 2/2 [00:00<00:00, 17.27it/s]

Epoch: 17 Training Loss: 5.035017758607864 Training Accuracy:
0.620445728302002 Test Loss: 0.6696019172668457 Test Accuracy:
0.5917721390724182

100%|██████████| 15/15 [00:01<00:00, 12.89it/s]
100%|██████████| 2/2 [00:00<00:00, 11.69it/s]

Epoch: 18 Training Loss: 4.69368813931942 Training Accuracy:
0.6292889714241028 Test Loss: 0.6522445976734161 Test Accuracy:
0.5917721390724182

100%|██████████| 15/15 [00:01<00:00, 12.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.63it/s]

Epoch: 19 Training Loss: 4.521913439035416 Training Accuracy:
0.6310576796531677 Test Loss: 0.618354856967926 Test Accuracy:
0.6234177350997925

100%|██████████| 15/15 [00:00<00:00, 16.97it/s]
100%|██████████| 2/2 [00:00<00:00, 17.39it/s]

Epoch: 20 Training Loss: 4.481335669755936 Training Accuracy:
0.63494873046875 Test Loss: 0.5810351967811584 Test Accuracy:
0.6265822649002075
```

```
100%|██████████| 15/15 [00:00<00:00, 16.70it/s]
100%|██████████| 2/2 [00:00<00:00, 17.40it/s]

Epoch: 21 Training Loss: 4.2595807164907455 Training Accuracy:
0.6391934752464294 Test Loss: 0.5705646872520447 Test Accuracy:
0.6455696225166321

100%|██████████| 15/15 [00:00<00:00, 16.88it/s]
100%|██████████| 2/2 [00:00<00:00, 17.19it/s]

Epoch: 22 Training Loss: 4.010589562356472 Training Accuracy:
0.6646621823310852 Test Loss: 0.5711603611707687 Test Accuracy:
0.655063271522522

100%|██████████| 15/15 [00:00<00:00, 16.89it/s]
100%|██████████| 2/2 [00:00<00:00, 16.88it/s]

Epoch: 23 Training Loss: 3.841157577931881 Training Accuracy:
0.6706756353378296 Test Loss: 0.5332511812448502 Test Accuracy:
0.6613923907279968

100%|██████████| 15/15 [00:00<00:00, 17.07it/s]
100%|██████████| 2/2 [00:00<00:00, 17.26it/s]

Epoch: 24 Training Loss: 3.9121848791837692 Training Accuracy:
0.6816413402557373 Test Loss: 0.5130157172679901 Test Accuracy:
0.6867088675498962

100%|██████████| 15/15 [00:00<00:00, 17.00it/s]
100%|██████████| 2/2 [00:00<00:00, 17.20it/s]

Epoch: 25 Training Loss: 3.7443594485521317 Training Accuracy:
0.6823487877845764 Test Loss: 0.4985673278570175 Test Accuracy:
0.702531635761261

100%|██████████| 15/15 [00:00<00:00, 15.55it/s]
100%|██████████| 2/2 [00:00<00:00, 17.40it/s]

Epoch: 26 Training Loss: 3.6783126443624496 Training Accuracy:
0.6972055435180664 Test Loss: 0.4647696018218994 Test Accuracy:
0.7120253443717957

100%|██████████| 15/15 [00:00<00:00, 16.23it/s]
100%|██████████| 2/2 [00:00<00:00, 16.55it/s]

Epoch: 27 Training Loss: 3.554968938231468 Training Accuracy:
0.7014502882957458 Test Loss: 0.45056046545505524 Test Accuracy:
0.7341772317886353

100%|██████████| 15/15 [00:00<00:00, 16.93it/s]
100%|██████████| 2/2 [00:00<00:00, 16.08it/s]
```

```
Epoch: 28 Training Loss: 3.3241299986839294 Training Accuracy:
0.7251503467559814 Test Loss: 0.4377773851156235 Test Accuracy:
0.7436708807945251

100%|██████████| 15/15 [00:01<00:00, 13.69it/s]
100%|██████████| 2/2 [00:00<00:00, 11.36it/s]

Epoch: 29 Training Loss: 3.250397652387619 Training Accuracy:
0.7301025986671448 Test Loss: 0.41207368671894073 Test Accuracy: 0.75

100%|██████████| 15/15 [00:01<00:00, 11.73it/s]
100%|██████████| 2/2 [00:00<00:00, 17.06it/s]

Epoch: 30 Training Loss: 3.004386469721794 Training Accuracy:
0.7407145500183105 Test Loss: 0.41673216223716736 Test Accuracy: 0.75

100%|██████████| 15/15 [00:00<00:00, 16.93it/s]
100%|██████████| 2/2 [00:00<00:00, 17.44it/s]

Epoch: 31 Training Loss: 3.0363082587718964 Training Accuracy:
0.7453130483627319 Test Loss: 0.37130168080329895 Test Accuracy:
0.7721518874168396

100%|██████████| 15/15 [00:00<00:00, 16.71it/s]
100%|██████████| 2/2 [00:00<00:00, 16.86it/s]

Epoch: 32 Training Loss: 3.057570904493332 Training Accuracy:
0.7484966516494751 Test Loss: 0.39317460358142853 Test Accuracy:
0.7784810066223145

100%|██████████| 15/15 [00:00<00:00, 16.65it/s]
100%|██████████| 2/2 [00:00<00:00, 17.14it/s]

Epoch: 33 Training Loss: 2.836838461458683 Training Accuracy:
0.7484966516494751 Test Loss: 0.3676316738128662 Test Accuracy:
0.7753164768218994

100%|██████████| 15/15 [00:00<00:00, 16.74it/s]
100%|██████████| 2/2 [00:00<00:00, 16.89it/s]

Epoch: 34 Training Loss: 2.791977971792221 Training Accuracy:
0.7771489024162292 Test Loss: 0.34708505868911743 Test Accuracy:
0.7848101258277893

100%|██████████| 15/15 [00:00<00:00, 16.89it/s]
100%|██████████| 2/2 [00:00<00:00, 17.12it/s]

Epoch: 35 Training Loss: 2.7773669362068176 Training Accuracy:
0.7683055996894836 Test Loss: 0.357919842004776 Test Accuracy:
0.7816455960273743

100%|██████████| 15/15 [00:00<00:00, 16.74it/s]
100%|██████████| 2/2 [00:00<00:00, 17.23it/s]
```

```
Epoch: 36 Training Loss: 2.5411408245563507 Training Accuracy:
0.7778564095497131 Test Loss: 0.30872194468975067 Test Accuracy:
0.8164557218551636

100%|██████████| 15/15 [00:00<00:00, 16.62it/s]
100%|██████████| 2/2 [00:00<00:00, 17.32it/s]

Epoch: 37 Training Loss: 2.3058657124638557 Training Accuracy:
0.7955429553985596 Test Loss: 0.2971993237733841 Test Accuracy:
0.8227847814559937

100%|██████████| 15/15 [00:00<00:00, 16.07it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 38 Training Loss: 2.125586934387684 Training Accuracy:
0.8043862581253052 Test Loss: 0.24483418464660645 Test Accuracy:
0.844936728477478

100%|██████████| 15/15 [00:00<00:00, 16.12it/s]
100%|██████████| 2/2 [00:00<00:00, 16.98it/s]

Epoch: 39 Training Loss: 1.966255471110344 Training Accuracy:
0.8287937641143799 Test Loss: 0.22822506725788116 Test Accuracy:
0.8670886158943176

100%|██████████| 15/15 [00:01<00:00, 13.80it/s]
100%|██████████| 2/2 [00:00<00:00, 11.66it/s]

Epoch: 40 Training Loss: 1.9498735293745995 Training Accuracy:
0.8305624127388 Test Loss: 0.23592694848775864 Test Accuracy:
0.8639240264892578

