ДОДАТОК А

Код генерації тексту

```
import numpy as np
import pandas as pd
import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Embedding

df = pd.read_csv('fake_news_500.csv')
fake_news_text = ' '.join(df[df['label'] == 'fake']['text'].values)

tokenizer = Tokenizer(char_level=True)
tokenizer.fit_on_texts(fake_news_text)
total_chars = len(tokenizer.word_index) + 1

sequences = tokenizer.texts_to_sequences([fake_news_text])[0]

seq_length = 100
X = []
y = []
for i in range(0, len(sequences) - seq_length):
    X.append(sequences[i:i + seq_length])
    y.append(sequences[i + seq_length])

X = np.array(X)
y = to_categorical(y, num_classes=total_chars)

print(f'Input shape: {X.shape}, Output shape: {y.shape}')
```

Рисунок А.1 – Код програми роботи токенізатора

```
model = Sequential()
model.add(Embedding(total_chars, 50, input_length=seq_length))
model.add(LSTM(100, return_sequences=True))
model.add(LSTM(100))
model.add(Dense(total_chars, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
model.summary()
```

Рисунок А.2 – Код програми RNN

```python
def generate_text(model, tokenizer, seq_length, seed_text, num_chars):
    result = []
    in_text = seed_text
    for _ in range(num_chars):
        encoded = tokenizer.texts_to_sequences([in_text])[0]
        encoded = np.array(encoded[-seq_length:])
        encoded = np.expand_dims(encoded, 0)
        y_pred = model.predict(encoded, verbose=0)
        y_pred = np.argmax(y_pred, axis=-1)
        out_char = tokenizer.index_word[y_pred[0]]
        in_text += out_char
        result.append(out_char)
    return ''.join(result)

seed_text = "Breaking news: "
generated_text = generate_text(model, tokenizer, seq_length, seed_text, 8000)

print(f'Generated text:\n{generated_text}')
```

```
Executed at 2024.06.01 13:49:12 in 3m 12s 690ms
```

```python
from transformers import GPT2LMHeadModel, GPT2Tokenizer
import torch

def generate_text(prompt, model_name='gpt2', max_length=100, temperature=0.7, num_return_sequences=1):

    tokenizer = GPT2Tokenizer.from_pretrained(model_name)
    model = GPT2LMHeadModel.from_pretrained(model_name)

    # Encode the prompt
    input_ids = tokenizer.encode(prompt, return_tensors='pt')

    # Generate text
    output = model.generate(
        input_ids,
        max_length=max_length,
        temperature=temperature,
        num_return_sequences=num_return_sequences,
        do_sample=True,
        top_k=50,
        top_p=0.95
    )

    # Decode the generated text
    generated_texts = [tokenizer.decode(output_seq, skip_special_tokens=True) for output_seq in output]

    return generated_texts
```

Рисунок А.3 – Код програми генерації тексту з GPT-2

```
class VAE(nn.Module):
    def __init__(self, vocab_size, embedding_dim, hidden_dim, latent_dim):
        super(VAE, self).__init__()
        self.vocab_size = vocab_size
        self.embedding_dim = embedding_dim
        self.hidden_dim = hidden_dim
        self.latent_dim = latent_dim

        self.embedding = nn.Embedding(vocab_size, embedding_dim)
        self.encoder = nn.LSTM(embedding_dim, hidden_dim, batch_first=True)
        self.hidden_to_mean = nn.Linear(hidden_dim, latent_dim)
        self.hidden_to_logvar = nn.Linear(hidden_dim, latent_dim)

        self.latent_to_hidden = nn.Linear(latent_dim, hidden_dim)
        self.decoder = nn.LSTM(embedding_dim, hidden_dim, batch_first=True)
        self.outputs_to_vocab = nn.Linear(hidden_dim, vocab_size)
```

Рисунок А.4 – Код програми генерації тексту з VAE

ДОДАТОК Б

Відомість кваліфікаційної роботи

Позначення	Найменування	Дод. відомості
	Текстові документи	
1.	Пояснювальна записка	70 с.
	Інші документи	
2.	Презентаційні матеріали	14 плакатів

.	.	Прізвище та ініціали.	Підп.	Дата			
Розробив		Писаренко К.В.			Створення та використання синтетичних текстових даних для покращення продуктивності моделей машинного навчання в умовах недостатньої доступності реальних даних	Шифр групи	Код напр./спец.
Перевірив		Вітько О.В.				СШІм-22-3	122
Н.контр.		Малєєва І.А.				ХНУРЕ кафедра ШІ	
Затв.		Філатов В.О.					