

ДОДАТОК А

Перелік джерел посилання за науковими напрямками керівника та науковців
кафедри програмної інженерії

10. Євгеній Мамочка, Андрій Єрохін. «Дослідження Текстових Генеративних Систем з Асоціативною Пам'яттю». Збірник 12ої Міжнародної науково-технічної конференції «Інформаційні системи та технології» (ICT-2023) Частина 2. Молодіжна секція.

11. Mamochka E. I., student, Gerasymchuk T. V. « NEURAL NETWORKS. GPT TECHNOLOGY. MIDJOURNEY». Збірник наукових статей. Питання сучасної модернізації науки та освіти. Харків – 2023р.

12. Filatov V. O., Yerokhin A. L., Zolotukhin O. V., Kudryavtseva M. S. Hybrid simulation models for complex decision-making problems with partial uncertainty. Information Extraction and Processing. 2022, 50(126), 78-86. DOI:<https://doi.org/10.15407/vidbir2022.50.078>

13. Dmytro Panchenko, Daniil Maksymenko, Olena Turuta, Andriy Yerokhin, Yana Daniil, Oleksii Turuta . Evaluation and Analysis of the NLP Model Zoo for Ukrainian Text Classification // Communications in Computer and Information Science, 2022, 1698 CCIS, pp. 109–123. DOI: 10.1007/978-3-031-20834-8_6

14. Daniil Maksymenko, Nataliia Saichyshyna, Oleksii Turuta, Olena Turuta, Andriy Yerokhin, Andrii Babii. Improving the Machine Translation Model in Specific Domains for the Ukrainian Language // International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2022, 2022-November, pp. 123–129. DOI: 10.1109/CSIT56902.2022.10000529

15. Extension Multi30K: Multimodal Dataset for Integrated Vision and Language Research in Ukrainian. Nataliia Saichyshyna, Daniil Maksymenko, Oleksii Turuta, Andriy Yerokhin, Andrii Babii, Olena Turuta / EACL 2023 - 2nd Ukrainian Natural Language Processing Workshop, UNLP 2023 - Proceedings of the Workshop, 2023, P.P. 54–61 / DOI: 10.18653/v1/2023.unlp-1.7

