

УДК 004.8:81'322

DOI <https://doi.org/10.24919/2308-4863/82-1-47>**Наталія ДЯЧУК,**

orcid.org/0000-0002-5905-6813

кандидат психологічних наук,

доцент кафедри англійської мови та прикладної лінгвістики

Житомирського державного університету імені Івана Франка

(Житомир, Україна) natadiachuk@gmail.com

МОВНІ ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ В ЗАВДАННЯХ АВТОМАТИЧНОГО АНОТУВАННЯ ТЕКСТІВ

Автоматичне анотування текстів є важливим інструментом сучасних інформаційних систем, який забезпечує ефективну обробку великих обсягів текстових даних. Актуальність цього дослідження зумовлена стрімким зростанням обсягів текстових даних у цифровому середовищі, які вимагають систематизації та аналізу, а також потребою підвищення точності й швидкості роботи з текстами. Визначено, що застосування мовних технологій штучного інтелекту, таких як трансформерні моделі BERT, GPT і XLM-R, сприяє автоматизації процесів анотування, відкриваючи нові можливості для інтеграції таких технологій у різні галузі: науку, освіту, медицину, право та бізнес.

Метою дослідження є аналіз можливостей і обмежень сучасних мовних моделей у завданнях автоматичного анотування текстів, визначення стратегічних викликів і розробка рекомендацій для їхньої інтеграції в інформаційні системи. У роботі використано методи порівняльного аналізу мовних моделей, аналізу літературних джерел і моделювання адаптації мовних технологій до практичних умов.

Доведено, що сучасні мовні моделі забезпечують високу точність і швидкість анотування, однак потребують розв'язання низки проблем: контекстуального розуміння, недостатньої адаптації до багатомовності та специфіки текстів різних галузей. Виявлено, що однією з найважливіших проблем є непрозорість алгоритмів, що ускладнює їх застосування в критично важливих сферах. Запропоновано рекомендації для підвищення ефективності мовних технологій, зокрема використання методів Explainable AI, додаткове навчання моделей на домен-специфічних текстах і розробка адаптивних інтерфейсів для користувачів.

У висновках зазначено, що впровадження мовних технологій автоматичного анотування текстів сприяє оптимізації інформаційних процесів, знижує витрати часу та підвищує ефективність роботи з даними. Перспективи подальших досліджень передбачають створення більш адаптивних моделей для багатомовного середовища, розробку алгоритмів із підвищеною прозорістю та інтеграцію цих технологій у глобальні інформаційні системи.

Ключові слова: штучний інтелект, автоматичне анотування, мовні технології, трансформерні моделі, контекстуальне розуміння, багатомовність, прозорість алгоритмів.

Nataliia DIACHUK,

orcid.org/0000-0002-5905-6813

PhD (Psychology),

Associate Professor of the Department of the English Language and Applied Linguistics

Zhytomyr Ivan Franko State University

(Zhytomyr, Ukraine) natadiachuk@gmail.com

LANGUAGE TECHNOLOGIES OF ARTIFICIAL INTELLIGENCE IN AUTOMATIC TEXT ANNOTATION TASKS

Automatic text annotation is a vital tool in modern information systems, ensuring efficient processing of large volumes of textual data. The relevance of this study is driven by the increasing amount of digital information requiring systematic analysis and the need to enhance the speed and accuracy of text processing. It has been established that applying language technologies powered by artificial intelligence, such as transformer models like BERT, GPT, and XLM-R, enables the automation of annotation processes, opening new opportunities for integrating these technologies into various fields, including science, education, medicine, law, and business.

The aim of this study is to analyze the capabilities and limitations of modern language models in addressing automatic text annotation tasks, identify key challenges, and develop recommendations for their integration into information systems. The research utilizes methods of comparative analysis of language models, literature review, and modeling the adaptation of language technologies to practical conditions.

It has been proven that modern language models deliver high accuracy and speed in annotation tasks but face challenges in contextual understanding, limited adaptation to multilingual environments, and domain-specific texts.

It was revealed that one of the primary challenges is algorithm transparency, which complicates their application in critical areas. Recommendations are proposed to enhance the efficiency of language technologies, including the use of Explainable AI approaches, additional training on domain-specific texts, and the development of adaptive user interfaces.

The conclusions emphasize that integrating automatic text annotation technologies facilitates the optimization of information processes, reduces time expenditures, and improves data processing efficiency. Prospects for further research include the creation of more adaptive models for multilingual environments, the development of algorithms with enhanced transparency, and the integration of these technologies into global information systems.

