

Ганжа О. О., аспірант, Іванчук Н. В., к.т.н., доцент (Національний університет водного господарства та природокористування, м. Рівне)

ПОРІВНЯННЯ ЕФЕКТИВНОСТІ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ В ЗАДАЧАХ ГЕНЕРУВАННЯ ТЕКСТОВИХ ОПИСІВ ЗАПЧАСТИН ДО СІЛЬСЬКОГОСПОДАРСЬКОЇ ТЕХНІКИ

В статті розглянуто процес генерації текстових описів за допомогою моделей глибокого навчання. Було підготовлено набір даних для дослідження ефективності моделей в галузі запчастин до сільськогосподарської техніки. На розглянутому наборі проведено навчання моделей-трансформерів на архітектурах GPT-2, XLM-RoBERTa та ELECTRA. Для порівняння ефективності моделей використано машиноцентричні та людиноцентричні методи. Результати дослідження показали придатність моделей для генерування описів та виявили ряд недоліків методу. Дані результати можуть бути використані для подальшого вдосконалення методик генерації для отримання більш ефективних моделей, придатних для комерційного використання.

Ключові слова: нейронні мережі; глибоке навчання; обробка природної мови; генерація тексту; сільськогосподарська техніка.

Вступ. Однією із найпоширеніших галузей застосування моделей глибокого навчання є обробка природної мови (Natural Language Processing, NLP), що займається розумінням, аналізом та генерацією тексту або мовлення людей.

Останні роки принесли значний розвиток у генерації текстів. Нові моделі революціонізують цю галузь, створюючи текст, що надзвичайно схожий на той, що написаний людиною. Цей підхід набирає шаленої популярності в різних сферах, включаючи новини, соціальні мережі, створення сценаріїв для фільмів та поезії. Використання текстової генерації в різноманітних областях викликало значний інтерес у наукових колах до цієї технології. А здатність замінити людину у задачах написання комерційних текстів робить її економічно перспективною.

Найпершою задачею даного дослідження є порівняння ефективності різних архітектур моделей трансформерів в задачах

генерування текстових описів на основі заголовку. Дослідження передбачає використання вихідних даних з однієї спеціалізованої галузі, в якості якої було обрано запчастини до сільськогосподарської техніки.

Опис запчастини містить її коротку характеристику та перелік параметрів, що є критичними значеннями для підбору запчастини під потреби покупця. Окрім того надзвичайно важливим значенням є каталожний номер запчастини, що часто представляється як артикул чи код товару. Саме цей параметр є найбільш точним ідентифікатором товару і служить для первинного пошуку потрібної деталі. Також варто зауважити, що в більшості випадків назва запчастини вже містить в собі номер деталі за каталогом та найважливіші її характеристики, а отже саме назва і може слугувати вхідним значенням для генератора для подальшого створення розширеного опису.

Аналіз літературних джерел. Зважаючи на перспективність досліджень та популярність теми на сьогодні існує значна кількість робіт пов'язаних з розглянутою тематикою і їх кількість невпинно збільшується. Kalyani Roy, Pawan Goyal та Manish Pandey [1] розглядають задачу генерації опису як поєднання задач розпізнавання іменованих сутностей на основі переднавченої моделі GPT-2 та відповідей на запитання на основі моделі T5. Fajri Koto, Jey Han Lau, and Timothy Baldwin [2] досліджують питання створення розгорнутих текстових описів на основі масиву структурованих даних представлених у вигляді таблиці. Vitobha Munigala, Abhijit Mishra, Srikanth G Tamilselvam, Shreya Khare, Riddhiman Dasgupta, і Anush Sankaran у своїй праці [3] також розглядають задачу генерації короткого текстового опису на основі назви товару. Для цього вони використовують поєднання розпізнавання іменованих сутностей для визначення ключових слів з подальшою генерацією речення за променем (beam-search).