ДОДАТОК Б

Результати перевірки на унікальність тексту в базі ХНУРЕ

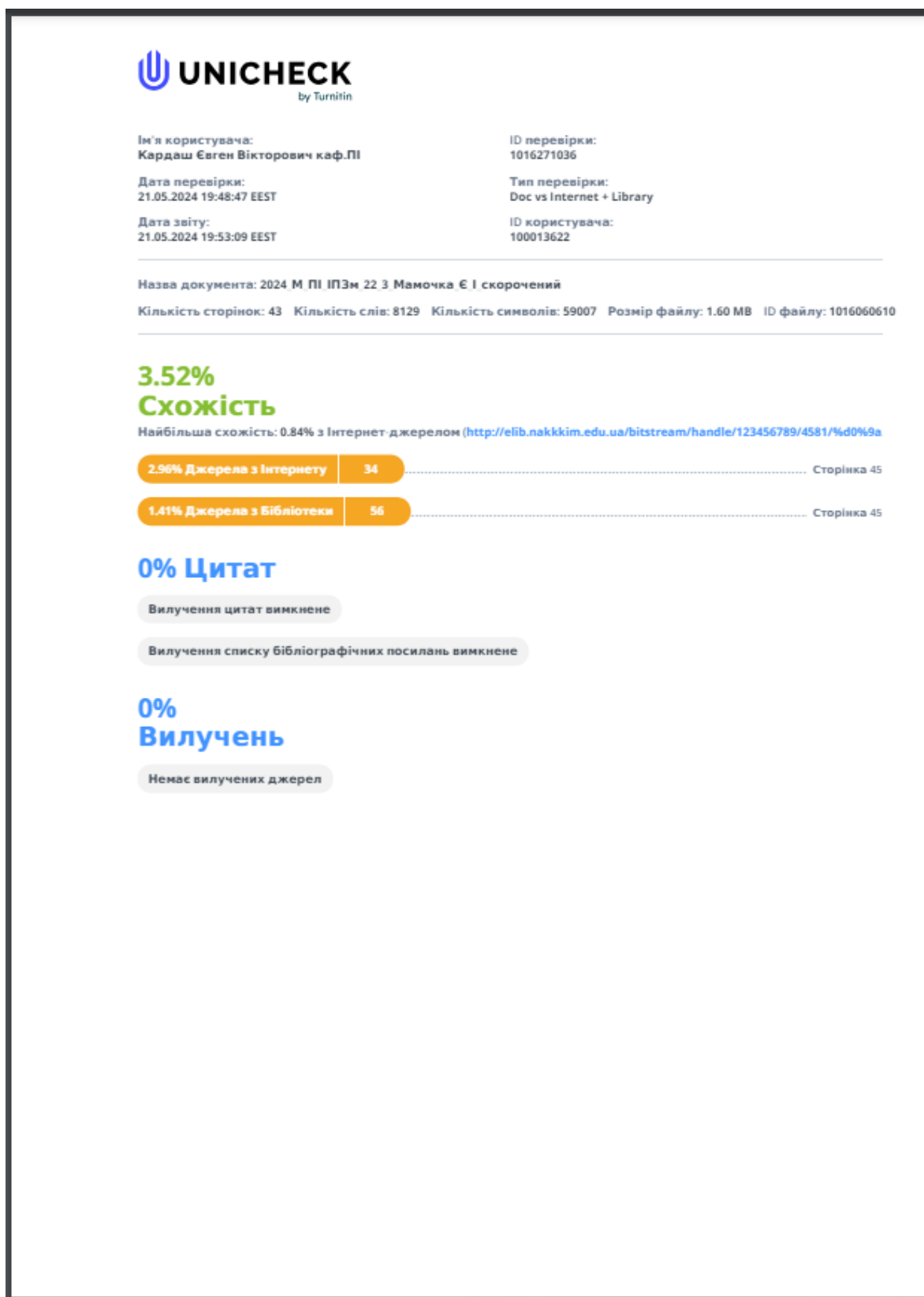


Рисунок Б.1 – Титульний аркуш звіту результатів перевірки на унікальність тексту в базі ХНУРЕ

ДОДАТОК В

Слайди презентації



**ДОСЛІДЖЕННЯ
ТЕКСТОВИХ
ГЕНЕРАТИВНИХ СИСТЕМ
З АСОЦІАТИВНОЮ
ПАМ'ЯТТЮ**

Виконав:
Мамочка Є. І

Керівник:
Єрохін А. Л.

Група:
ІПЗм-22-3

Дата захисту:
07.05.2023

NURE
Харківський національний
університет
радіоелектроніки

Рисунок В.1 – Перший слайд презентації



ОСНОВНІ ВІДОМОСТІ

2

ОБ'ЄКТ
Генеративні системи з різними типами пам'яті.

МЕТА
Дослідити сферу машинного навчання, нейронні мережі та їх типи пам'яті, провести практичне дослідження генеративних системи з різними типами пам'яті.

ЗАВДАННЯ
Виконати порівняння двох генеративних систем з метою виявлення доцільності використання асоціативної пам'яті у даних системах.

Рисунок В.2 – Другий слайд презентації



Рисунок В.3 – Третій слайд презентації

АНАЛІЗ ГАЛУЗІ

Машинне навчання – це наука про розробку алгоритмів та моделей, що дозволяють комп'ютерам виконувати завдання без конкретних інструкцій, спираючись на аналіз шаблонів та логічні висновки.