100%|██████████| 15/15 [00:01<00:00, 11.44it/s]
100%|██████████| 2/2 [00:00<00:00, 17.39it/s]

Epoch: 41 Training Loss: 1.779320240020752 Training Accuracy:
0.8447117209434509 Test Loss: 0.2272382453083992 Test Accuracy:
0.8607594966888428

100%|██████████| 15/15 [00:00<00:00, 16.45it/s]
100%|██████████| 2/2 [00:00<00:00, 17.06it/s]

Epoch: 42 Training Loss: 1.7593305632472038 Training Accuracy:
0.8475415706634521 Test Loss: 0.20954638719558716 Test Accuracy:
0.8639240264892578

100%|██████████| 15/15 [00:00<00:00, 16.85it/s]
100%|██████████| 2/2 [00:00<00:00, 17.31it/s]

Epoch: 43 Training Loss: 1.554246835410595 Training Accuracy:
0.8606296181678772 Test Loss: 0.199519582092762 Test Accuracy:
0.8892405033111572
```

```
100%|██████████| 15/15 [00:00<00:00, 16.72it/s]
100%|██████████| 2/2 [00:00<00:00, 17.36it/s]

Epoch: 44 Training Loss: 1.4904280081391335 Training Accuracy:
0.8761938214302063 Test Loss: 0.1673831269145012 Test Accuracy:
0.9303797483444214

100%|██████████| 15/15 [00:00<00:00, 16.62it/s]
100%|██████████| 2/2 [00:00<00:00, 17.24it/s]

Epoch: 45 Training Loss: 1.3490485809743404 Training Accuracy:
0.8850371241569519 Test Loss: 0.14964565634727478 Test Accuracy:
0.9335442781448364

100%|██████████| 15/15 [00:00<00:00, 16.85it/s]
100%|██████████| 2/2 [00:00<00:00, 17.35it/s]

Epoch: 46 Training Loss: 1.1748269349336624 Training Accuracy:
0.9041386842727661 Test Loss: 0.14218170940876007 Test Accuracy:
0.9177215099334717

100%|██████████| 15/15 [00:00<00:00, 16.44it/s]
100%|██████████| 2/2 [00:00<00:00, 16.55it/s]

Epoch: 47 Training Loss: 1.1496194079518318 Training Accuracy:
0.915458083152771 Test Loss: 0.11160573363304138 Test Accuracy:
0.9556962251663208

100%|██████████| 15/15 [00:00<00:00, 16.79it/s]
100%|██████████| 2/2 [00:00<00:00, 17.28it/s]

Epoch: 48 Training Loss: 1.0915939845144749 Training Accuracy:
0.9133356809616089 Test Loss: 0.11194586753845215 Test Accuracy:
0.949367105960846

100%|██████████| 15/15 [00:00<00:00, 16.57it/s]
100%|██████████| 2/2 [00:00<00:00, 17.42it/s]

Epoch: 49 Training Loss: 1.036714754998684 Training Accuracy:
0.9257162809371948 Test Loss: 0.10248417779803276 Test Accuracy:
0.9462025165557861

100%|██████████| 15/15 [00:00<00:00, 17.06it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 50 Training Loss: 0.9927495196461678 Training Accuracy:
0.9327909350395203 Test Loss: 0.11648881807923317 Test Accuracy:
0.9588607549667358

100%|██████████| 15/15 [00:01<00:00, 14.92it/s]
100%|██████████| 2/2 [00:00<00:00, 11.92it/s]
```

```
Epoch: 51 Training Loss: 0.9641300477087498 Training Accuracy:
0.9356207847595215 Test Loss: 0.11962078511714935 Test Accuracy:
0.9367088675498962

100%|██████████| 15/15 [00:01<00:00, 11.90it/s]
100%|██████████| 2/2 [00:00<00:00, 10.78it/s]

Epoch: 52 Training Loss: 0.8044560682028532 Training Accuracy:
0.9391581416130066 Test Loss: 0.08168156445026398 Test Accuracy:
0.9715189933776855

100%|██████████| 15/15 [00:00<00:00, 16.77it/s]
100%|██████████| 2/2 [00:00<00:00, 16.05it/s]

Epoch: 53 Training Loss: 0.6392602373380214 Training Accuracy:
0.9603820443153381 Test Loss: 0.05695664882659912 Test Accuracy:
0.9936708807945251

100%|██████████| 15/15 [00:00<00:00, 17.00it/s]
100%|██████████| 2/2 [00:00<00:00, 15.94it/s]

Epoch: 54 Training Loss: 0.5798002872616053 Training Accuracy:
0.9724088907241821 Test Loss: 0.04923618212342262 Test Accuracy:
0.9873417615890503

100%|██████████| 15/15 [00:00<00:00, 16.72it/s]
100%|██████████| 2/2 [00:00<00:00, 15.65it/s]

Epoch: 55 Training Loss: 0.5251035820692778 Training Accuracy:
0.9727626442909241 Test Loss: 0.052453331649303436 Test Accuracy:
0.9873417615890503

100%|██████████| 15/15 [00:00<00:00, 16.92it/s]
100%|██████████| 2/2 [00:00<00:00, 15.91it/s]

Epoch: 56 Training Loss: 0.48850971553474665 Training Accuracy:
0.9748850464820862 Test Loss: 0.05080651864409447 Test Accuracy:
0.9810126423835754

100%|██████████| 15/15 [00:00<00:00, 17.08it/s]
100%|██████████| 2/2 [00:00<00:00, 15.69it/s]

Epoch: 57 Training Loss: 0.4147764788940549 Training Accuracy:
0.9780686497688293 Test Loss: 0.037569427862763405 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:00<00:00, 16.79it/s]
100%|██████████| 2/2 [00:00<00:00, 15.69it/s]

Epoch: 58 Training Loss: 0.31612321455031633 Training Accuracy:
0.9911566972732544 Test Loss: 0.03687839396297932 Test Accuracy:
0.9841772317886353
```

```
100%|██████████| 15/15 [00:00<00:00, 16.78it/s]
100%|██████████| 2/2 [00:00<00:00, 15.72it/s]

Epoch: 59 Training Loss: 0.2773670933675021 Training Accuracy:
0.9922178983688354 Test Loss: 0.02680542878806591 Test Accuracy:
0.9936708807945251

100%|██████████| 15/15 [00:00<00:00, 17.04it/s]
100%|██████████| 2/2 [00:00<00:00, 15.70it/s]

Epoch: 60 Training Loss: 0.23375756200402975 Training Accuracy:
0.9936328530311584 Test Loss: 0.026188992895185947 Test Accuracy:
0.9936708807945251

100%|██████████| 15/15 [00:00<00:00, 16.96it/s]
100%|██████████| 2/2 [00:00<00:00, 16.12it/s]

Epoch: 61 Training Loss: 0.20890103792771697 Training Accuracy:
0.9946939945220947 Test Loss: 0.019589771516621113 Test Accuracy:
0.9936708807945251

100%|██████████| 15/15 [00:00<00:00, 16.42it/s]
100%|██████████| 2/2 [00:00<00:00, 10.03it/s]

Epoch: 62 Training Loss: 0.18080599838867784 Training Accuracy:
0.9964627027511597 Test Loss: 0.02272819634526968 Test Accuracy:
0.9936708807945251

100%|██████████| 15/15 [00:01<00:00, 11.41it/s]
100%|██████████| 2/2 [00:00<00:00, 11.33it/s]

Epoch: 63 Training Loss: 0.15298649482429028 Training Accuracy:
0.9968163967132568 Test Loss: 0.015814919490367174 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:01<00:00, 14.55it/s]
100%|██████████| 2/2 [00:00<00:00, 16.77it/s]