Key words: artificial intelligence, automatic annotation, language technologies, transformer models, contextual understanding, multilingualism, algorithm transparency.

Постановка проблеми. Автоматичне анотування текстів – це важливий напрям розвитку мовних технологій штучного інтелекту, орієнтований на підвищення ефективності обробки великих обсягів інформації. В умовах глобальної цифровізації стрімко зростає обсяг текстових даних, що потребують систематизації, аналізу та стислого представлення. Традиційні підходи до анотування вимагають значних часових і людських ресурсів, що обмежує можливості їхнього застосування в умовах високої динаміки інформаційних процесів. Штучний інтелект, зокрема, методи обробки природної мови, пропонує інструменти для автоматизації цих процесів, забезпечуючи точність, швидкість і адаптивність до різних мов і тематик.

Автоматичне анотування текстів є важливим інструментом для розв'язання актуальних наукових і практичних завдань. Наукові аспекти стосуються розробки алгоритмів, здатних урахувати контекстуальну семантику, багатомовність і специфіку текстів різних доменів. Практичні завдання зосереджені на інтеграції цих алгоритмів в інформаційні системи, що використовуються в науковій, освітній, юридичній та інших сферах. Такі технології сприяють підвищенню ефективності аналізу даних, автоматизації рутинних завдань та розширенню можливостей роботи з великими обсягами текстової інформації.

Отже, автоматичне анотування текстів є не лише інноваційним завданням штучного інтелекту, але й важливим кроком до оптимізації інформаційних процесів у сучасному суспільстві.

Аналіз останніх досліджень і публікацій. Управління соціокультурними інноваціями через призму медіа та філософії культури дає змогу зрозуміти, як медійні технології формують нові культурні процеси та впливають на суспільну свідомість.

Н. А. Цимбал розглядає роль прикладної лінгвістики у впровадженні мовних технологій, підкреслюючи їхнє значення для систематизації текстів у наукових і освітніх контекстах (Цимбал, 2019). Подібні питання аналізують О. В. Бармак, О. В. Мазурець та А. В. Живілік, які запропонували модель автоматизованого анотування тек-

стів, що значно підвищує точність стислих текстових резюме за допомогою алгоритмів машинного навчання (Бармак, Мазурець, Живілік, 2017). О. В. Захарова дослідила семантичне анотування великих даних. Науковиця довела, що аналіз великих обсягів інформації сприяє оптимізації аналітичних процесів (Захарова, 2020).

Т. П. Голуб, О. О. Коваленко, О. І. Назаренко, Л. М. Жигжитова висвітлюють підходи до автоматичного реферування текстів, наголошуючи на необхідності адаптації моделей до специфічних галузевих умов (Golub, Kovalenko, Nazarenko, Zhygzhytova, 2021). М. О. Фант порівнює неконтрольовані методи оцінювання екстрактивних анотацій, акцентуючи на їхніх перевагах у випадках, коли попередня розмітка даних відсутня (Фант, 2024).

L. C. Cante, S. D'Angelo та B. Di Martino здійснили огляд інструментів для анотування текстів, порівнюючи їхні можливості та обмеження (Cante, D'Angelo, Di Martino, 2024). D. Baviskar, S. Ahirrao, V. Potdar, K. Kotecha акцентують на автоматизованій обробці неструктурованих документів за допомогою штучного інтелекту, підкреслюючи його здатність ефективно працювати з великими обсягами складних текстових структур (Baviskar, Ahirrao, Potdar, Kotecha, 2021).

S. P. PATTYAM аналізує методи обробки природної мови для автоматизації текстового аналізу, зокрема системи для виявлення семантичних структур (Pattyam, 2021). Натомість Н. Bashiri, Н. Naderi у своєму критичному аналізі трансформерних моделей підкреслюють їхню ефективність у завданнях анотування текстів (Bashiri, Naderi, 2024).

C.-M. Chen, Y.-T. Chen, C.-Y. Liu (Chen, Chen, Liu, 2019) досліджують створення автоматизованих систем для текстового анотування в цифрових гуманітарних дослідженнях, указуючи на їхню ефективність у підвищенні продуктивності дослідників (Chen, Chen, Liu, 2019). C. Zong, R. Xia, J. Zhang (Zong, Xia, Zhang, 2021) пропонують систематизацію підходів до текстового майнінгу, демонструючи переваги інтеграції автоматичного анотування в процеси обробки текстів (Zong, Xia, Zhang, 2021).