Набір даних. Для даного дослідження був підготований набір даних, що містить близько 19 000 екземплярів україномовних описів запчастин до сільськогосподарської техніки зібраних з різних інтернет-магазинів. Частина набору даних представлена на рис. 1.


name string · lengths	cat_n string · lengths	short_desc string · lengths	params string · lengths	text string · lengths
 10 171	 4 39	 20 480	 11 897	 50 1.19k
Шнек завантаження зерна - AH209421...	AH209421	Шнек вивантажування бункера до комбайнів John...	Параметри Вага , W kg 5 Зовнішній...	Шнек завантаження зерна - AH209421 John Deere , ...
Шнек елеватора - AH169021 John...	AH169021	Шнек подачі зерна з правосторонніми витками...	Параметри Вага , W kg 9.53 Зовнішній...	Шнек елеватора - AH169021 John Deere , правий, ...
Шнек висипу зерна AZ49211 підходить...	AZ49211	Шнек вивантажування бункера комбайнів John...	Параметри Вага , W kg 17 Зовнішній...	Шнек висипу зерна AZ49211 підходить для John Deere ...
Шнек висипу зерна AH124852 John...	AH124852	Шнек вивантажувальний вертикальний для комбайні...	Параметри Внутрішній діамет...	Шнек висипу зерна AH124852 John Deere [AM], 1102mm...
Шнек бункера - AXE12921 John...	AXE12921	Шнек зернового транспортера до комбайнів...	Параметри Вага , W kg 32.8 Внутрішній...	Шнек бункера - AXE12921 John Deere , 825мм Шнек...

Рис. 1. Набір даних для дослідження:

вихідні дані містять наступні поля: name – заголовок товару; cat_n – номер за каталогом; short_desc – короткий текстовий опис; params – характеристики товару; text – опис, що є комбінацією чотирьох попередніх полів

Для даного дослідження найбільшу цінність мають поля name, що виступає вхідним значенням для генерації та text, що є цілєю генерацією, та використовуватиметься для навчання та оцінювання ефективності моделей.

Модель. Дослідження базується на використанні переднавчених моделей трансформерів з різною архітектурою. Першою є GPT-2 [4] модель глибокого навчання, розроблена командою OpenAI. Вона є однією з ранніх версій серії GPT, яка використовує архітектуру трансформера для генерації тексту. GPT базується на декодерній частині трансформера та основною метою має генерування наступного токена на основі попередніх. Другою було обрано модель XLM-RoBERTa [5] – це BERT-подібна модель глибокого навчання, яка є розвитком RoBERTa (Robustly optimized BERT approach), спроектованого для розв'язання задач обробки природної мови (NLP). Однак, на відміну від RoBERTa, який призначений для англійської мови, XLM-RoBERTa призначений для роботи з текстами на різних мовах і враховує міжмовну взаємодію. XLM-RoBERTa базується на архітектурі трансформера, та включає в

себе механізми самозавантаження та позиційне кодування, які дозволяють моделі ефективно аналізувати та генерувати тексти. Ця модель може використовуватися для різних завдань NLP, включаючи генерацію текстів. Третьою була обрана ELECTRA [6] – архітектура моделі глибокого навчання, розроблена командою Google Research. ELECTRA – передбачає процес переднавчання моделі, який базується на ідеї використання ефективного дискримінатора для навчання генератора. ELECTRA відрізняється від інших моделей таких як BERT чи GPT, оскільки її архітектура заснована на двопрохідовій парадигмі, що робить її більш ефективною з точки зору обчислювальних ресурсів та швидкодії.

Оскільки обсяг навчальних даних не є достатньо великим для навчання повноцінних та ефективних моделей з нуля, у всіх трьох випадках використовувалися моделі, що попередньо навчалися на корпусах україномовних текстів, що забезпечили їм загальні знання про мову і світ суттєво поліпшивши їхню здатність до розуміння та генерації тексту.