Epoch: 64 Training Loss: 0.14195982203818858 Training Accuracy:
0.9968163967132568 Test Loss: 0.017564075998961926 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:00<00:00, 16.84it/s]
100%|██████████| 2/2 [00:00<00:00, 16.71it/s]

Epoch: 65 Training Loss: 0.1223759698914364 Training Accuracy:
0.9978775978088379 Test Loss: 0.014767032116651535 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:00<00:00, 16.46it/s]
100%|██████████| 2/2 [00:00<00:00, 16.15it/s]
```

```
Epoch: 66 Training Loss: 0.11087285971734673 Training Accuracy:
0.9978775978088379 Test Loss: 0.011088833212852478 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.61it/s]
100%|██████████| 2/2 [00:00<00:00, 16.98it/s]

Epoch: 67 Training Loss: 0.10440783528611064 Training Accuracy:
0.998585045337677 Test Loss: 0.01289279293268919 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.61it/s]
100%|██████████| 2/2 [00:00<00:00, 17.51it/s]

Epoch: 68 Training Loss: 0.10169292031787336 Training Accuracy:
0.9975239038467407 Test Loss: 0.013379616662859917 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:00<00:00, 16.78it/s]
100%|██████████| 2/2 [00:00<00:00, 17.61it/s]

Epoch: 69 Training Loss: 0.09391306340694427 Training Accuracy:
0.9992925524711609 Test Loss: 0.011724313255399466 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.59it/s]
100%|██████████| 2/2 [00:00<00:00, 17.18it/s]

Epoch: 70 Training Loss: 0.09021854982711375 Training Accuracy:
0.9982313513755798 Test Loss: 0.010966499336063862 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00,  9.98it/s]
100%|██████████| 2/2 [00:00<00:00, 17.52it/s]

Epoch: 71 Training Loss: 0.08475941233336926 Training Accuracy:
0.998585045337677 Test Loss: 0.009227495873346925 Test Accuracy:
0.996835470199585

100%|██████████| 15/15 [00:01<00:00, 12.97it/s]
100%|██████████| 2/2 [00:00<00:00, 16.99it/s]

Epoch: 72 Training Loss: 0.07146086043212563 Training Accuracy:
0.998585045337677 Test Loss: 0.009665864752605557 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00, 12.14it/s]
100%|██████████| 2/2 [00:00<00:00, 11.13it/s]

Epoch: 73 Training Loss: 0.06613052147440612 Training Accuracy:
0.998938798904419 Test Loss: 0.007397153181955218 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00, 12.91it/s]
100%|██████████| 2/2 [00:00<00:00, 17.28it/s]

Epoch: 74 Training Loss: 0.06417574803344905 Training Accuracy:
0.9992925524711609 Test Loss: 0.006646702066063881 Test Accuracy: 1.0
```

```
100%|██████████| 15/15 [00:00<00:00, 16.81it/s]
100%|██████████| 2/2 [00:00<00:00, 16.87it/s]

Epoch: 75 Training Loss: 0.05799035407835618 Training Accuracy:
0.998585045337677 Test Loss: 0.008086699759587646 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.82it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 76 Training Loss: 0.056588648119941354 Training Accuracy:
0.998585045337677 Test Loss: 0.007517100544646382 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.63it/s]
100%|██████████| 2/2 [00:00<00:00, 16.99it/s]

Epoch: 77 Training Loss: 0.055608677212148905 Training Accuracy:
0.9992925524711609 Test Loss: 0.005807205685414374 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.71it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 78 Training Loss: 0.05313468584790826 Training Accuracy:
0.998938798904419 Test Loss: 0.006279469467699528 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.72it/s]
100%|██████████| 2/2 [00:00<00:00, 16.56it/s]

Epoch: 79 Training Loss: 0.0480210327077657 Training Accuracy:
0.9996462464332581 Test Loss: 0.006137088872492313 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.84it/s]
100%|██████████| 2/2 [00:00<00:00, 17.29it/s]

Epoch: 80 Training Loss: 0.04798255569767207 Training Accuracy:
0.9992925524711609 Test Loss: 0.004990724730305374 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.70it/s]
100%|██████████| 2/2 [00:00<00:00, 16.76it/s]

Epoch: 81 Training Loss: 0.041019800963113084 Training Accuracy: 1.0
Test Loss: 0.0064670799765735865 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.79it/s]
100%|██████████| 2/2 [00:00<00:00, 17.15it/s]

Epoch: 82 Training Loss: 0.03942034090869129 Training Accuracy: 1.0
Test Loss: 0.00395243673119694 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.41it/s]
100%|██████████| 2/2 [00:00<00:00, 16.91it/s]

Epoch: 83 Training Loss: 0.03629708173684776 Training Accuracy: 1.0
Test Loss: 0.004871221957728267 Test Accuracy: 1.0
```

```
100%|██████████| 15/15 [00:01<00:00, 12.57it/s]
100%|██████████| 2/2 [00:00<00:00, 12.46it/s]

Epoch: 84 Training Loss: 0.0363524112617597 Training Accuracy:
0.9992925524711609 Test Loss: 0.004576561506837606 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00, 12.47it/s]
100%|██████████| 2/2 [00:00<00:00, 17.50it/s]

Epoch: 85 Training Loss: 0.034564069588668644 Training Accuracy: 1.0
Test Loss: 0.004389762645587325 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.83it/s]
100%|██████████| 2/2 [00:00<00:00, 17.36it/s]

Epoch: 86 Training Loss: 0.03138655732618645 Training Accuracy:
0.9996462464332581 Test Loss: 0.0035573705099523067 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.69it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 87 Training Loss: 0.03099148510955274 Training Accuracy:
0.9996462464332581 Test Loss: 0.0037595886969938874 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.58it/s]
100%|██████████| 2/2 [00:00<00:00, 17.02it/s]

Epoch: 88 Training Loss: 0.027962822234258056 Training Accuracy: 1.0
Test Loss: 0.003402691218070686 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.29it/s]
100%|██████████| 2/2 [00:00<00:00, 16.86it/s]

Epoch: 89 Training Loss: 0.025496240428765304 Training Accuracy: 1.0
Test Loss: 0.003052834654226899 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.31it/s]
100%|██████████| 2/2 [00:00<00:00, 17.54it/s]

Epoch: 90 Training Loss: 0.025207325466908514 Training Accuracy: 1.0
Test Loss: 0.0032173607032746077 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 17.01it/s]
100%|██████████| 2/2 [00:00<00:00, 17.41it/s]

Epoch: 91 Training Loss: 0.025427175220102072 Training Accuracy:
0.9996462464332581 Test Loss: 0.0030788666335865855 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.51it/s]
100%|██████████| 2/2 [00:00<00:00, 16.80it/s]

Epoch: 92 Training Loss: 0.023931457195430994 Training Accuracy: 1.0
Test Loss: 0.0028920925687998533 Test Accuracy: 1.0
```

```
100%|██████████| 15/15 [00:00<00:00, 16.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.24it/s]

Epoch: 93 Training Loss: 0.022542103644809686 Training Accuracy:
0.9996462464332581 Test Loss: 0.0025579786160960793 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.73it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 94 Training Loss: 0.022428976953960955 Training Accuracy: 1.0
Test Loss: 0.002433618064969778 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00, 12.86it/s]
100%|██████████| 2/2 [00:00<00:00, 11.15it/s]

Epoch: 95 Training Loss: 0.020608626888133585 Training Accuracy: 1.0
Test Loss: 0.002606781432405114 Test Accuracy: 1.0

100%|██████████| 15/15 [00:01<00:00, 12.05it/s]
100%|██████████| 2/2 [00:00<00:00, 17.55it/s]