W. S. El-Kassas, C. R. Salama, A. A. Rafea, A. E. Hassanien аналізують методи автоматичного реферування, зокрема з використанням нейронних мереж, які забезпечують високу його точність для великих текстових масивів (El-Kassas, Salama, Rafea, Hassanien, 2021). W. Kryściński, B. McCann, C. Xiong, R. Socher у своїй роботі оцінюють нейронні моделі текстового реферування, акцентуючи на можливостях їхньої адаптації до конкретних завдань (Kryściński, McCann, Xiong, Socher, 2019).

R. G. Goriparthi досліджує багатомовне текстове реферування та переклад, демонструючи значення інструментів штучного інтелекту для роботи з текстами в глобальному контексті (Goriparthi, 2021).

Аналіз цих робіт свідчить про значний прогрес у сфері мовних технологій, орієнтованих на автоматичне анування текстів. Основні результати показують, що застосування штучного інтелекту сприяє підвищенню точності, швидкості та ефективності текстової обробки, що є визначальним фактором розвитку сучасних інформаційних систем.

Попри значні досягнення в дослідженні автоматичного анування текстів, окремі аспекти, зокрема його концептуальні засади та роль у сучасних інформаційних процесах, залишаються недостатньо розкритими. Більшість наукових робіт зосереджуються на технічних аспектах, тоді як питання інтеграції цих технологій у системи управління інформацією досі опрацьовані фрагментарно, що створює значні перешкоди для їх ефективного практичного застосування.

Дослідження ефективності мовних моделей, зокрема трансформерів (BERT, GPT), також потребує додаткової уваги. Відсутність порівняльного аналізу продуктивності цих моделей за різних обставин, зокрема в багатомовних середовищах та для специфічних доменів, значно обмежує їхню здатність до адаптації в реальних практичних умовах.

Адаптація до багатомовності та домен-специфічних текстів залишається викликом, особливо для мов із низьким ресурсом чи вузькоспеціалізованих галузей. Сучасні підходи, такі як донавчання, часто не дають необхідної точності, що підкреслює потребу в нових методах.

Контекстуальність і прозорість алгоритмів є ще однією нерозв'язаною проблемою. Сучасні моделі часто працюють як «чорні ящики», що знижує довіру до їхніх результатів і обмежує застосування в критичних галузях.

Пропоноване дослідження спрямоване на подолання цих викликів через розроблення нових

підходів до адаптації моделей, підвищення прозорості алгоритмів за допомогою Explainable AI та створення ефективних методів інтеграції мовних технологій в інформаційні системи. Це забезпечить розширення можливостей автоматичного анування текстів та його впровадження в різних галузях.

Мета статті – проаналізувати можливості й обмеження мовних технологій штучного інтелекту в розв'язанні завдань автоматичного анування текстів та запропонувати рекомендації щодо їхнього удосконалення й застосування в сучасних інформаційних системах.

Завдання статті:

1. Описати концептуальні основи автоматичного анування текстів і проаналізувати ефективність сучасних мовних моделей.
2. Дослідити адаптацію мовних технологій до багатомовного середовища та домен-специфічних текстів.
3. Визначити основні виклики та розробити рекомендації для інтеграції мовних технологій в інформаційні системи.

Виклад основного матеріалу. Автоматичне анування текстів – основна частина сучасних мовних технологій штучного інтелекту, спрямованих на оптимізацію обробки великих обсягів інформації (Бармак, Мазурець, Живілік, 2017). Цей процес полягає у створенні короткого опису основних ідей або змісту тексту без втрати контекстуальної важливості, що забезпечує зручний доступ до інформації в умовах стрімкого зростання обсягів текстових даних. Завдяки використанню алгоритмів обробки природної мови автоматичне анування сприяє зменшенню витрат часу, оптимізації робочих процесів та підвищенню ефективності пошуку релевантної інформації (табл. 1).

На практиці автоматичне анування текстів уже активно застосовується в наукових бібліотеках, інформаційних системах та цифрових архівах. Воно значно спрощує управління великими обсягами інформації, даючи можливість автоматизувати завдання, які раніше вимагали участі експертів. Наприклад, системи автоматичного анування на основі трансформерів успішно використовуються в науковій сфері для прискорення аналізу літератури, у юридичній галузі – для обробки та систематизації документів, а також у комерційних цілях – для структурування відгуків клієнтів і коментарів у соціальних мережах.