4



Рисунок В.4 – Четвертий слайд презентації

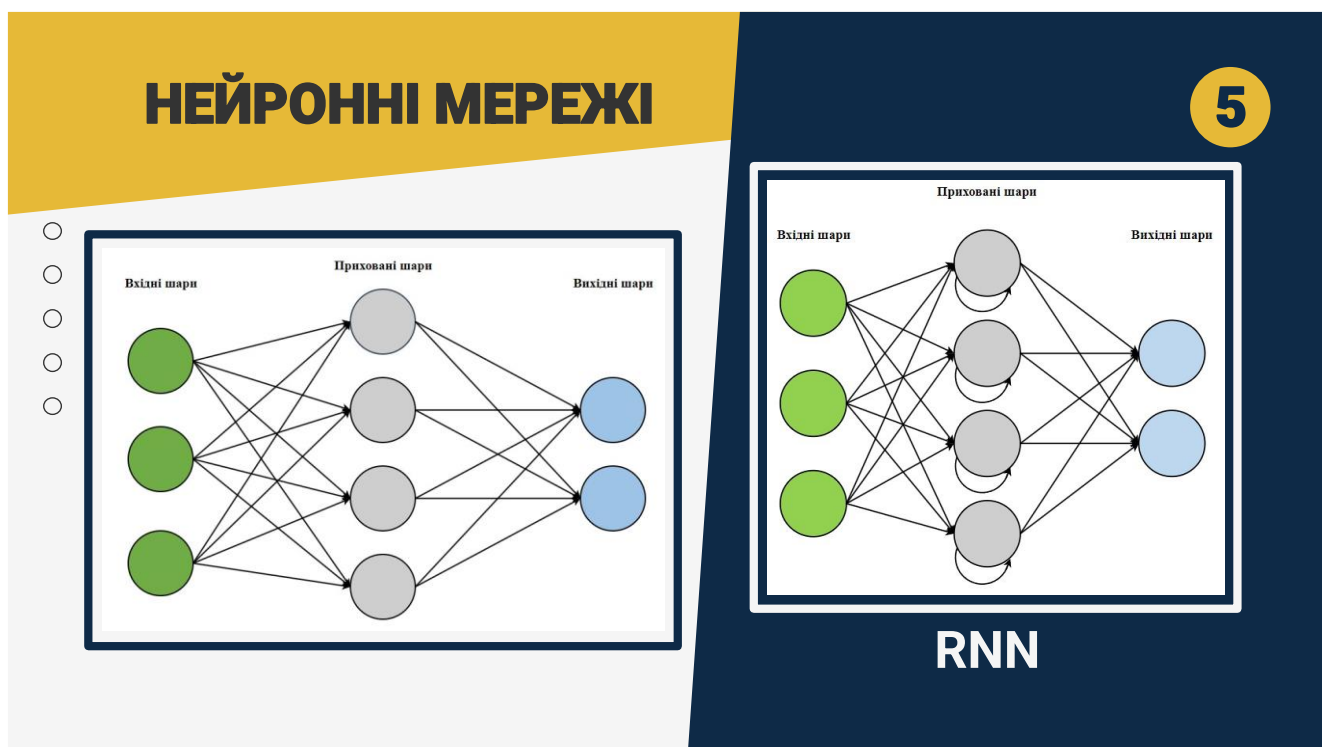


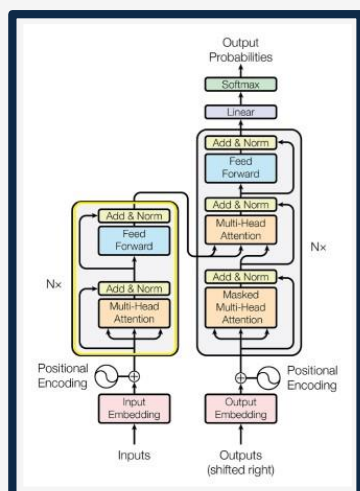
Рисунок В.5 – П'ятий слайд презентації



Рисунок В.6 – Шостий слайд презентації

ТРАНСФОРМЕР

7

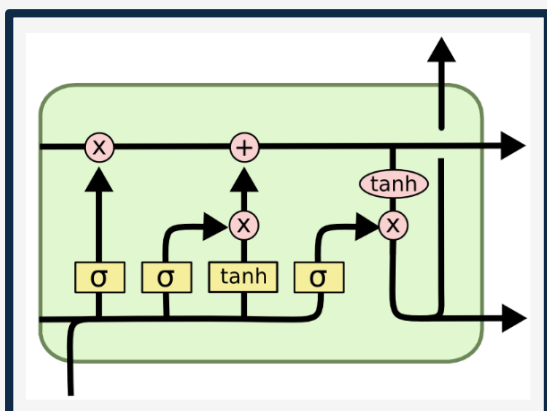


МЕХАНІЗМ УВАГИ

Рисунок В.7 – Сьомий слайд презентації

LONG SHORT TERM MEMORY


8



GATE

Рисунок В.8 – Восьмий слайд презентації

GPT-2 ASM. ОСНОВНІ ВІДОМОСТІ ПРО РЕАЛІЗАЦІЮ



9

MAIN

GPT-2

MODEL

MEMORY

VISUALIZATOR




Рисунок В.9 – Дев’ятий слайд презентації

РЕАЛІЗАЦІЯ ПАМ’ЯТІ

10



**СТРУКТУРА
АСОЦІАТИВНОЇ
ПАМ’ЯТІ**



**СТРУКТУРА МУЛЬТИМОДАЛЬНИХ
ГОЛІВ**

Рисунок В.10 – Десятий слайд презентації

13

АПРОБАЦІЯ ДОСЛІДЖЕННЯ

Матеріали дослідження були опубліковані у двох збірниках з конференцій




ICT-2023



**ЗБІРНИК НАУКОВИХ СТАТЕЙ
ПИТАННЯ СУЧАСНОЇ
МОДЕРНІЗАЦІЇ НАУКИ ТА
ОСВІТИ**

Рисунок В.13 – Тринадцятий слайд презентації



**Харківський Національний
Університет Радіоелектроніки**

Дякую за те, що прослухали мою доповідь для захисту магістерського дослідження

