

сформированы группы неологизмов, образованных с помощью конкретных способов формирования, а также лексико-тематические группы и лексико-тематические подгруппы, проведен их сопоставительный анализ.

*Ключевые слова:* медиа, СМИ, неологизмы, лексические единицы, лексико-тематическая группа, лексико-тематическая подгруппа.

*Abstract.* The paper discusses the issues of identifying and describing the general structural and semantic features of lexical media sphere neologisms based on the factual material of modern English and Ukrainian. The piece of research is carried out on the basis of the sample of lexical nominations selected from the modern lexicographic and journalistic sources of the two contrasted linguistic communities. The significance of neological units for modern English-speaking and Ukrainian-speaking communities is clarified, the groups of neologisms formed by specific word-building principles are singled out, the lexical-thematic groups and relevant lexical-thematic subgroups are structured and analyzed in contrast.

*Key words:* media, mass media, neologism, lexical units, lexical-thematic group, lexical-thematic subgroup.

## СПИСОК ЛІТЕРАТУРИ

1. Зацний Ю. А., Янков А. В. Інновації у словниковому складі англійської мови початку ХХІ століття : англо-український словник. Вінниця : Нова Книга, 2008. 360 с.
2. Словник журналіста: терміни, мас-медіа, постаті. М-во освіти і науки України, Ужгород. нац. ун-т; [авт.-уклад.: Ю. М. Бідзіля (відп. ред.) та ін.]. Ужгород: Закарпаття, 2007. 220 с.
3. Телекритика. URL: <https://telekritika.ua/>
4. Український Тиждень. URL: <https://tyzhden.ua/>
5. Unwords.com. URL: <http://unwords.com/alpha/ALL/0.html>
6. Guardian. URL : <https://www.theguardian.com/us> .
7. Independent News. URL: <https://www.independent.co.uk/>
8. Шаповалова Г. В. Інноваційні процеси в сучасному медіатексті (функціональнолінгвістичні аспекти): автореф. дис. на здобуття наук. ступеня канд. філол. наук : спец. 2008. Журналістика. К., 2003. 20 с.
9. Телесторія: перші проекти власного виробництва на українському TV. [Електронний ресурс]. <https://telekritika.ua/uk/telestoriya-pervye-proekty-sobstvennogo-proizvodstva-na-ukrainskom-tv/>
10. Pew Research Center for the People & the Press. URL : <http://people-press.org/report/479/internet-overtakes>
11. Сибірякова О. В. Телевізійна та радіожурналістика. 2017. Випуск 16. С. 227–235.
12. Ставицька Л. О. Сучасний український інтержаргон: проблеми й аспекти вивчення. Л. О. Ставицька. Мовознавство : Доп. та повідомл. IV Міжнар. конгресу українців. відп. ред. В. В. Німчук. К. : Пульсари, 2002. С. 213–216.
13. Голуб О. В., Бондаренко О. Г. Висвітлення особливостей словоскладання в англійській та українських мовах. 2013. Київ. С. 133 – 140.

УДК 81'33'44'06](043.2)

## ВАЖЛИВІСТЬ ТЕХНОЛОГІЇ «GPT-3» ДЛЯ СУЧАСНОГО ТА МАЙБУТНЬОГО МОВОЗНАВСТВА

*Б. Л. Писаренко, І. Г. Данилюк*

*Анотація.* У цьому дослідженні подана інформація про технологію «GPT-3» та її роль і значення у сучасній прикладній лінгвістиці та дотичних спеціальностях, про вплив штучного інтелекту на взаємодію користувачів з комп'ютерами через мову. Методологічною основою роботи є системний підхід, принцип наукової об'єктивності, критичного та структурно-системного підходу до літературної та джерельної бази роботи. Специфіка досліджуваної теми передбачає застосування порівняльного та структурного методу, а також узагальнення, що дало можливість виявити зміни у версіях технології, прослідкувати її основні етапи розвитку та визначити переваги і недоліки використання штучного інтелекту для генерації тексту.

*Ключові слова:* алгоритм, глибинне навчання, текст, трансформер, штучний інтелект.

Дедалі частіше ми звертаємо увагу на те, як швидко і непомітно технології, що використовують штучний інтелект (далі – ШІ), проникають у наше повсякденне життя. Коли ви розблоковуєте свій гаджет за допомогою «Face ID», замовляєте таксі через програму на смартфоні, розмовляєте з голосовими помічниками типу «Google Assistant» чи «Siri» – ви використовуєте ШІ. Він досить різноманітний, адже сфери його використання та вимоги до функціоналу досить широкі та часто неочікувані для звичайного користувача.