Важливість цієї технології зумовлена її здатністю значно скорочувати час на обробку текстів, підвищувати точність роботи з інформацією та забезпечувати доступність для багатомовного

Порівняння характеристик традиційного та автоматичного анотування текстів

Характеристика	Традиційне анотування	Автоматичне анотування
Часові витрати	Значні, залежать від обсягу тексту та кваліфікації експерта	Мінімальні, виконуються швидко, незалежно від обсягу тексту
Якість результатів	Залежить від експертного рівня, можливі суб'єктивні помилки	Забезпечується алгоритмом, можливі похибки через недостатнє врахування контексту
Адаптація до нових тем і мов	Вимагає спеціалізованої підготовки експертів	Залежить від навчання моделі на відповідних наборах даних
Масштабованість процесу	Обмежена через людський ресурс	Висока, забезпечує одночасну обробку великих обсягів текстової інформації
Контекстуальне розуміння	Може бути глибоким, залежить від досвіду експерта	Частково враховується за допомогою моделей типу BERT чи GPT

Джерело: власна розробка автора

аналізу. Зміни в умовах сучасного інформаційного середовища, зокрема зростання вимог до швидкості аналізу даних, інтеграція великих мовних моделей і потреба в адаптивних системах, зробили автоматичне анотування незамінним інструментом у багатьох галузях.

Автоматичне анотування текстів значно залежить від використання сучасних мовних моделей, які базуються на алгоритмах глибокого навчання. Ці моделі сприяють забезпеченню високої точності й контекстуального розуміння текстів. Найбільш значущими досягненнями в цій галузі є трансформерні архітектури, зокрема моделі BERT (Bidirectional Encoder Representations from Transformers) та GPT (Generative Pre-trained Transformer), які показали високу ефективність у завданнях, пов'язаних із класифікацією, резюмуванням та анотуванням текстів. Завдяки здатності обробляти великі обсяги даних і враховувати семантичний контекст, ці моделі забезпечують глибше розуміння тексту, ніж традиційні алгоритми (табл. 2).

У сучасних умовах ці мовні моделі широко використовуються для автоматичного анотування текстів у різних сферах. Наприклад, BERT активно застосовується в наукових базах даних для автоматичної класифікації статей за тематикою, що значно спрощує роботу дослідників. GPT-3 демонструє виняткову ефективність у створенні стислих і точних анотацій для новинних статей та юридичних документів, що особливо корисно в умовах, коли обсяги інформації перевищують можливості людини.

Практичні приклади стосуються використання T5 для анотування мультимовних текстів у бібліотеках та цифрових архівах, що дає можливість забезпечувати доступ до знань користувачам різних мовних груп. У комерційних цілях DistilBERT

часто інтегрується в мобільні додатки та CRM-системи, де важлива швидка обробка тексту без значних обчислювальних ресурсів. RoBERTa використовується у фінансових системах для аналізу великих обсягів текстових даних, наприклад, у звітах та аналітиці, що вимагає глибокого контекстуального розуміння (Cante, D'Angelo, Di Martino, 2024).

Використання цих моделей не лише підвищує ефективність анотування текстів, але й відкриває можливості для автоматизації інформаційних процесів у глобальних масштабах, мінімізуючи людські ресурси та забезпечуючи точність і швидкість роботи з даними.

Адаптація мовних технологій до багатомовного середовища та домен-специфічних текстів є одним з основних викликів у сучасній обробці природної мови. Використання універсальних моделей, таких як BERT або GPT, не завжди забезпечує необхідну точність у контексті специфічних мов або професійних доменів, оскільки вони можуть бути недостатньо навчені на даних, що відображають ці особливості. Багатомовні моделі, такі як mBERT або XLM-R, здатні працювати з декількома мовами, але їхня продуктивність часто залежить від наявності великих корпусів текстів для навчання кожною мовою. У домен-специфічних текстах складність зростає через використання вузькоспеціалізованої термінології та стилістичних особливостей, що потребує додаткової адаптації або створення нових моделей (табл. 3).

