

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)Кафедра Інформатики
(повна назва)Рівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«____» _____ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУстудентові Стебаєву Ігорю Миколайовичу
(прізвище, ім'я, по батькові)1. Тема роботи Дослідження методів створення системи багатомовного перекладу

затверджена наказом по університету від 3 листопада 2023 року № 1280Ст

2. Термін подання студентом роботи до екзаменаційної комісії 25 грудня 2023 р.3. Вихідні дані до роботи науково-методична та науково-технічна література, матеріали конференцій, дані інтернет-мережі, бібліотека OpenCV, мова програмування Python, середовище розроблення Notebook.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Огляд основних методів створення системи багатомовного перекладу2. Комп'ютерна модель «Encoder-Decoder». Дослідження методів3. Протестувати розроблений застосунок та провести аналіз результатів.4. Виявити перспективи подальшої роботи.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Актуальність дослідження, об'єкт та мета дослідження, постановка задачі дослідження, вихідні дані дослідження, етапи розроблення, результат тестування, висновки, перспективи подальших досліджень, апробація результатів роботи.

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	03.11.2023	
2	Аналіз завдання, підбір літератури	03.11.23-05.11.23	
3	Аналіз літератури з досліджуваної проблеми	06.11.23-09.11.23	
4	Аналіз моделі «Encoder-Decoder»	10.11.23-21.11.23	
5	Розробка методів моделі «Encoder-Decoder»	22.12.23-01.12.23	
6	Програмна реалізація	02.12.23-08.12.23	
7	Оформлення пояснювальної записки	09.12.23-10.12.23	
8	Перевірка на плагіат	11.12.2023	
9	Рецензування	15.12.2023	
10	Підготовка презентації та доповіді	17.12.2023	
11	Занесення роботи в електронний архів	25.12.2023	
12	Попередній захист кваліфікаційної роботи	04.01.2024	

Дата видачі завдання 3 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

проф. Кузьомін О. Я.
(посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 63 с., 1 табл., 56 джерел.

ВЕЛИКОМОВНА МОДЕЛЬ, МЕТОДИ ДЛЯ БАГАТОМОВНОГО ПЕРЕКЛАДУ, НЕЙРОННИЙ МАШИННИЙ ПЕРЕКЛАД (NMT).

Об'єктом дослідження є методи для багатомовного перекладу з використанням штучного інтелекту.

Метою роботи є дослідження технологій та алгоритмів для багатомовного перекладу з використанням штучного інтелекту, які дозволяють оптимізувати процес перекладу.

Використано методи числового моделювання та аналітичного обґрунтування. Проведено дослідження та аналіз методів для перекладу української мови. Досліджено методи «Збір і підготовка даних», «Нейронний машинний переклад (NMT)», «Архітектура моделі», «Навчання», «Точне налаштування», «Оцінка та показники» та інші. Розроблено алгоритм перекладу української мови.

У результаті дослідження здійснена програмна реалізація системи для багатомовного перекладу з використанням штучного інтелекту.

LARGE LANGUAGE MODEL, METHODS FOR MULTILINGUAL TRANSLATION, NEURAL MACHINE TRANSLATION (NMT)..

The object of research is methods for multilingual translating using artificial intelligence.

The purpose of the work is the research of technologies and algorithms for multilingual translation using artificial intelligence, which allow to optimize the translation process.

Numerical modeling and analytical reasoning methods were used. A study and analysis of methods for the translation of the Ukrainian language was carried out. The methods "Collection and preparation of data", "Neural machine translation (NMT)", "Model architecture", "Learning", "Fine tuning", "Evaluation and indicators" and others were studied. The Ukrainian language translation algorithm has been developed.