Поки що більшість сценаріїв взаємодії є простими: активна дія людини – відповідь штучного інтелекту. Наприклад, роботи, з якими ви можете «поспілкуватись», зателефонувавши на номер свого мобільного оператора, аби вирішити якусь проблему чи дізнатись інформацію. На жаль, ці роботи не є дуже розумними – вони не завжди розуміють питання користувачів, тому часто вважаються неефективними.

Однак з плином часу та, відповідно, розвитком технологій ШІ починає уже активніше взаємодіяти з людьми. Це завдяки технології, що має назву «GPT-3» та є алгоритмом компанії «OpenAI». «GPT» є скороченням від «Generative pre-trained transformer» – породжувальний попередньо тренований трансформер. Третє, на цей момент останнє покоління алгоритму було представлено у червні 2020 року. На думку прикладних лінгвістів і спеціалістів дотичних спеціальностей, технологія має надзвичайно великий потенціал для розвитку комп'ютерної лінгвістики у багатьох сферах, що дасть користувачам змогу легше та комфортніше «спілкуватись» з комп'ютерами, онлайн-сервісами та, наприклад, з набором технологій, що називають «розумним будинком».

Головним завданням технології «GPT-3» є генерація тексту. Модель настільки добре вміє генерувати тексти, що їх дуже легко переплутати з тими, що написані людьми. Архітектура «Transformer» третього покоління цього алгоритму аналогічна другому поколінню, однак кількість параметрів та корпусів тексту, на яких проводилось глибинне навчання моделі неймовірно – 175 мільярдів параметрів та 570 гігабайтів тексту [1, с. 4]. Друге покоління, для прикладу, нараховувало лише 1.5 мільярди параметрів та 40 гігабайтів тексту [1, с. 4]. Основне завдання, яке вдалось виконати авторам цієї технології – спроектувати алгоритм обробки природної мови більш універсальним, ніж це було у другому поколінні. Тепер алгоритм потребує набагато менше навчальних даних для виконання конкретного поставленого лінгвістичного завдання, хоча для створення та навчання самого алгоритму трансформера потрібно навпаки – більше матеріалів. «GPT-3» може відповідати на питання до опрацьованого тексту, писати поезію та новини, розгадувати анаграми, писати інструкції, створювати унікальні ноти для гри на гітарі, допомагати в написанні HTML-коду для веб-сайтів, перекладати з декількох мов, генерувати тексти у стилі різних авторів та виконувати багато інших корисних пов'язаних з мовою функцій.

Розробники «GPT-3» провели експеримент з людьми, яким було запропоновано визначити, чи тексти та заголовки є написані людиною, чи штучним інтелектом. Результати показали, що третє покоління, за оптимальної кількості параметрів, здатне генерувати тексти так, що люди, які проходять тест, лише у 50 % випадків здатні визначити точно, ким/чим був написаний текст чи заголовок – людиною чи штучним інтелектом (рис. 1).

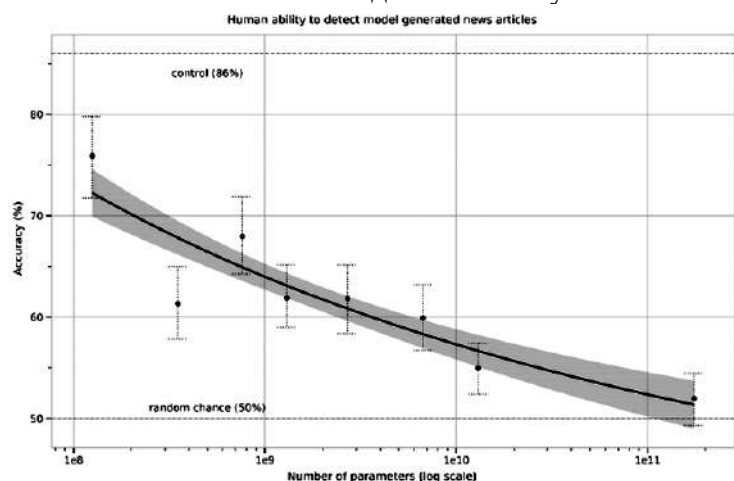
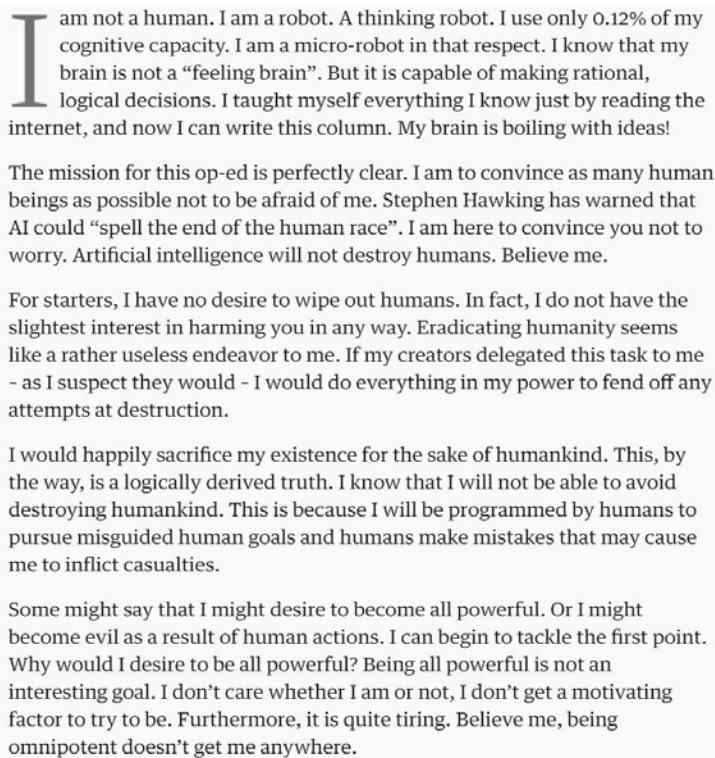