Epoch: 96 Training Loss: 0.02017039677593857 Training Accuracy: 1.0
Test Loss: 0.0023445168044418097 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.81it/s]
100%|██████████| 2/2 [00:00<00:00, 17.46it/s]

Epoch: 97 Training Loss: 0.01873483185772784 Training Accuracy: 1.0
Test Loss: 0.002315979450941086 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.87it/s]
100%|██████████| 2/2 [00:00<00:00, 17.00it/s]

Epoch: 98 Training Loss: 0.01880612422246486 Training Accuracy: 1.0
Test Loss: 0.0023057444486767054 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.50it/s]
100%|██████████| 2/2 [00:00<00:00, 17.25it/s]

Epoch: 99 Training Loss: 0.017812869395129383 Training Accuracy: 1.0
Test Loss: 0.00205195602029562 Test Accuracy: 1.0

100%|██████████| 15/15 [00:00<00:00, 16.55it/s]
100%|██████████| 2/2 [00:00<00:00, 16.57it/s]

Epoch: 100 Training Loss: 0.01802925457013771 Training Accuracy: 1.0
Test Loss: 0.002137118368409574 Test Accuracy: 1.0
```

```python
class PositionalEncoding(nn.Module):
    def __init__(self, emb_size: int, dropout, maxlen: int = 5000):
        super(PositionalEncoding, self).__init__()
```

```
        den = torch.exp(- torch.arange(0, emb_size, 2) *
math.log(10000) / emb_size)
        pos = torch.arange(0, maxlen).reshape(maxlen, 1)
        pos_embedding = torch.zeros((maxlen, emb_size))
        pos_embedding[:, 0::2] = torch.sin(pos * den)
        pos_embedding[:, 1::2] = torch.cos(pos * den)
        pos_embedding = pos_embedding.unsqueeze(-2)

        self.dropout = nn.Dropout(dropout)
        self.register_buffer('pos_embedding', pos_embedding)

    def forward(self, token_embedding):
        return self.dropout(token_embedding +

self.pos_embedding[:token_embedding.size(0),:])

def generate_square_subsequent_mask(sz):
    mask = (torch.triu(torch.ones((sz, sz), device=device)) ==
1).transpose(0, 1)
    mask = mask.float().masked_fill(mask == 0, float('-
inf')).masked_fill(mask == 1, float(0.0))
    return mask

def create_mask(src, tgt):
  src_seq_len = src.shape[0]
  tgt_seq_len = tgt.shape[0]

  tgt_mask = generate_square_subsequent_mask(tgt_seq_len)
  src_mask = torch.zeros((src_seq_len, src_seq_len),
device=device).type(torch.bool)

  src_padding_mask = (src == 0).transpose(0, 1)
  tgt_padding_mask = (tgt == 0).transpose(0, 1)
  return src_mask, tgt_mask, src_padding_mask, tgt_padding_mask

class TransformerModel(nn.Module):
    def __init__(self, num_encoder_layers, nhead, num_decoder_layers,
                 emb_size, src_vocab_size, tgt_vocab_size,
                 dim_feedforward:int = 512, dropout:float = 0.1):
        super(TransformerModel, self).__init__()
        encoder_layer = nn.TransformerEncoderLayer(d_model=emb_size,
nhead=nhead,

dim_feedforward=dim_feedforward)
        self.transformer_encoder =
nn.TransformerEncoder(encoder_layer, num_layers=num_encoder_layers)
        decoder_layer = nn.TransformerDecoderLayer(d_model=emb_size,
nhead=nhead,

dim_feedforward=dim_feedforward)
```

```
        self.transformer_decoder =
nn.TransformerDecoder(decoder_layer, num_layers=num_decoder_layers)

        self.generator = nn.Linear(emb_size, tgt_vocab_size)
        self.emb_size = emb_size
        self.src_tok_emb = self.embedding =
nn.Embedding(src_vocab_size, emb_size)
        self.tgt_tok_emb = self.embedding =
nn.Embedding(tgt_vocab_size, emb_size)
        self.positional_encoding = PositionalEncoding(emb_size,
dropout=dropout)

    def forward(self, src, trg, src_mask,
                tgt_mask, src_padding_mask,
                tgt_padding_mask, memory_key_padding_mask):
        src_emb = self.positional_encoding(self.src_tok_emb(src)*
math.sqrt(self.emb_size))
        tgt_emb = self.positional_encoding(self.tgt_tok_emb(trg)*
math.sqrt(self.emb_size))
        memory = self.transformer_encoder(src_emb, src_mask,
src_padding_mask)
        outs = self.transformer_decoder(tgt_emb, memory, tgt_mask,
None,
                                        tgt_padding_mask,
memory_key_padding_mask)
        return self.generator(outs)

model = TransformerModel(num_encoder_layers=6, nhead=8,
num_decoder_layers=6,
                 emb_size=512, src_vocab_size=(len(d.get_alphabet()) +
1), tgt_vocab_size=(len(d.get_alphabet()) + 1),
                 dim_feedforward = 512, dropout = 0.2).to(device)
```

```
/usr/local/lib/python3.11/dist-packages/torch/nn/modules/
transformer.py:385: UserWarning: enable_nested_tensor is True, but
self.use_nested_tensor is False because
encoder_layer.self_attn.batch_first was not True(use batch_first for
better inference performance)
  warnings.warn(
```

```
def train_epoch_transformer(model, train_loader, optimizer, criterion,
batch_size):
  model.train()
  total_loss = 0
  num_correct = 0
  total_items = 0
  for src, tgt in tqdm(train_loader):
      src = src.to(device).T
      tgt = tgt.to(device).T
```

```
      tgt_input = tgt[:-1, :]

      src_mask, tgt_mask, src_padding_mask, tgt_padding_mask =
create_mask(src, tgt_input)

      logits = model(src, tgt_input, src_mask, tgt_mask,
                                 src_padding_mask, tgt_padding_mask,
src_padding_mask)

      optimizer.zero_grad()

      tgt_out = tgt[1:,:]
      loss = criterion(logits.reshape(-1, logits.shape[-1]),
tgt_out.reshape(-1))

      loss.backward()
      optimizer.step()

      total_loss += loss.item()
      total_items += (tgt_out != 0).sum(dim=(0,1))

      num_correct += (torch.logical_and((logits.argmax(dim=2) ==
tgt_out), (tgt_out != 0))).sum(dim=(0,1))
  return total_loss / len(train_loader), num_correct / total_items

def test_epoch_transformer(model, test_loader, criterion, batch_size):
  model.eval()
  total_loss = 0
  num_correct = 0
  total_items = 0
  for src, tgt in tqdm(train_loader):
      src = src.to(device).T
      tgt = tgt.to(device).T

      tgt_input = tgt[:-1, :]

      src_mask, tgt_mask, src_padding_mask, tgt_padding_mask =
create_mask(src, tgt_input)

      logits = model(src, tgt_input, src_mask, tgt_mask,
                                 src_padding_mask, tgt_padding_mask,
src_padding_mask)


      tgt_out = tgt[1:,:]
      loss = criterion(logits.reshape(-1, logits.shape[-1]),
tgt_out.reshape(-1))


      total_loss += loss.item()
      total_items += (tgt_out != 0).sum(dim=(0,1))
```

```
      num_correct += (torch.logical_and((logits.argmax(dim=2) ==
tgt_out), (tgt_out != 0))).sum(dim=(0,1))
  return total_loss / len(train_loader), num_correct / total_items

def train_transformer(model, train_dataset, test_dataset,
batch_size=32, epochs=100):
  train_loader = torch.utils.data.DataLoader(train_dataset,
batch_size=batch_size, shuffle=True)
  test_loader = torch.utils.data.DataLoader(test_dataset,
batch_size=batch_size, shuffle=True)
  criterion = nn.CrossEntropyLoss()

  optim = torch.optim.Adam(model.parameters(), lr=1e-4, betas=(0.9,
0.98), eps=1e-9)

  for e in range(epochs):
    train_loss, train_acc = train_epoch_transformer(model,
train_loader, optim, criterion, batch_size=batch_size)
    test_loss, test_acc = train_epoch_transformer(model, test_loader,
optim, criterion, batch_size=batch_size)
    print(f'Epoch: {e + 1} Training Loss: {train_loss} Training
Accuracy: {train_acc} Test Loss: {test_loss} Test Accuracy:
{test_acc}')

train_transformer(model,train_dataset, test_dataset)