На практиці багатомовні моделі, такі як mBERT і XLM-R, застосовуються в глобальних інформаційних системах для автоматизації анотування та аналізу текстів. Наприклад, XLM-R активно використовується для створення багатомовних баз знань у наукових бібліотеках, забезпечуючи рівний доступ до інформації для різних мовних груп

Таблиця 2

Ефективність сучасних мовних моделей у завданнях автоматичного анотування текстів

Модель	Основні характеристики	Результати в завданнях анотування
BERT	Двоспрямований аналіз тексту, що враховує контекст до й після слова	Висока точність у визначенні основних тем і розумінні семантики тексту
GPT-3	Генеративна модель, здатна створювати нові анотації на основі аналізу попередніх текстів	Висока якість автоматично створених анотацій, які за характеристиками максимально наближені до зразків, створених людиною
T5 (Text-to-Text Transfer Transformer)	Універсальна модель, яка перетворює всі NLP завдання в текстовий формат, зокрема анотування	Здатність виконувати анотування для багатомовних текстів із високою адаптивністю
DistilBERT	Легка версія BERT із меншою кількістю параметрів для швидшої обробки тексту без значного зниження точності	Оптимальний вибір для реальних застосунків з обмеженими ресурсами
RoBERTa	Покращена версія BERT із розширеним набором даних для навчання та більшою увагою до семантики тексту	Забезпечує найвищу точність у контекстуальному анотуванні текстів

Джерело: сформовано автором на підставі [Цимбал, 2019; Bashiri, Naderi, 2024; Cante, D'Angelo, Di Martino, 2024; Langston, Ashford, 2024]

Таблиця 3

Методи й моделі адаптації мовних технологій до багатомовного середовища та домен-специфічних текстів

Модель/Метод	Особливості адаптації до багатомовності	Особливості адаптації до доменів
mBERT	Універсальна підтримка понад 100 мов, враховує загальні закономірності між мовами	Вимагає додаткового донавчання на домен-специфічних текстах для підвищення точності
XLM-R (XLM-RoBERTa)	Покращена багатомовна модель із глибоким контекстуальним розумінням	Забезпечує адаптацію до доменів через використання великих попередньо підготовлених корпусів текстів
Fine-tuning на домен-специфічних корпусах	Навчання базової моделі на корпусах, що включають специфічну термінологію та мовні конструкції	Ефективно адаптує модель до вузькоспеціалізованих завдань, таких як юридичні або медичні тексти
Translation Alignment	Використання перекладів для створення узгоджених багатомовних корпусів	Корисно для адаптації до доменів із доступними багатомовними ресурсами, але обмежено специфічністю перекладів
Custom Tokenizers	Модифікація токенизаторів для роботи з мовами зі складними морфологічними структурами	Покращує якість аналізу текстів у специфічних доменах, враховуючи особливості термінів і синтаксису

Джерело: власна розробка автора

(Baviskar, Ahirrao, Potdar, Kotecha, 2021). Модель Fine-tuning є стандартним підходом у медичній сфері, де навчання на спеціалізованих корпусах, що містять медичну термінологію, сприяє підвищенню точності діагностичних систем (Zong, Xia, Zhang, 2021).

Особливості домен-специфічної адаптації на практиці добре ілюструються в юридичній галузі, де створюються кастомізовані токенизатори для обробки складних юридичних конструкцій.

Translation Alignment часто використовується в міжнародних організаціях для забезпечення семантичної відповідності текстів різними мовами. Такі підходи, хоча й залежать від якості навчальних даних, демонструють високу ефективність у складних багатомовних і домен-специфічних завданнях, сприяючи глобалізації інформаційних процесів і інтеграції вузькоспеціалізованих знань.

Застосування штучного інтелекту в автоматичному анотуванні текстів постає перед низкою

викликів, які обмежують ефективність та надійність таких технологій у реальних умовах. Одним із найважливіших є забезпечення контекстуальності анотацій. Традиційні алгоритми часто працюють із текстами фрагментарно, що призводить до втрати змістовних зв'язків між їхніми елементами. Навіть сучасні трансформерні моделі, такі як BERT або GPT, не завжди можуть повністю врахувати приховані семантичні зв'язки, особливо в складних текстах із неоднозначною термінологією чи стилістикою.

Багатомовність є ще одним викликом, оскільки моделі, що використовуються для автоматичного анотування, здебільшого розробляються на основі англійських корпусів. Це призводить до зниження точності для текстів іншими мовами, особливо для тих, що мають складну морфологію, унікальні синтаксичні конструкції або недостатньо представлені в навчальних наборах даних. У таких випадках необхідність донавчання моделей на специфічних мовних корпусах створює додаткове навантаження на обчислювальні ресурси та потребує більше часу.