Зворотній зв'язок:
E-mail: yevhenii.mamochka@nure.ua

Рисунок В.14 – Чотирнадцятий слайд презентації

ДОДАТОК Г

Тексти наукових публікацій за темою кваліфікаційної роботи

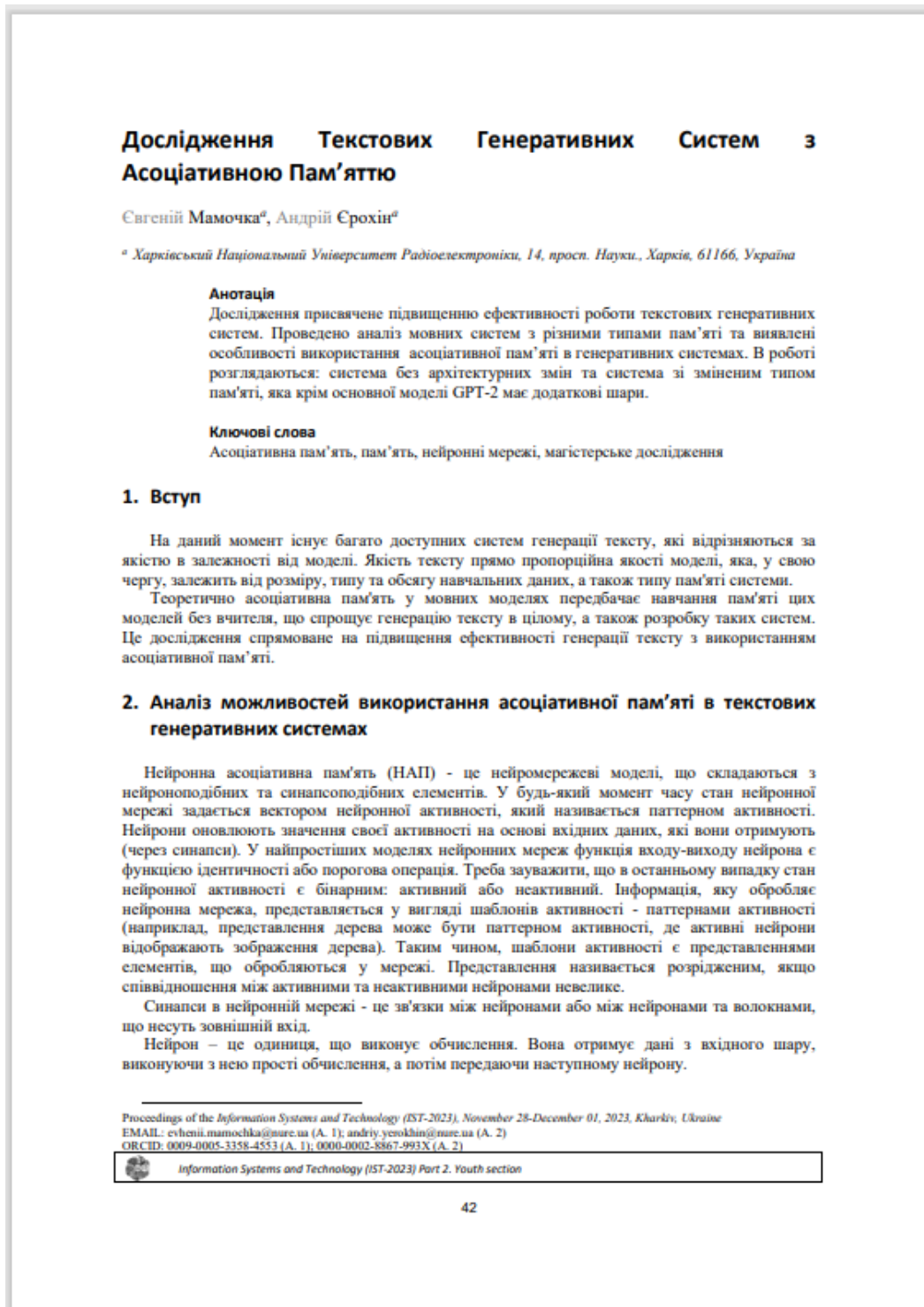


Рисунок Г.1 – Тези магістерського дослідження з 12ої Міжнародної науково-технічної конференції «Інформаційні системи та технології» (ІСТ-2023) Частина 2. Молодіжна секція. Перша сторінка

Різні функції пам'яті визначаються тим, як можна отримати доступ до вивчених шаблонів вибірково звертаючись до вхідного зразка. Функція розпізнавання зразків означає класифікування вхідних зразків на два класи: знайомі зразки та решту. Асоціація шаблонів описує функцію асоціювання певних вхідних зразків з певними зразками у пам'яті, тобто кожна пам'ять складається з пари вхідного і бажаного вихідного зразків.

Асоціативна пам'ять у нейронних мережах стосується їхньої здатності запам'ятовувати та асоціювати вхідні дані чи образи. Цей термін охоплює кілька аспектів:

1) пам'ять образів – мережі можуть запам'ятовувати та розпізнавати конкретні образи у вхідних даних, виявляючи шаблони та закономірності;

2) асоціативне запам'ятовування – мережі відновлюють інформацію з пам'яті, користуючись частинами вхідних даних або ключовими елементами, що дозволяє їм асоціювати вхідні дані з певними виходами;

3) навчання з учителем і без учителя – мережі можуть навчатися з учителем, коли надаються пари вхід-вихід, або без учителя, коли вони самостійно виявляють закономірності в даних;

4) зіставлення схожих образів – система порівнює вхідні дані з відомими зразками, що дозволяє їй робити висновки або прогнози на основі схожості.

Перевага нейронної асоціативної пам'яті над іншими алгоритмами зберігання шаблонів, наприклад, таблицями пошуку хеш-кодів, полягає в тому, що доступ до пам'яті може бути відмовостійким по відношенню до змін вхідного шаблону. Для асоціації шаблонів це означає, що вихідний шаблон може бути створений для набору вхідних шаблонів, які є найближчі до вхідного шаблону, представленого під час навчання.