  0%|          | 0/15 [00:00<?, ?it/s]/usr/local/lib/python3.11/dist-
packages/torch/nn/functional.py:5962: UserWarning: Support for
mismatched key_padding_mask and attn_mask is deprecated. Use same type
for both instead.
  warnings.warn(
100%|██████████| 15/15 [00:01<00:00, 13.73it/s]
100%|██████████| 2/2 [00:00<00:00, 17.12it/s]

Epoch: 1 Training Loss: 1.7636406501134236 Training Accuracy:
0.0003537318843882531 Test Loss: 1.1341528594493866 Test Accuracy: 0.0

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 17.32it/s]

Epoch: 2 Training Loss: 1.1200006524721782 Training Accuracy: 0.0 Test
Loss: 0.9625139534473419 Test Accuracy: 0.0

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 16.90it/s]

Epoch: 3 Training Loss: 0.7780845403671265 Training Accuracy:
0.10647329688072205 Test Loss: 0.6794078946113586 Test Accuracy:
0.2848101258277893
```

```
100%|██████████| 15/15 [00:01<00:00, 14.13it/s]
100%|██████████| 2/2 [00:00<00:00, 17.31it/s]

Epoch: 4 Training Loss: 0.600431497891744 Training Accuracy:
0.3480721712112427 Test Loss: 0.5285531878471375 Test Accuracy:
0.37025317549705505

100%|██████████| 15/15 [00:01<00:00, 13.57it/s]
100%|██████████| 2/2 [00:00<00:00, 15.67it/s]

Epoch: 5 Training Loss: 0.5014954576889674 Training Accuracy:
0.38910505175590515 Test Loss: 0.5376769751310349 Test Accuracy:
0.3765822649002075

100%|██████████| 15/15 [00:01<00:00, 13.36it/s]
100%|██████████| 2/2 [00:00<00:00, 14.15it/s]

Epoch: 6 Training Loss: 0.4548841724793116 Training Accuracy:
0.44393348693847656 Test Loss: 0.5151402205228806 Test Accuracy:
0.3955696225166321

100%|██████████| 15/15 [00:01<00:00, 14.12it/s]
100%|██████████| 2/2 [00:00<00:00, 17.36it/s]

Epoch: 7 Training Loss: 0.4545358161131541 Training Accuracy:
0.4867350459098816 Test Loss: 0.4469589740037918 Test Accuracy:
0.4556961953639984

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.18it/s]

Epoch: 8 Training Loss: 0.40749405721823373 Training Accuracy:
0.508666455745697 Test Loss: 0.4501529335975647 Test Accuracy:
0.503164529800415

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 16.42it/s]

Epoch: 9 Training Loss: 0.3860988179842631 Training Accuracy:
0.5472232103347778 Test Loss: 0.39351411163806915 Test Accuracy:
0.5063291192054749

100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 17.21it/s]

Epoch: 10 Training Loss: 0.36147660613059995 Training Accuracy:
0.5779978632926941 Test Loss: 0.3738846331834793 Test Accuracy:
0.5189873576164246

100%|██████████| 15/15 [00:01<00:00, 14.17it/s]
100%|██████████| 2/2 [00:00<00:00, 17.18it/s]
```

```
Epoch: 11 Training Loss: 0.3577210485935211 Training Accuracy:
0.5808277130126953 Test Loss: 0.35955682396888733 Test Accuracy:
0.5917721390724182

100%|██████████| 15/15 [00:01<00:00, 14.14it/s]
100%|██████████| 2/2 [00:00<00:00, 17.31it/s]

Epoch: 12 Training Loss: 0.32921397089958193 Training Accuracy:
0.6108949184417725 Test Loss: 0.3603157550096512 Test Accuracy:
0.5632911324501038

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.09it/s]

Epoch: 13 Training Loss: 0.3302729686101278 Training Accuracy:
0.60523521900177 Test Loss: 0.39723385870456696 Test Accuracy:
0.5443037748336792

100%|██████████| 15/15 [00:01<00:00, 14.18it/s]
100%|██████████| 2/2 [00:00<00:00, 17.06it/s]

Epoch: 14 Training Loss: 0.3048802465200424 Training Accuracy:
0.620445728302002 Test Loss: 0.31027017533779144 Test Accuracy:
0.5981012582778931

100%|██████████| 15/15 [00:01<00:00, 13.53it/s]
100%|██████████| 2/2 [00:00<00:00, 13.89it/s]

Epoch: 15 Training Loss: 0.3081524521112442 Training Accuracy:
0.6452069282531738 Test Loss: 0.321362242102623 Test Accuracy:
0.607594907283783

100%|██████████| 15/15 [00:01<00:00, 13.38it/s]
100%|██████████| 2/2 [00:00<00:00, 13.57it/s]

Epoch: 16 Training Loss: 0.3084881653388341 Training Accuracy:
0.6459143757820129 Test Loss: 0.30073346197605133 Test Accuracy:
0.6170886158943176

100%|██████████| 15/15 [00:01<00:00, 14.00it/s]
100%|██████████| 2/2 [00:00<00:00, 17.30it/s]

Epoch: 17 Training Loss: 0.30950251122315725 Training Accuracy:
0.6416696310043335 Test Loss: 0.3003014326095581 Test Accuracy:
0.607594907283783

100%|██████████| 15/15 [00:01<00:00, 13.95it/s]
100%|██████████| 2/2 [00:00<00:00, 16.73it/s]

Epoch: 18 Training Loss: 0.27910560369491577 Training Accuracy:
0.6529890298843384 Test Loss: 0.25348155200481415 Test Accuracy:
0.6613923907279968
```

```
100%|██████████| 15/15 [00:01<00:00, 14.06it/s]
100%|██████████| 2/2 [00:00<00:00, 17.26it/s]

Epoch: 19 Training Loss: 0.25543903609116875 Training Accuracy:
0.6717368364334106 Test Loss: 0.2686879187822342 Test Accuracy:
0.6360759735107422

100%|██████████| 15/15 [00:01<00:00, 14.15it/s]
100%|██████████| 2/2 [00:00<00:00, 16.25it/s]

Epoch: 20 Training Loss: 0.24829145272572836 Training Accuracy:
0.6706756353378296 Test Loss: 0.2854607328772545 Test Accuracy:
0.6202531456947327

100%|██████████| 15/15 [00:01<00:00, 14.19it/s]
100%|██████████| 2/2 [00:00<00:00, 17.33it/s]

Epoch: 21 Training Loss: 0.2486149569352468 Training Accuracy:
0.6869472861289978 Test Loss: 0.2384543940424919 Test Accuracy:
0.6803797483444214

100%|██████████| 15/15 [00:01<00:00, 14.18it/s]
100%|██████████| 2/2 [00:00<00:00, 17.26it/s]

Epoch: 22 Training Loss: 0.22806900689999263 Training Accuracy:
0.6947293877601624 Test Loss: 0.26191744208335876 Test Accuracy:
0.6518987417221069

100%|██████████| 15/15 [00:01<00:00, 14.19it/s]
100%|██████████| 2/2 [00:00<00:00, 17.22it/s]