Прозорість алгоритмів є важливою проблемою, оскільки сучасні моделі глибокого навчання часто діють як «чорні ящики», де неможливо чітко зрозуміти, як алгоритм ухвалює певні рішення. Це особливо критично у випадках, коли автоматичні анотації використовуються для прийняття важливих рішень, наприклад у юридичних чи медичних текстах. Недостатня пояснюваність результатів може призводити до недовіри користувачів і обмежувати можливості впровадження таких систем у важливих сферах (Kryściński, McCann, Xiong, Socher, 2019).

З огляду на сучасні реалії, ці виклики значно ускладнюють інтеграцію автоматичного анотування текстів у глобальні інформаційні системи. Цей процес потребує комплексного підходу, що враховує технічні, організаційні та користувацькі аспекти (El-Kassas, Salama, Rafea, Hassanien, 2021). Одним з основних етапів є вибір мовної моделі, що відповідає специфіці завдань. Наприклад, для багатомовних систем доцільно використовувати моделі на зразок XLM-R або mBERT, які забезпечують високу точність анотування текстів різними мовами. Для домен-специфічних систем необхідно здійснювати додаткове навчання моделей на корпусах даних, що містять релевантну термінологію та стилістичні особливості.

Не менш важливим є створення адаптивного інтерфейсу, який дозволить користувачам контролювати результати автоматичного анотування та за необхідності вносити корективи. Це спри-

ятиме підвищенню довіри до системи та забезпечить гнучкість її використання. Для інтеграції в наявні інформаційні системи рекомендується використовувати API-рішення (API – Application Programming Interface – інтерфейс програмування додатків, програмний інтерфейс програми), які забезпечують сумісність мовних технологій із різними платформами та полегшують їхнє оновлення.

Важливо також передбачити методи для підвищення прозорості алгоритмів. Використання підходів Explainable AI дає можливість роз'яснювати, яким чином модель ухвалює рішення, що особливо важливо для критичних галузей, таких як медицина, право чи фінанси. Крім того, варто забезпечити регулярний моніторинг продуктивності інтегрованих моделей для вчасного оновлення алгоритмів та усунення можливих помилок.

На практиці такі рекомендації вже успішно впроваджуються в автоматизованих бібліотечних системах, системах управління контентом та цифрових архівах. Наприклад, інтеграція BERT у наукові бази даних дозволила значно спростити пошук і класифікацію статей, а використання API-рішень на основі GPT забезпечило адаптацію текстових аналізаторів у багатомовних корпоративних системах. Розробка та впровадження подібних технологій значно підвищують ефективність інформаційних систем, забезпечуючи автоматизацію рутинних завдань, оптимізацію часу на пошук даних та гнучкість у роботі з різними типами текстів.

Висновки. Здійснене нами дослідження дозволило дійти висновків, згідно з якими автоматичне анотування текстів є важливим етапом розвитку мовних технологій штучного інтелекту, здатним оптимізувати обробку великих обсягів текстової інформації. Проаналізовано сучасні мовні моделі, такі як BERT, GPT, XLM-R, та їхню адаптацію до багатомовних середовищ і домен-специфічних текстів. Визначено, що основними проблемами є забезпечення контекстуальності анотацій, недостатня продуктивність моделей у багатомовному контексті через обмеженість навчальних даних, а також непрозорість алгоритмів, що ускладнює їхнє впровадження в критичних сферах.

Запропоновано рекомендації для інтеграції мовних технологій у сучасні інформаційні системи, серед яких вибір адаптованих моделей, використання API для спрощення інтеграції, створення адаптивних інтерфейсів для взаємодії з користувачами та впровадження підходів Explainable AI для підвищення прозорості алгоритмів.