Нейронні мережі можуть досить ефективно працювати з різними типами даних, такими як зображення, звук, відео, текст тощо. Тоді з точки зору підвищення ефективності роботи текстових генеративних систем важливою та актуальною задачею є саме дослідження асоціативного типу пам'яті. Використання асоціативної пам'яті у мовних моделях є доцільним, оскільки це спрощує перевірку асоціації між словами та контекстом під час генерації тексту.

3. Мета дослідження і постановка задачі дослідження

Головна мета дослідження – порівняння двох мовних систем з різними типами пам'яті та виявлення доцільності використання саме асоціативної пам'яті в генеративних системах.

Задачі, які необхідно виконати під час проведення дослідження:

- 1) знайти мовну модель, яка відповідає конкретним вимогам дослідження;
- 2) адаптувати обрану мовну модель для роботи з асоціативною пам'яттю, що становить основу дослідження;
- 3) детально задокументувати процес розробки системи;
- 4) підготувати дані для навчання мовної системи;
- 5) провести навчання розроблених систем на конкретних вхідних даних;
- 6) порівняти дві отримані нейронні мережі на однакових даних з метою визначення ефективності використання асоціативної пам'яті.

Навчання даних систем виконувалось на сервері Microsoft Azure.

Через велику кількість даних в матрицях, було визнано доцільним розпаралелити дії, для чого було використано один із видів оптимізації, а саме – TPU.

Для перевірки даних систем використовувалась метрика *accuracy*, а саме – порівняння очікуваних даних і даних нейронної мережі.

4. Реалізація задач дослідження

Отримана система є трансформером GPT-2, тобто системою для обробки текстових послідовностей, таких як у завданнях машинного перекладу та автоматичного реферування. Основна його особливість – механізм уваги, який виявляє взаємозв'язки між частинами вхідних і вихідних даних. Модель GPT-2 використовує *transformer attention*, що дає змогу ефективно обробляти вхідні дані, враховуючи всі позиції в послідовності.

Рисунок Г.2 – Тези магістерського дослідження з 12ої Міжнародної науково-технічної конференції «Інформаційні системи та технології» (ICT-2023) Частина 2. Молодіжна секція. Друга сторінка

У результаті ми отримали дві системи: одна система без архітектурних змін, тобто має свій звичайний вид GPT-2, а друга система – зі зміненим типом пам'яті, тобто окрім основної моделі GPT-2 має додаткову мультимодальну "голову" входу. Вдосконалена система має такі додаткові шари (1):

Memory output -> Embedding -> Bidirectional LSTM -> Attention Mechanism. (1)

де *Memory output* – вивід додаткової моделі без вчителя, яка повинна створювати асоціації на основі автосоціативного принципу і поновлювати образи даних.

Спочатку дані обробляються системою пам'яті, далі дані стискаються за подібним змістом. На наступному кроці двонаправлена LSTM виконує роботу з даними (прогнозування послідовності, класифікація даних, розпізнавання іменованих даних, позначення частин мови), а на останньому кроці – вибираються важливі дані.

У результаті до двох розроблених систем ставляться однакові запитання (а також N - грами в процесі навчання). У трансформерів є ймовірність наступного слова: якщо текст буде написаний з більшим значенням ймовірності, то він вважається більш правильним також, якщо метрика точності буде високою. Тобто дві моделі дають схожу відповідь, але там, де точність відповіді більше, то там і краще текст. Окрім того враховується її відповідність до даних, на яких система навчалася.

5. Висновки

В результаті проведення дослідження виявилось, що асоціативна пам'ять є особливо ефективною при роботі з завданнями, які передбачають роботу з контекстом, інтерпретацію великих обсягів інформації та генерацію змісту з урахуванням попередніх знань. Проте ефективність цього методу може варіюватися в залежності від конкретної задачі, і деякі завдання можуть бути менш чутливими до використання асоціативної пам'яті.