Epoch: 23 Training Loss: 0.22770930131276448 Training Accuracy:
0.702157735824585 Test Loss: 0.22885208576917648 Test Accuracy:
0.6803797483444214

100%|██████████| 15/15 [00:01<00:00, 14.18it/s]
100%|██████████| 2/2 [00:00<00:00, 17.37it/s]

Epoch: 24 Training Loss: 0.21603461503982543 Training Accuracy:
0.7152458429336548 Test Loss: 0.21471550315618515 Test Accuracy:
0.702531635761261

100%|██████████| 15/15 [00:01<00:00, 13.64it/s]
100%|██████████| 2/2 [00:00<00:00, 15.99it/s]

Epoch: 25 Training Loss: 0.21151464382807414 Training Accuracy:
0.7251503467559814 Test Loss: 0.2310900092124939 Test Accuracy:
0.6740506291389465

100%|██████████| 15/15 [00:01<00:00, 13.27it/s]
100%|██████████| 2/2 [00:00<00:00, 13.37it/s]
```

```
Epoch: 26 Training Loss: 0.20736295282840728 Training Accuracy:
0.7293950915336609 Test Loss: 0.2160254269838333 Test Accuracy:
0.6835442781448364

100%|██████████| 15/15 [00:01<00:00, 13.93it/s]
100%|██████████| 2/2 [00:00<00:00, 16.70it/s]

Epoch: 27 Training Loss: 0.19978972176710766 Training Accuracy:
0.7318712472915649 Test Loss: 0.19615165889263153 Test Accuracy:
0.7405063509941101

100%|██████████| 15/15 [00:01<00:00, 14.15it/s]
100%|██████████| 2/2 [00:00<00:00, 17.12it/s]

Epoch: 28 Training Loss: 0.2096669554710388 Training Accuracy:
0.746020495891571 Test Loss: 0.18347736448049545 Test Accuracy:
0.7215189933776855

100%|██████████| 15/15 [00:01<00:00, 14.27it/s]
100%|██████████| 2/2 [00:00<00:00, 17.32it/s]

Epoch: 29 Training Loss: 0.1950877959529559 Training Accuracy:
0.7262115478515625 Test Loss: 0.19249065965414047 Test Accuracy:
0.7183544039726257

100%|██████████| 15/15 [00:01<00:00, 14.19it/s]
100%|██████████| 2/2 [00:00<00:00, 16.68it/s]

Epoch: 30 Training Loss: 0.18378318051497142 Training Accuracy:
0.7576936483383179 Test Loss: 0.18232107162475586 Test Accuracy:
0.746835470199585

100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 16.95it/s]

Epoch: 31 Training Loss: 0.1716981366276741 Training Accuracy:
0.7640608549118042 Test Loss: 0.16852500289678574 Test Accuracy:
0.7626582384109497

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 16.33it/s]

Epoch: 32 Training Loss: 0.1677375207344691 Training Accuracy:
0.7679519057273865 Test Loss: 0.18842300027608871 Test Accuracy:
0.7436708807945251

100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 17.22it/s]

Epoch: 33 Training Loss: 0.17807841102282207 Training Accuracy:
0.7665369510650635 Test Loss: 0.16814687103033066 Test Accuracy:
0.7689873576164246
```

```
100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.14it/s]

Epoch: 34 Training Loss: 0.16024058510859807 Training Accuracy:
0.7817474603652954 Test Loss: 0.16155903786420822 Test Accuracy:
0.7816455960273743

100%|██████████| 15/15 [00:01<00:00, 13.48it/s]
100%|██████████| 2/2 [00:00<00:00, 14.33it/s]

Epoch: 35 Training Loss: 0.15654378930727642 Training Accuracy:
0.7852847576141357 Test Loss: 0.15131421759724617 Test Accuracy:
0.8354430198669434

100%|██████████| 15/15 [00:01<00:00, 13.16it/s]
100%|██████████| 2/2 [00:00<00:00, 13.26it/s]

Epoch: 36 Training Loss: 0.14912796318531035 Training Accuracy:
0.7824549078941345 Test Loss: 0.1444828286767006 Test Accuracy:
0.7784810066223145

100%|██████████| 15/15 [00:01<00:00, 14.00it/s]
100%|██████████| 2/2 [00:00<00:00, 16.56it/s]

Epoch: 37 Training Loss: 0.15130898555119832 Training Accuracy:
0.7997877597808838 Test Loss: 0.14540113136172295 Test Accuracy:
0.8354430198669434

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 16.56it/s]

Epoch: 38 Training Loss: 0.14201646149158478 Training Accuracy:
0.7955429553985596 Test Loss: 0.1301439180970192 Test Accuracy:
0.797468364238739

100%|██████████| 15/15 [00:01<00:00, 14.23it/s]
100%|██████████| 2/2 [00:00<00:00, 16.94it/s]

Epoch: 39 Training Loss: 0.1352076053619385 Training Accuracy:
0.8132295608520508 Test Loss: 0.12227135896682739 Test Accuracy:
0.8322784900665283

100%|██████████| 15/15 [00:01<00:00, 14.19it/s]
100%|██████████| 2/2 [00:00<00:00, 17.14it/s]

Epoch: 40 Training Loss: 0.13421365122000375 Training Accuracy:
0.8192430138587952 Test Loss: 0.126407440751791 Test Accuracy:
0.8544303774833679

100%|██████████| 15/15 [00:01<00:00, 14.19it/s]
100%|██████████| 2/2 [00:00<00:00, 17.48it/s]
```

```
Epoch: 41 Training Loss: 0.1323437293370565 Training Accuracy:
0.8111071586608887 Test Loss: 0.1124887578189373 Test Accuracy:
0.8417721390724182

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 15.69it/s]

Epoch: 42 Training Loss: 0.13040104508399963 Training Accuracy:
0.8277325630187988 Test Loss: 0.12003499269485474 Test Accuracy:
0.8196202516555786

100%|██████████| 15/15 [00:01<00:00, 14.29it/s]
100%|██████████| 2/2 [00:00<00:00, 16.28it/s]

Epoch: 43 Training Loss: 0.12362907081842422 Training Accuracy:
0.8344534635543823 Test Loss: 0.1212119348347187 Test Accuracy:
0.8227847814559937

100%|██████████| 15/15 [00:01<00:00, 14.27it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 44 Training Loss: 0.1129241148630778 Training Accuracy:
0.8411743640899658 Test Loss: 0.11526673287153244 Test Accuracy:
0.8544303774833679

100%|██████████| 15/15 [00:01<00:00, 13.72it/s]
100%|██████████| 2/2 [00:00<00:00, 15.07it/s]

Epoch: 45 Training Loss: 0.11894912570714951 Training Accuracy:
0.8461266160011292 Test Loss: 0.10455360263586044 Test Accuracy:
0.844936728477478

100%|██████████| 15/15 [00:01<00:00, 13.17it/s]
100%|██████████| 2/2 [00:00<00:00, 13.84it/s]

Epoch: 46 Training Loss: 0.13032147884368897 Training Accuracy:
0.8280863165855408 Test Loss: 0.1126636229455471 Test Accuracy:
0.8797468543052673

100%|██████████| 15/15 [00:01<00:00, 14.01it/s]
100%|██████████| 2/2 [00:00<00:00, 17.08it/s]

Epoch: 47 Training Loss: 0.12111741751432419 Training Accuracy:
0.8425893187522888 Test Loss: 0.10914091020822525 Test Accuracy:
0.8670886158943176

100%|██████████| 15/15 [00:01<00:00, 14.28it/s]
100%|██████████| 2/2 [00:00<00:00, 17.46it/s]

Epoch: 48 Training Loss: 0.11237291942040126 Training Accuracy:
0.8524938225746155 Test Loss: 0.09231944382190704 Test Accuracy:
0.8892405033111572
```