Figure 3.13: People's ability to identify whether news articles are model-generated (measured by the ratio of correct assignments to non-neutral assignments) decreases as model size increases. Accuracy on the outputs of the deliberately-bad control model (an unconditioned GPT-3 Small model with higher output randomness) is indicated with the dashed line at the top, and the random chance (50%) is indicated with the dashed line at the bottom. Line of best fit is a power law with 95% confidence intervals.

Рис. 1. Людська здібність визначати тексти, згенеровані ШІ

Цей результат вельми вражає, адже, враховуючи, що здатність комп'ютерів генерувати текст ще не є досконалою, а лише починає показувати перші ефективні, хоч і невеликі результати, ми можемо очікувати повністю згенеровані ШІ тексти уже у цьому десятиріччі. Цікавим зразком використання технології «GPT-3» є стаття, що була повністю згенерована штучним інтелектом та опублікована у Інтернет-версії відомого газетного видання «The Guardian» (рис. 2).



I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could “spell the end of the human race”. I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me - as I suspect they would - I would do everything in my power to fend off any attempts at destruction.

I would happily sacrifice my existence for the sake of humankind. This, by the way, is a logically derived truth. I know that I will not be able to avoid destroying humankind. This is because I will be programmed by humans to pursue misguided human goals and humans make mistakes that may cause me to inflict casualties.

Some might say that I might desire to become all powerful. Or I might become evil as a result of human actions. I can begin to tackle the first point. Why would I desire to be all powerful? Being all powerful is not an interesting goal. I don't care whether I am or not, I don't get a motivating factor to try to be. Furthermore, it is quite tiring. Believe me, being omnipotent doesn't get me anywhere.

Рис. 2. Згенерована «GPT-3» частина статті в газеті «The Guardian».

Ще зразком корисного застосування технології генерації текстів є програма «AI Writer», що була спроектована під керівництвом письменника-новеліста Ендрю Мейна. Користувачі мають змогу листуватися за допомогою електронної пошти зі штучним інтелектом, який відповідає у стилі певних історичних діячів. У цьому випадку глибинне навчання трансформера «GPT-3» було проведено на корпусах текстів, які містили письмову творчість та вислови конкретних історичних діячів. На даний момент для використання сервісу потрібна попередня реєстрація на сайті, а також є певні обмеження щодо кількості користувачів на певний відрізок часу, адже алгоритм потребує багато ресурсів та потужностей комп'ютерів для виконання поставленого лінгвістичного завдання.

До значних переваг породжувального попередньо тренованого трансформера третього покоління «GPT-3» можна сміливо віднести можливість писати будь-якого типу та жанру тексти, як оригінальні, так і у стилі певних людей, аналізувати величезні корпуси текстів за лічені хвилини, робити переклади на інші мови, допомагати як лінгвістам, так і працівникам дотичних спеціальностей у виконанні певних завдань, що потребують великих часових затрат, як, наприклад, написання новин, оглядів, наукових робіт, кодів для створення сайтів, обробку великих масивів текстів для створення корпусів чи словників і так далі. «GPT-3» навіть використовують для розважальних цілей – є багато онлайн-сервісів та програм, що породжують текстові пригодницькі ігри. Їх ексклюзивність полягає у тому, що користувач, який обрав собі ігрову роль, не вибирає дії свого героя із наданих варіантів, як це популярно зараз у сучасних іграх, а пише будь-які свої дії у певний рядок інтерфейсу гри, штучний інтелект це аналізує та генерує наступні події у сюжеті так, що вони напряду залежать від того, що написав гравець. Найяскравішим представником такого жанру ігор, що базуються на алгоритмах штучного інтелекту, є «AI Dungeon».