6. Література

- [1] Haykin S. Neural Networks. A comprehensive Foundation. - 2ed. - 1999. -P. 690.
- [2] Zhao H. Global asymptotic stability of Hopfield neural network involving distributed delays - 2004. - Vol. 17, N 1. - P. 48 - 53.
- [3] Rini,J.M., Schulze-Gahmen,U. and Wilson,I.A. (1992) Science, 255, 959–965.
- [4] Narayanan,A., Sellers,B.D. and Jacobson,M.P. (2009) J. Mol. Biol., 388, 941–953.
- [5] Martin,A.C., Cheetham,J.C. and Rees,A.R. (1989) Proc. Natl Acad. Sci. USA, 86, 9268–9272.
- [6] Chothia,C. and Lesk,A.M. (1987) J. Mol. Biol., 196, 901–917.
- [7] Grefenstette,J.J. (ed.), Proceedings of the First International Conference on Genetic Algorithms. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 101–111.
- [8] Slyadnikov E. E. (2007). Physical model and associative memory of the cytoskeletal microtubule dipole system. Journal of Technical Physics, 77 (7), 77–86.

Рисунок Г.3 – Тези магістерського дослідження з 12ої Міжнародної науково-технічної конференції «Інформаційні системи та технології» (ІСТ-2023) Частина 2. Молодіжна секція. Третя сторінка

NEURAL NETWORKS. GPT TECHNOLOGY. MIDJOURNEY

Mamochka E. I., student,

Gerasymchuk T. V., Associate Professor,

Kharkiv National University of Radio Electronics

A neural network can be called a program, which is based on the principle of the brain function.

A neural network is a type of machine learning in which a program works on the principle of the human brain. No one knows 100% exactly how the brain works, but it is believed that this is the most approximate but simplified version. The neural network itself consists of a combination of neurons – layers. There are incoming, hidden and outgoing layers.

We should note that neural networks can be:

1. Single-layer (perceptron) structure of a neural network. It is a structure of neuronal interaction in which signals from the input layer are immediately directed to the output layer, which, in fact, not only converts the signal, but also immediately outputs a response.

2. Multi-layer (Deep) Neural Network. Here, in addition to the output and input layers, there are several other hidden intermediate layers. The number of these layers depends on the degree of complexity of the neural network. It is more like the structure of a biological neural network.

In addition to the number of layers, neural networks can be classified according to the direction of information distribution along synapses between neurons, but first you need to understand what neurons and synapses are.

Рисунок Г.4 – Всеукраїнська «Annual All-Ukrainian Bachelor & Master Students' Conference» та міжнародна «the International Contest of Students' Scientific Papers in English» конференції. Перша сторінка

A neuron is a unit that performs calculations. It receives data from the input layer by performing simple calculations with it, and then transmits it to the next neuron.

A synapse is a connection between neurons, and each synapse has its own weight. This is why the input data is modified during transmission. During processing, the information transmitted by the synapse with a high weight indicator will become overwhelming.

That is, the result is influenced not by neurons, but specifically by synapses, which give a set of weights of input data, because the neurons themselves constantly perform exactly the same calculations. The scales are set in random order. Currently, neural networks are developing very rapidly, and such systems can be distinguished, for example, GPT, MidJourney.

MidJourney system is a system for generating images. The user provides the system with a text or image, based on which another image is generated in accordance with the request. The system consists of several neural networks already trained on large amounts of data, performing their own functions for generating, processing, and improving images.

If the user enters the text, the first neural network converts it into a vector (an array of data), then the diffusion model (used to convert embedding an image into an image) creates a small initial image, and then the convolutional neural network performs an improvement (magnification) of the image, which is already ready to be provided to the user. Four options are created for such images so that the user can choose the best option.

The process of generating an image into an image uses only a diffusion and magnification model. When entering an image and text, the system performs vectorization of the text and image separately (the image vector is larger and more important than the text vector) and then, thanks to the diffusion model, it is combined into a single whole. The improvement process is similar to other options.