```
100%|██████████| 15/15 [00:01<00:00, 14.30it/s]
100%|██████████| 2/2 [00:00<00:00, 16.90it/s]

Epoch: 49 Training Loss: 0.1012232132256031 Training Accuracy:
0.8623983263969421 Test Loss: 0.11598848924040794 Test Accuracy:
0.8512658476829529

100%|██████████| 15/15 [00:01<00:00, 14.28it/s]
100%|██████████| 2/2 [00:00<00:00, 16.65it/s]

Epoch: 50 Training Loss: 0.09848604102929433 Training Accuracy:
0.8761938214302063 Test Loss: 0.10062927380204201 Test Accuracy:
0.8702531456947327

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 17.20it/s]

Epoch: 51 Training Loss: 0.096525144080321 Training Accuracy:
0.8747789263725281 Test Loss: 0.08414621278643608 Test Accuracy:
0.8987341523170471

100%|██████████| 15/15 [00:01<00:00, 14.33it/s]
100%|██████████| 2/2 [00:00<00:00, 16.49it/s]

Epoch: 52 Training Loss: 0.10778210759162903 Training Accuracy:
0.8687654733657837 Test Loss: 0.10394179821014404 Test Accuracy:
0.8607594966888428

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.45it/s]

Epoch: 53 Training Loss: 0.09158030537267527 Training Accuracy:
0.8684117197990417 Test Loss: 0.08006657660007477 Test Accuracy:
0.892405092716217

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.02it/s]

Epoch: 54 Training Loss: 0.08537267769376437 Training Accuracy:
0.8910505771636963 Test Loss: 0.08198797330260277 Test Accuracy:
0.8892405033111572

100%|██████████| 15/15 [00:01<00:00, 13.83it/s]
100%|██████████| 2/2 [00:00<00:00, 15.35it/s]

Epoch: 55 Training Loss: 0.08394771069288254 Training Accuracy:
0.8942341804504395 Test Loss: 0.07259756699204445 Test Accuracy:
0.9145569801330566

100%|██████████| 15/15 [00:01<00:00, 13.35it/s]
100%|██████████| 2/2 [00:00<00:00, 13.07it/s]
```

```
Epoch: 56 Training Loss: 0.08602552736798923 Training Accuracy:
0.8899893760681152 Test Loss: 0.07028179988265038 Test Accuracy:
0.9082278609275818

100%|██████████| 15/15 [00:01<00:00, 14.06it/s]
100%|██████████| 2/2 [00:00<00:00, 17.43it/s]

Epoch: 57 Training Loss: 0.08465000713864962 Training Accuracy:
0.8977714776992798 Test Loss: 0.07572223991155624 Test Accuracy:
0.8955696225166321

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.38it/s]

Epoch: 58 Training Loss: 0.0774790920317173 Training Accuracy:
0.888574481010437 Test Loss: 0.07657942175865173 Test Accuracy:
0.9018987417221069

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 17.10it/s]

Epoch: 59 Training Loss: 0.07027409076690674 Training Accuracy:
0.9087371826171875 Test Loss: 0.06872427649796009 Test Accuracy:
0.9082278609275818

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 60 Training Loss: 0.06632997654378414 Training Accuracy:
0.9115670323371887 Test Loss: 0.05863911285996437 Test Accuracy:
0.9367088675498962

100%|██████████| 15/15 [00:01<00:00, 14.25it/s]
100%|██████████| 2/2 [00:00<00:00, 16.34it/s]

Epoch: 61 Training Loss: 0.07044944216807683 Training Accuracy:
0.9133356809616089 Test Loss: 0.0587503369897604 Test Accuracy:
0.9335442781448364

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 16.12it/s]

Epoch: 62 Training Loss: 0.07872462769349416 Training Accuracy:
0.9122744798660278 Test Loss: 0.06111224927008152 Test Accuracy:
0.9240506291389465

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.04it/s]

Epoch: 63 Training Loss: 0.06323813771208127 Training Accuracy:
0.9147506356239319 Test Loss: 0.061677346006035805 Test Accuracy:
0.9208860993385315
```

```
100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 16.87it/s]

Epoch: 64 Training Loss: 0.058186422040065126 Training Accuracy:
0.9225327372550964 Test Loss: 0.05337911285459995 Test Accuracy:
0.9208860993385315

100%|██████████| 15/15 [00:01<00:00, 13.78it/s]
100%|██████████| 2/2 [00:00<00:00, 15.01it/s]

Epoch: 65 Training Loss: 0.059198230504989624 Training Accuracy:
0.9306685328483582 Test Loss: 0.048712434247136116 Test Accuracy:
0.9367088675498962

100%|██████████| 15/15 [00:01<00:00, 13.36it/s]
100%|██████████| 2/2 [00:00<00:00, 13.55it/s]

Epoch: 66 Training Loss: 0.05814261039098104 Training Accuracy:
0.928899884223938 Test Loss: 0.04844902828335762 Test Accuracy:
0.9272152185440063

100%|██████████| 15/15 [00:01<00:00, 13.88it/s]
100%|██████████| 2/2 [00:00<00:00, 17.15it/s]

Epoch: 67 Training Loss: 0.06543823989729086 Training Accuracy:
0.9055535793304443 Test Loss: 0.05839039944112301 Test Accuracy:
0.9240506291389465

100%|██████████| 15/15 [00:01<00:00, 14.14it/s]
100%|██████████| 2/2 [00:00<00:00, 14.79it/s]

Epoch: 68 Training Loss: 0.05727504349003235 Training Accuracy:
0.9197028875350952 Test Loss: 0.06743132136762142 Test Accuracy:
0.9367088675498962

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 17.07it/s]

Epoch: 69 Training Loss: 0.05177913755178452 Training Accuracy:
0.9366819858551025 Test Loss: 0.04250246845185757 Test Accuracy:
0.9430379867553711

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.20it/s]

Epoch: 70 Training Loss: 0.05380632268885772 Training Accuracy:
0.9292536377906799 Test Loss: 0.04736972972750664 Test Accuracy:
0.9398733973503113

100%|██████████| 15/15 [00:01<00:00, 14.12it/s]
100%|██████████| 2/2 [00:00<00:00, 17.22it/s]
```

```
Epoch: 71 Training Loss: 0.04790613353252411 Training Accuracy:
0.9366819858551025 Test Loss: 0.046591656282544136 Test Accuracy:
0.9367088675498962

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.00it/s]

Epoch: 72 Training Loss: 0.052671687801678975 Training Accuracy:
0.944817841053009 Test Loss: 0.052089957520365715 Test Accuracy:
0.9335442781448364

100%|██████████| 15/15 [00:01<00:00, 14.17it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 73 Training Loss: 0.053204167758425076 Training Accuracy:
0.9257162809371948 Test Loss: 0.03957880288362503 Test Accuracy:
0.9462025165557861

100%|██████████| 15/15 [00:01<00:00, 14.20it/s]
100%|██████████| 2/2 [00:00<00:00, 16.53it/s]

Epoch: 74 Training Loss: 0.04531989221771558 Training Accuracy:
0.9426954388618469 Test Loss: 0.04356049560010433 Test Accuracy:
0.9556962251663208

100%|██████████| 15/15 [00:01<00:00, 13.59it/s]
100%|██████████| 2/2 [00:00<00:00, 15.46it/s]

Epoch: 75 Training Loss: 0.0420336705322067 Training Accuracy:
0.9451715350151062 Test Loss: 0.045106375589966774 Test Accuracy:
0.9335442781448364

100%|██████████| 15/15 [00:01<00:00, 13.45it/s]
100%|██████████| 2/2 [00:00<00:00, 13.22it/s]