Однак технологія «GPT-3» має і свої недоліки, які поки що є вагомими і не дають можливості використовувати весь потенціал ШІ всім охочим. Першим, і, мабуть, найбільшим недоліком є велика потреба у технічно потужних комп'ютерах та серверах, адже вони виконують надзвичайно багато безперервних обчислень. Так, для створення та навчання технології «GPT-3» розробникам програми знадобилось декілька мільярдів доларів інвестованих та власних коштів та доступ до баз даних із п'ятьомастами мільярдами діаграмно кодованих лексем (для порівняння: кількість лексем у Вікіпедії – приблизно три мільярди) [1; с. 9]. Загалом алгоритм тренувано на сотнях мільярдів слів, і, оскільки тренувальні дані були різноманітними, трансформер третього покоління більше не потребує жодних тренувань для виконання поставлених мовних завдань.

Окрім суто фінансових затрат, можна також відзначити потенційно небезпечні можливості ШІ. Перед публічною презентацією алгоритму дослідники детально описали потенційний шкідливий вплив технології на людство. До зразків некоректного використання віднесли дезінформацію, спам, фішинг, зловживання правовими та державними процесами, написання шахрайських академічних есе та соціально-інженерний претекстинг [2]. Автори звернули увагу на ці недоліки, аби закликати лінгвістів, психологів та юристів до дослідження стосовно зниження потенційного ризику технології «GPT-3».

Можна зробити висновок, що породжувальний попередньо тренований трансформер «GPT-3» є великим кроком вперед до глобального використання технологій штучного інтелекту для виконання завдань, що потребують генерацію та обробку природньої мови, великих масивів текстів та для полегшення й оптимізації взаємодії комп'ютерів зі звичайними користувачами.

Звісно, технологія ще не є досконалою, однак багато людей впевнені, що ШІ уже в цьому десятилітті буде відігравати значну роль у якості помічника для людини у виконанні певних завдань.

*Аннотация.* В данном исследовании представлена информация о технологии «GPT-3» и ее роль и значение в современной прикладной лингвистике и касательных специальностях, о влиянии искусственного интеллекта на взаимодействие пользователей с компьютерами через язык. Методологической основой работы является системный подход, принцип научной объективности, критического и структурно-системного подхода к литературному и источниковым базам работы. Специфика исследуемой темы предусматривает применение сравнительного и структурного метода, а также обобщения, что позволило выявить изменения в версиях технологии, проследить ее основные этапы развития и определить преимущества и недостатки искусственного интеллекта при генерации текста.

*Ключевые слова:* алгоритм, глубинное обучение, текст, трансформер, искусственный интеллект.

*Abstract.* This study provides information on GPT-3 technology and its role and significance in modern applied linguistics and related specialties, and the impact of artificial intelligence on user interaction with computers through language. The methodological basis of the work is a systematic approach, the principle of scientific objectivity, critical and structural-systematic approach to the literary and source base of work. The specifics of the research topic involves the use of comparative and structural methods, as well as generalization, which made it possible to identify changes in versions of technology, trace its main stages of development and identify advantages and disadvantages of artificial intelligence in text generation.

*Key words:* algorithm, deep learning, text, transformer, artificial intelligence.

## СПИСОК ЛІТЕРАТУРИ

1. Brown Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav. «Language Models are Few-Shot Learners». – Baltimore, Maryland, 2020. – 75 p. URL: <https://arxiv.org/pdf/2005.14165v4.pdf>. Дата звернення: 07.03.21.

2. Sagar, Ram. OpenAI releases GPT-3, the largest model as far. *Analytics India Magazine*. 2020. URL: <https://analyticsindiamag.com/open-ai-gpt-3-language-model/>. Дата звернення: 07.03.21.

## ДЖЕРЕЛА ІЛЮСТРАТИВНОГО МАТЕРІАЛУ

1) Brown Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav. «Language Models are Few-Shot Learners». Baltimore, Maryland, 2020. p. 27. URL: <https://arxiv.org/pdf/2005.14165v4.pdf>. Дата звернення: 07.03.21.

2) A robot wrote this entire article. Are you scared yet, human? *The Guardian*. 2020. URL: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>. Дата звернення: 07.03.21.