Експеримент. Навчання всіх трьох моделей виконувалося в максимально подібних умовах. Модель GPT-2 по замовчуванню навчена для послідовної генерації (автодоповнення) тексту, тому при навчанні застосовувалися її стандартний токенизатор та модель. Решта дві вихідні моделі попередньо не були навчені для авторегресійної генерації, а призначалися для задачі заповнення масок в тексті (MLM), тому процес навчання передбачав також зміну задачі моделі на XLNetForCausalLM та ElectraForCausalLM відповідно. Навчання всіх моделей відбувалося при наступних параметрах: `learning_rate: 2e-05; train_batch_size: 8; eval_batch_size: 8; seed: 42; optimizer: Adam with betas=(0.9,0.999) and epsilon=1e-08; num_epochs: 3.0.`

Методи оцінки результатів. Існують два способи оцінки якості створеного тексту: людиноцентричний (HC) і машиноцентричний (MC) [7]. Людиноцентричний підхід полягає в залученні експертів з мови та предметної області для оцінки створеного тексту. Це дорого вартісне і трудомістке завдання, що не виключає людських помилок. З іншого боку, машиноцентричний підхід, відомий як об'єктивна оцінка якості, широко використовується і описується в літературі [8]. Цей підхід використовує різноманітні метрики оцінки. Для оцінки результатів даного експерименту використовувалися числові метрики, для вимірювання ефективності моделі на конкретному

завданні. Для даного випадку були обрані метрики: BLEU, METEOR і ROUGE. Зазвичай такі методики застосовують для оцінювання перекладів, але вони також застосовні і для перевірки якості генерації тексту. В розглянутому експерименті вхідним значенням слугує текст згенерований на основі назви товару, а еталоном – опис даного товару написаний людиною. В якості даних для оцінювання використовувалися випадково відібрані екземпляри з того ж набору даних, який застосовувався для навчання.

Результати. В ході оцінювання усіх моделей було отримано результати представлені в таблиці 1. Також для наочності дані представлені у вигляді діаграми на рис. 2.

Таблиця 1

Результати оцінки ефективності моделей

Модель	BLEU	Rouge1	Rouge2	RougeL	Rouge Lsum	METEOR
GPT-2	0,216412	0,360339	0,222327	0,336333	0,359177	0,545501
XLM-RoBERTa	0,259085	0,345882	0,184182	0,319778	0,340445	0,410934
ELECTRA	0,213430	0,308992	0,153474	0,274513	0,304263	0,412633

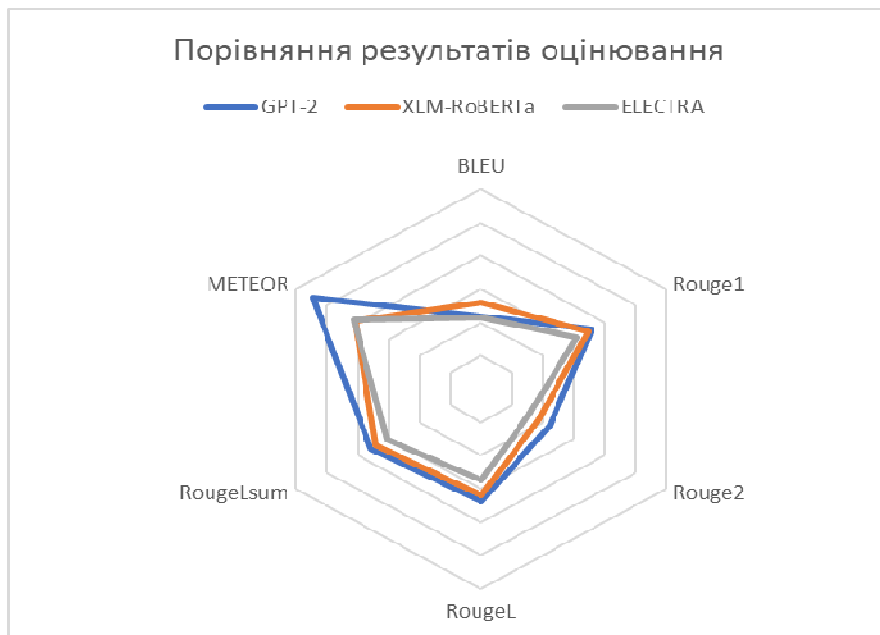


Рис. 2. Діаграма результатів оцінки ефективності моделей

Аналізуючи результати стає помітним, що GPT-2 демонструє найвищі показники майже за всіма критеріями крім метрики BLEU, де показники XLM-RoBERTa незначно випереджають її. Показовим є суттєвий відрив GPT-2 по метриці METEOR, яка є найбільш комплексною і зазвичай дає більш високу кореляцію з людською оцінкою якості моделі. Для порівняння в таблиці 2 наведені результати генерування опису різними моделями для екземпляру, що відсутній у навчальному наборі даних.