Перспективи подальших досліджень передбачають розроблення моделей, здатних ефек-

тивніше працювати з багатомовними текстами та враховувати специфіку вузькоспеціалізованих доменів. Додаткової уваги потребує інтеграція механізмів, які забезпечують пояснюваність результатів роботи моделей, що сприятиме ширшому використанню технологій автоматичного анотування в різних сферах діяльності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бармак О. В., Мазурець О. В., Живілік А. В. Інформаційна технологія автоматизованого анотування та реферування цифрових текстів. *Вісник Хмельницького національного університету. Технічні науки*. 2017. № 4. С. 147–158. URL: <https://elar.khmnu.edu.ua/server/api/core/bitstreams/c985cd37-019d-406b-bddd-f2829284638a/content> (дата звернення: 22.09.2024).
2. Захарова О. В. Основні аспекти семантичного анотування великих даних. *Проблеми програмування*. 2020. № 4. С. 22–33. URL: <http://dspace.nbuv.gov.ua/handle/123456789/180491> (дата звернення: 22.09.2024).
3. Фант М. О. Порівняльний аналіз неконтрольованих методів оцінки екстрактивних анотацій. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2024. Т. 35 (74). № 2. С. 190–195. URL: https://www.tech.vernadskyjournals.in.ua/journals/2024/2_2024/2_2024.pdf#page=200 (дата звернення: 22.09.2024).
4. Цимбал Н. А. *Прикладна лінгвістика: навчальний посібник*. Умань: Візаві, 2019. 106 с. URL: <https://dspace.udpu.edu.ua/handle/123456789/11809> (дата звернення: 22.09.2024).
5. Bashiri H., Naderi H. Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*. 2024. Vol. 66. P. 7305–7361. DOI: <https://doi.org/10.1007/s10115-024-02214-3>
6. Bavisar D., Ahirrao S., Potdar V., Kotecha K. Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. *IEEE Access*. 2021. Vol. 9. P. 72894–72936. DOI: <https://doi.org/10.1109/ACCESS.2021.3072900>
7. Cante L. C., D'Angelo S., Di Martino B. Text Annotation Tools: A Comprehensive Review and Comparative Analysis. In: *Complex, Intelligent and Software Intensive Systems: Proceedings of the 18th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2024)*. Springer Nature, 2024. P. 353–363.
8. Chen C.-M., Chen Y.-T., Liu C.-Y. Development and evaluation of an automatic text annotation system for supporting digital humanities research. *Library Hi Tech*. 2019. Vol. 37. № 3. P. 436–455. DOI: <https://doi.org/10.1108/LHT-10-2017-0219>
9. El-Kassas W. S., Salama C. R., Rafea A. A., Hassanien A. E. Automatic Text Summarization: A Comprehensive Survey. *Expert Systems with Applications*. 2021. Vol. 165. Art. 113679. DOI: <https://doi.org/10.1016/j.eswa.2020.113679>
10. Golub T. P., Kovalenko O. O., Nazarenko O. I., Zhygzyhtova L. M. Approaches to automatic summarization and annotation. *Науковий вісник Дрогобицького державного педагогічного університету імені Івана Франка. Серія: Філологічні науки (мовознавство)*. 2021. № 16. С. 13–17. URL: http://ddpu-filolvisnyk.com.ua/uploads/arkhiv-nomerov/2021/NV_2021_16/2.pdf (дата звернення: 22.09.2024).
11. Goriparthi R. G. AI-Driven Natural Language Processing for Multilingual Text Summarization and Translation. *Revista de Inteligencia Artificial en Medicina*. 2021. Vol. 12. № 1. P. 513–535. URL: <https://redcrevistas.com/index.php/Revista/article/view/226> (дата звернення: 22.09.2024).
12. Kryściński W., McCann B., Xiong C., Socher R. Neural Text Summarization: A Critical Evaluation. *arXiv preprint*. 2019. Art. arXiv:1908.08960. URL: <https://arxiv.org/abs/1908.08960> (дата звернення: 22.09.2024).
13. Langston O., Ashford B. Automated Summarization of Multiple Document Abstracts and Contents Using Large Language Models. *TechRxiv*. 2024. DOI: <https://doi.org/10.36227/techrxiv.172262754.45577350/v1>
14. Pattayam S. P. AI-Enhanced Natural Language Processing: Techniques for Automated Text Analysis, Sentiment Detection, and Conversational Agents. *Journal of Artificial Intelligence Research and Applications*. 2021. Vol. 1. № 1. P. 371–406. URL: <https://aimlstudies.co.uk/index.php/jaira/article/view/225> (дата звернення: 22.09.2024).
15. Zong C., Xia R., Zhang J. *Text Data Mining*. Singapore: Springer, 2021. URL: <https://link.springer.com/book/10.1007/978-981-16-0100-2> (дата звернення: 22.09.2024).