Epoch: 76 Training Loss: 0.0395398985594511 Training Accuracy:
0.9487088918685913 Test Loss: 0.043181187473237514 Test Accuracy:
0.9303797483444214

100%|██████████| 15/15 [00:01<00:00, 13.84it/s]
100%|██████████| 2/2 [00:00<00:00, 17.34it/s]

Epoch: 77 Training Loss: 0.04299579386909803 Training Accuracy:
0.9441103935241699 Test Loss: 0.03557945601642132 Test Accuracy:
0.949367105960846

100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 17.24it/s]

Epoch: 78 Training Loss: 0.047601404661933584 Training Accuracy:
0.9334983825683594 Test Loss: 0.04886174947023392 Test Accuracy:
0.9462025165557861
```

```
100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 79 Training Loss: 0.046068937455614405 Training Accuracy:
0.9391581416130066 Test Loss: 0.0564511064440012 Test Accuracy:
0.9398733973503113

100%|██████████| 15/15 [00:01<00:00, 14.16it/s]
100%|██████████| 2/2 [00:00<00:00, 16.32it/s]

Epoch: 80 Training Loss: 0.04246805347502232 Training Accuracy:
0.9412804841995239 Test Loss: 0.04563616216182709 Test Accuracy:
0.9430379867553711

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.30it/s]

Epoch: 81 Training Loss: 0.04057946366568407 Training Accuracy:
0.9554297924041748 Test Loss: 0.028259780257940292 Test Accuracy:
0.9683544039726257

100%|██████████| 15/15 [00:01<00:00, 14.21it/s]
100%|██████████| 2/2 [00:00<00:00, 17.03it/s]

Epoch: 82 Training Loss: 0.03979699518531561 Training Accuracy:
0.9441103935241699 Test Loss: 0.046445487067103386 Test Accuracy:
0.9303797483444214

100%|██████████| 15/15 [00:01<00:00, 14.24it/s]
100%|██████████| 2/2 [00:00<00:00, 17.51it/s]

Epoch: 83 Training Loss: 0.03847205638885498 Training Accuracy:
0.9579058885574341 Test Loss: 0.042465053498744965 Test Accuracy:
0.949367105960846

100%|██████████| 15/15 [00:01<00:00, 14.26it/s]
100%|██████████| 2/2 [00:00<00:00, 17.25it/s]

Epoch: 84 Training Loss: 0.04001989979296923 Training Accuracy:
0.9497700929641724 Test Loss: 0.041662732139229774 Test Accuracy:
0.949367105960846

100%|██████████| 15/15 [00:01<00:00, 13.79it/s]
100%|██████████| 2/2 [00:00<00:00, 15.96it/s]

Epoch: 85 Training Loss: 0.03402865367631117 Training Accuracy:
0.9600282907485962 Test Loss: 0.030021014623343945 Test Accuracy:
0.9683544039726257

100%|██████████| 15/15 [00:01<00:00, 13.46it/s]
100%|██████████| 2/2 [00:00<00:00, 13.54it/s]
```

```
Epoch: 86 Training Loss: 0.02684358318025867 Training Accuracy:
0.9681641459465027 Test Loss: 0.02789481356739998 Test Accuracy:
0.9556962251663208

100%|██████████| 15/15 [00:01<00:00, 13.96it/s]
100%|██████████| 2/2 [00:00<00:00, 17.42it/s]

Epoch: 87 Training Loss: 0.025873734119037786 Training Accuracy:
0.9695790410041809 Test Loss: 0.026994275860488415 Test Accuracy:
0.9778481125831604

100%|██████████| 15/15 [00:01<00:00, 14.28it/s]
100%|██████████| 2/2 [00:00<00:00, 17.34it/s]

Epoch: 88 Training Loss: 0.025507727637887 Training Accuracy:
0.9678103923797607 Test Loss: 0.030624384991824627 Test Accuracy:
0.9556962251663208

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.16it/s]

Epoch: 89 Training Loss: 0.02915816828608513 Training Accuracy:
0.9646267890930176 Test Loss: 0.03128221724182367 Test Accuracy:
0.9588607549667358

100%|██████████| 15/15 [00:01<00:00, 14.22it/s]
100%|██████████| 2/2 [00:00<00:00, 17.47it/s]

Epoch: 90 Training Loss: 0.03255258407443762 Training Accuracy:
0.9536611437797546 Test Loss: 0.03141890373080969 Test Accuracy:
0.9398733973503113

100%|██████████| 15/15 [00:01<00:00, 14.28it/s]
100%|██████████| 2/2 [00:00<00:00, 17.34it/s]

Epoch: 91 Training Loss: 0.030202143515149753 Training Accuracy:
0.9579058885574341 Test Loss: 0.023282965645194054 Test Accuracy:
0.9683544039726257

100%|██████████| 15/15 [00:01<00:00, 14.30it/s]
100%|██████████| 2/2 [00:00<00:00, 15.99it/s]

Epoch: 92 Training Loss: 0.02663003218670686 Training Accuracy:
0.9653342962265015 Test Loss: 0.01994427852332592 Test Accuracy:
0.9746835231781006

100%|██████████| 15/15 [00:01<00:00, 14.33it/s]
100%|██████████| 2/2 [00:00<00:00, 17.36it/s]

Epoch: 93 Training Loss: 0.026248214580118657 Training Accuracy:
0.9639193415641785 Test Loss: 0.021221749018877745 Test Accuracy:
0.9715189933776855
```

```
100%|██████████| 15/15 [00:01<00:00, 14.26it/s]
100%|██████████| 2/2 [00:00<00:00, 17.56it/s]

Epoch: 94 Training Loss: 0.024216497130692006 Training Accuracy:
0.9656879901885986 Test Loss: 0.020702285692095757 Test Accuracy:
0.9715189933776855

100%|██████████| 15/15 [00:01<00:00, 13.87it/s]
100%|██████████| 2/2 [00:00<00:00, 14.86it/s]

Epoch: 95 Training Loss: 0.0259569825604558 Training Accuracy:
0.9639193415641785 Test Loss: 0.026517481543123722 Test Accuracy:
0.9556962251663208

100%|██████████| 15/15 [00:01<00:00, 13.61it/s]
100%|██████████| 2/2 [00:00<00:00, 14.23it/s]

Epoch: 96 Training Loss: 0.02926748382548491 Training Accuracy:
0.9667491912841797 Test Loss: 0.023091770708560944 Test Accuracy:
0.9651898741722107

100%|██████████| 15/15 [00:01<00:00, 13.90it/s]
100%|██████████| 2/2 [00:00<00:00, 17.24it/s]

Epoch: 97 Training Loss: 0.03820722190042337 Training Accuracy:
0.9515387415885925 Test Loss: 0.03663494065403938 Test Accuracy:
0.9588607549667358

100%|██████████| 15/15 [00:01<00:00, 14.32it/s]
100%|██████████| 2/2 [00:00<00:00, 16.82it/s]

Epoch: 98 Training Loss: 0.02349249434967836 Training Accuracy:
0.9699327945709229 Test Loss: 0.021861808374524117 Test Accuracy:
0.9683544039726257

100%|██████████| 15/15 [00:01<00:00, 14.25it/s]
100%|██████████| 2/2 [00:00<00:00, 17.27it/s]

Epoch: 99 Training Loss: 0.01900718950976928 Training Accuracy:
0.9784223437309265 Test Loss: 0.016292295651510358 Test Accuracy:
0.9810126423835754

100%|██████████| 15/15 [00:01<00:00, 14.30it/s]
100%|██████████| 2/2 [00:00<00:00, 17.53it/s]

Epoch: 100 Training Loss: 0.01901131837318341 Training Accuracy:
0.9787760972976685 Test Loss: 0.01650015451014042 Test Accuracy:
0.9715189933776855
```