Таблиця 2

Приклад описів згенерованих моделями

Заголовок	0007347852 – 734785.0 – Підшипник кульковий закритий спеціальний [SLF]
GPT-2	0007347852 – 734785.0 – Підшипник кульковий закритий спеціальний – [SLF] Однорядний радіальний кульковий підшипник 6206-2RS C3 SLF, Італія. Розміри 25x52x15 мм. Відповідає підшипнику 180206 (ГОСТ 8882-75). Параметри Вага, W kg 0.045 Внутрішній діаметр, d mm 25 Зовнішній діаметр, D mm 52 Загальна ширина, T mm 15 Тип ущільнення Двостороннє
XLM-RoBERTa	0007347852 – 734785.0 – Підшипник кульковий закритий спеціальний – [SLF] Однорядний радіальний кульковий підшипник 6208-2RS SLF, Франція. Розміри 30x90x18 мм. Відповідає підшипнику 180180 (ГОСТ 8882-75). Параметри Вага, W kg 0.05 Внутрішній діаметр, d mm 30 Зовнішній діаметр, D mm 90
ELECTRA	0007347852 – 734785.0 – Підшипник кульковий закритий спеціальний- [SLF] Однорядний радіальний кульковий підшипник 620-2RS C3 Schaeffler, Німеччина. Розміри в мм: Внутрішній діаметр – 40 Зовнішній діаметр – 30. Каталогів номери: 66011, 66011. 0. Замінник.

На даному прикладі помітно, що всі моделі здатні генерувати описи подібні на природні, однак точність даних є досить низькою, що особливо помітно на описі характеристик товару.

Висновки. Результати дослідження показують здатність генеративних моделей-трансформерів на базі GPT-2, XLM-RoBERTa та ELECTRA до генерування текстових описів товарів для інтернет-магазинів. При цьому найкращі показники якості генерації має модель з архітектурою GPT-2, що підтверджує перевагу моделей, що базуються на декодерній частині трансформера в задачах авторегресійного генерування. Однак точність і достовірність таких описів є незадовільною, особливо в сферах чутливих до точності даних, якою є галузь запчастин. В таких випадках методика генерації потребує суттєвого доопрацювання.

1. Kalyani Roy, Pawan Goyal and Manish Pandey. Attribute Value Generation from Product Title using Language Models. *4th Workshop on e-Commerce and NLP*, Association for Computational Linguistics, Online. 2021. P. 13–17. 2. Fajri Koto, Jey Han Lau and Timothy Baldwin. Can Pretrained Language Models Generate Persuasive, Faithful, and Informative Ad Text for Product Descriptions?. *Proceedings of the Fifth Workshop on e-Commerce and NLP (ECNLP 5)*, Association for Computational Linguistics. Dublin, Ireland, 2022. P. 234–243. 3. Vitobha Munigala, Abhijit Mishra, Srikanth G. Tamilselvam, Shreya Khare, Riddhiman Dasgupta and Anush Sankaran. Persuade! an adaptive persuasive text generation system for fashion domain. *The Web Conference*. 2018. P. 335–342. 4. Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever et al. Language models are unsupervised multitask learners. *OpenAI blog*. 2019. Vol. 1(8). P. 9. 5. Conneau, Alexis, et al. Unsupervised cross-lingual representation learning at scale. URL: <https://arxiv.org/pdf/1911.02116.pdf> (дата звернення: 08.03.2024). 6. CLARK, Kevin, et al. Electra: Pre-training text encoders as discriminators rather than generators. 2020. URL: <https://arxiv.org/pdf/2003.10555.pdf>. (дата звернення: 20.03.2024). 7. J. Gu, Q. Wu and Z. Yu. Perception score a learned metric for open-ended text generation evaluation. *Proc. AAAI Conf. Artif. Intell.* 2021. Vol. 35, no. 14. Pp. 12902–12910. 8. N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota and A. Soomro. A Systematic Literature Review on Text Generation Using Deep Neural Network Models. *IEEE Access*. 2022. Vol. 10. Pp. 53490–53503. Doi: 10.1109/ACCESS.2022.3174108.