REFERENCES

1. Barmak, O. V., Mazurets, O. V., & Zhivlyuk, A. V. (2017) Informatsiina tekhnolohiia avtomatyzovanoho anotovannia ta referuvannia tsyfrovyykh tekstiv [Information technology of automated annotation and summarization of digital texts]. *Visnyk Khmelnytskoho natsionalnogo universytetu. Tekhnichni nauky – Bulletin of Khmelnytskyi National University. Technical Sciences*, 4, 147 – 158. Retrieved from <https://elar.khmnu.edu.ua/server/api/core/bitstreams/c985cd37-019d-406b-bddd-f2829284638a/content> [in Ukrainian].
2. Zakharova, O. V. (2020) Osnovni aspekty semantychnoho anotovannia velykykh danykh [Key aspects of semantic annotation of big data]. *Problemy prohramuvannia – Problems of Programming*, 4, 22–33. Retrieved from <http://dspace.nbuv.gov.ua/handle/123456789/180491> [in Ukrainian].
3. Fant, M. O. (2024) Porivnialnyi analiz nekontrolovanykh metodiv otsinky ekstraktyvnykh anotatsii [Comparative analysis of unsupervised methods for evaluating extractive annotations]. *Vcheni zapysky TNU imeni V. I. Vernadskoho. Seriya: Tekhnichni nauky – Scientific Notes of TNU Named After V. I. Vernadsky. Series: Technical Sciences*, 35(74), 190–195. Retrieved from https://www.tech.vernadskyjournals.in.ua/journals/2024/2_2024/2_2024.pdf#page=200 [in Ukrainian].
4. Tsymbal, N. A. (2019) *Prykladna linhvistyka: navchalnyi posibnyk* [Applied Linguistics: A Textbook]. Uman: Vizavi. Retrieved from <https://dspace.udpu.edu.ua/handle/123456789/11809> [in Ukrainian].
5. Bashiri, H., & Naderi, H. (2024) Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, 66, 7305–7361. DOI: <https://doi.org/10.1007/s10115-024-02214-3>

6. Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021) Efficient automated processing of unstructured documents using artificial intelligence: A systematic literature review and future directions. *IEEE Access*, 9. 72894–72936. DOI: <https://doi.org/10.1109/ACCESS.2021.3072900>
7. Cante, L. C., D'Angelo, S., & Di Martino, B. (2024) Text annotation tools: A comprehensive review and comparative analysis. In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 18th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2024)* (pp. 353–363). Springer Nature.
8. Chen, C.-M., Chen, Y.-T., & Liu, C.-Y. (2019) Development and evaluation of an automatic text annotation system for supporting digital humanities research. *Library Hi Tech*, 37(3). 436–455. DOI: <https://doi.org/10.1108/LHT-10-2017-0219>
9. El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Hassanien, A. E. (2021) Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165. 113679. DOI: <https://doi.org/10.1016/j.eswa.2020.113679>
10. Golub, T. P., Kovalenko, O. O., Nazarenko, O. I., & Zhygzhytova, L. M. (2021) Approaches to automatic summarization and annotation. *Naukovi visnyk Drohobyskoho derzhavnoho pedahohichnoho universytetu imeni Ivana Franka. Serii: Filolohichni nauky (movoznavstvo) – Scientific Bulletin of Drohobych State Pedagogical University Named After Ivan Franko. Series: Philological Sciences (Linguistics)*, 16. 13–17. Retrieved from http://ddpu-filolvisnyk.com.ua/uploads/arkhiv-nomerov/2021/NV_2021_16/2.pdf
11. Goriparthi, R. G. (2021) AI-driven natural language processing for multilingual text summarization and translation. *Revista de Inteligencia Artificial en Medicina*, 12(1). 513–535. Retrieved from <https://redcrevistas.com/index.php/Revista/article/view/226>
12. Kryściński, W., McCann, B., Xiong, C., & Socher, R. (2019) Neural text summarization: A critical evaluation. *arXiv preprint. arXiv:1908.08960*. Retrieved from <https://arxiv.org/abs/1908.08960>
13. Langston, O., & Ashford, B. (2024) Automated summarization of multiple document abstracts and contents using large language models. *TechRxiv*. DOI: <https://doi.org/10.36227/techrxiv.172262754.45577350/v1>
14. Pattayam, S. P. (2021) AI-enhanced natural language processing: Techniques for automated text analysis, sentiment detection, and conversational agents. *Journal of Artificial Intelligence Research and Applications*, 1(1). 371–406. Retrieved from <https://aimlstudies.co.uk/index.php/jaira/article/view/225>
15. Zong, C., Xia, R., & Zhang, J. (2021) *Text data mining*. Singapore: Springer. Retrieved from <https://link.springer.com/book/10.1007/978-981-16-0100-2>