The GPT system is a text generation system created by OpenAI. There is such a version of GPT as ChatGPT, based on the system version 3.5. The user enters a

Рисунок Г.5 – Всеукраїнська «Annual All-Ukrainian Bachelor & Master Students' Conference» та міжнародна «the International Contest of Students' Scientific Papers in English» конференції. Друга сторінка

request, and the system analyzes the text and creates a response based on it and its knowledge base. Briefly speaking about the principle of operation, we should note that when generating the text continuation using GPT the following happens:

1. the input text is tokenized into a sequence of numbers (tokens).

2. the token list passes through the Embedding Layer and turns into an embedding list (very similar to word2vec).

3. A positive embedding is added to each embedding.

Unlike recurrent networks, the transformer architecture is not sensitive to the order of input tokens, that is, even if you mix words in places, the output will still be the same (permutation invariance).

But in speech word order is very important! To take it into account, I had to come up with a crutch - positive encoding. This mechanism allows transformers to "see" the order of incoming tokens.

4. Next, the list of embeddings begins its journey through several identical blocks (Transformer Decoder Block).

5. after the list of embeddings passes through the last block, the embedding corresponding to the last token is Matrix multiplied by the same input, but already transposed Embedding Layer, and after applying SoftMax, we get the probability distribution of the next token.

6. from this distribution we select the next token (for example, using the argmax function).

7. then we add this token to the input text and repeat steps 1-6.

Currently, the development of such systems is bringing humanity closer to creating a strong artificial intelligence that will help us solve everyday problems and the development of humanity as a whole. Neural networks are our future!

Literature:

1. Neural networks by Simon Heikin.
2. Artificial Neural Networks: calculus-Novotarsky M. A., Nesterenko B. B.

Рисунок Г.6 – Всеукраїнська «Annual All-Ukrainian Bachelor & Master Students' Conference» та міжнародна «the International Contest of Students' Scientific Papers in English» конференції. Третя сторінка

ДОДАТОК Е

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи		
студент <small>(ініціали)</small>	програмної інженерії <small>(кафедра)</small>	ПЗМ-22-3 <small>(група)</small>
Мамочка Є. І. <small>(прізвище, ім'я, по батькові)</small>		
Зауваження		
Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
7.1.20	Заголовки структурних елементів звіту та заголовки розділів треба друкувати з абзацного відступу великими літерами напівжирним шрифтом без крапки в кінці. Дозволено їх розміщувати посередині рядка.	6
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунок	
	7.6 Таблиці	
	7.7 Переліки	
7.7.2	Якщо подають переліки одного рівня підпорядкованості, на які у звіті немає посилань, то перед кожним із переліків ставлять знак «тире». Якщо у звіті є посилання на переліки, підпорядкованість позначають малими літерами української абетки, далі — арабськими цифрами, далі — через знаки «тире». Після цифри або літери певної позиції переліку ставлять круглу дужку.	10, далі за текстом. Дивиться методичні вказівки
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
7.10.4	Номер формули чи рівняння друкують на їх рівні праворуч у крайньому положенні в круглих дужках. У багаторядкових формулах або рівняннях їхній номер проставляють на рівні останнього рядка.	30
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	
6.2.2	Якщо додатки є продовженням тексту основної частини звіту, нумерація сторінок додатків — це продовження нумерації сторінок звіту. Кожний додаток повинен мати заголовок, який друкують згорі малими літерами з першої великої симетрично до тексту сторінки. Над заголовком, але посередині рядка, друкують слово «ДОДАТОК» і відповідну велику літеру української абетки, крім літер Г, Є, З, І, І, Й, О, Ч, Ъ, яка позначає додаток. Текст кожного	53
Методичні вказівки до виконання кваліфікаційної роботи магістра...	Слово ДОДАТОК та його назва розташовуються посередині сторінки без абзацного відступу, пишеться заголовними літерами звичайним начертанням(не жирним). Дивиться методичні вказівки.	53, далі за текстом.
Експерт <small>(підпис)</small>		Нечволод В.Ю. <small>(прізвище, ініціали)</small>
	22.05.2024	

Рисунок Е.1 – Експертний висновок результатів перевірки кваліфікаційної роботи