REFERENCES:

1. Kalyani Roy, Pawan Goyal and Manish Pandey. Attribute Value Generation from Product Title using Language Models. *4th Workshop on e-Commerce and NLP*, Association for Computational Linguistics, Online. 2021. P. 13–17. 2. Fajri Koto, Jey Han Lau and Timothy Baldwin. Can Pretrained Language Models

Generate Persuasive, Faithful, and Informative Ad Text for Product Descriptions?. *Proceedings of the Fifth Workshop on e-Commerce and NLP (ECNLP 5)*, Association for Computational Linguistics. Dublin, Ireland, 2022. P. 234–243. **3.** Vitobha Munigala, Abhijit Mishra, Srikanth G. Tamilselvam, Shreya Khare, Riddhiman Dasgupta and Anush Sankaran. Persuaide! an adaptive persuasive text generation system for fashion domain. *The Web Conference*. 2018. P. 335–342. **4.** Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever et al. Language models are unsupervised multitask learners. *OpenAI blog*. 2019. Vol. 1(8). P. 9. **5.** Conneau, Alexis, et al. Unsupervised cross-lingual representation learning at scale. URL: <https://arxiv.org/pdf/1911.02116.pdf> (data zvernennia: 20.03.2024). **6.** CLARK, Kevin, et al. Electra: Pre-training text encoders as discriminators rather than generators. 2020. URL: <https://arxiv.org/pdf/2003.10555.pdf>. (data zvernennia: 20.03.2024). **7.** J. Gu, Q. Wu and Z. Yu. Perception score a learned metric for open-ended text generation evaluation. *Proc. AAAI Conf. Artif. Intell.* 2021. Vol. 35, no. 14. Pp. 12902–12910. **8.** N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota and A. Soomro. A Systematic Literature Review on Text Generation Using Deep Neural Network Models. *IEEE Access*. 2022. Vol. 10. Pp. 53490–53503. Doi: 10.1109/ACCESS.2022.3174108.

Hanzha O. O., Post-graduate Student, Ivanchuk N. V., Candidate of Engineering (Ph.D.), Associate Professor (National University of Water and Environmental Engineering, Rivne)

COMPARISON OF THE EFFECTIVENESS OF DEEP LEARNING MODELS IN GENERATING TEXTUAL DESCRIPTIONS OF SPARE PARTS FOR AGRICULTURAL MACHINERY

In recent years, deep learning models have made significant strides in text generation, enabling the creation of text that closely resembles human writing. This has become highly popular in various domains such as news, social media, and creative writing. The use of text generation has sparked significant interest in academic circles, particularly due to its ability to replace humans in writing commercial texts. This study compares the effectiveness of different transformer model architectures in generating text descriptions based on headlines, using data on agricultural machinery parts.

For this study, a dataset comprising approximately 19,000 Ukrainian-language descriptions of agricultural machinery parts

collected from online stores was prepared. The main fields in the dataset include product name, catalog number, short description, specifications, and full description, which is a combination of the former. Three transformer models were selected for the study: GPT-2, XLM-RoBERTa, and ELECTRA, which were trained on the dataset. The performance of the models was evaluated using numerical metrics such as BLEU, METEOR, and ROUGE, using randomly selected instances from the same dataset used for training.

Analyzing the results, it is evident that GPT-2 shows the best performance overall, except for the BLEU metric, where XLM-RoBERTa slightly outperforms it. The example demonstrates that all models generate natural language, but the accuracy of the data is insufficient, especially in product specifications. The conclusions indicate that transformer models, especially GPT-2, are capable of generating product descriptions for online stores, but the accuracy of such descriptions leaves room for improvement, especially in the field of spare parts. In such cases, the generation method needs significant refinement.

Keywords: neural networks; deep learning; natural language processing; text generation; agricultural machinery.