As a result of the research, the software implementation of the system for the multilingual translation using artificial intelligence was carried out.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Огляд основних методів створення системи багатомовного перекладу ...	10
1.1 Розвиток великомовної моделі	10
1.2 Основні тези дослідження великомовної моделі.....	12
1.3 Основні методи дослідження великомовної моделі.....	13
1.4 Аналіз літературних джерел	15
1.5 Постановка задачі дослідження.....	21
2 Модель «Encoder-Decoder». Дослідження методів.....	22
2.1 Модель «Encoder-Decoder»	22
2.2 Збір і підготовка даних (Data Collection and Preparation)	23
2.3 Нейронний машинний переклад (NMT).....	24
2.4 Ахітектура моделі (Model Architecture).....	26
2.5 Навчання (Training).....	27
2.6 Точне налаштування (Fine-Tuning)	28
2.7 Оцінка та показники (Evaluation and Metrics).....	29
2.8 Постредагування (Post-Editing)	30
2.9 Постійне вдосконалення (Continuous Improvement).....	32
2.10 Масштабованість і розгортання (Scalability and Deployment).....	33
2.11 Етичні міркування (Ethical Considerations)	34
2.12 Відгуки користувачів (User Feedback)	35
2.13 Обмеження великомовної моделі.....	36
2.14 Переваги та недоліки розглянутих методів	37
3 Комп'ютерна модель для багатомовного перекладу.....	41
3.1 Обґрунтування вибору середовища програмної реалізації	41
3.2 Програмна реалізація.....	45
3.3 Інструкція користувача	49
3.4 Тестування та аналіз розробленої моделі.....	50

	6
Висновки	53
Перелік джерел посилання	57

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ**

NMT – Neural Machine Translation (нейронний машинний переклад)

ШІ – штучний інтелект

LLM – Large Language Model (велика мовна модель)

NLP – Neuro-Linguistic Programming (Нейролінгвістичне програмування)

ВСТУП

Тема дослідження «Дослідження методів створення системи багатомовного перекладу» належить до сфери машинного перекладу та штучного інтелекту. Ця тема є актуальною і важливою в сучасному світі, оскільки переклад є необхідною складовою спілкування між різними мовними спільнотами, а великомовні моделі відкривають нові можливості для автоматизації та поліпшення якості перекладу.

Зростання глобалізації та міжнародного співробітництва робить мовний переклад необхідним для бізнесу, науки, освіти та культурного обміну. Великомовні моделі, засновані на штучному інтелекті, можуть ефективно вирішувати завдання перекладу та забезпечувати доступ до інформації для різних аудиторій та мовних груп [1–4].

Проте ця тема також породжує питання етики, конфіденційності та якості перекладу, які потребують детального дослідження та розв'язання. Дослідження великомовних моделей для перекладу української мови ставить перед собою завдання покращити доступність перекладу та забезпечити його якість у відповідності з вимогами сучасного світу [5, 6].

Дослідження великомовної моделі для багатомовного перекладу з використанням штучного інтелекту має велику актуальність з кількох ключових причин [7]:

– зростаючий попит на переклади: Україна має активну економіку та багато контактів з іноземними партнерами, що призводить до зростаючої потреби в перекладах. Переклади необхідні в різних сферах, від бізнесу та науки до медицини та культури [8];

– штучний інтелект і машинне навчання: розвиток штучного інтелекту та машинного навчання в останні роки привів до можливостей побудови високоефективних моделей перекладу, які можуть покращити точність та швидкість перекладів;

– зростаюча складність мови: українська мова має свою власну складну граматику та багатий лексикон. Переклад на українську мову теж вимагає глибокого розуміння мови. Великомовні моделі можуть допомогти покращити якість перекладів [9];

– міжнародні комунікації: зростаюча глобалізація та міжнародні відносини вимагають високоякісних перекладів для забезпечення ефективної комунікації між різними культурами і мовами;

– підтримка інтернет–контенту: з кожним днем збільшується кількість інтернет–контенту, який потребує перекладу на різні мови, включаючи українську. Великомовні моделі можуть полегшити цей процес;

– доступність технологій: завдяки розвитку технологій, великомовні моделі стають більш доступними для різних організацій і дослідників, що підвищує їхню актуальність та реалізовує їхній потенціал [10, 11].

Враховуючи ці фактори, дослідження та розробка великомовних моделей для перекладу української мови є актуальним та важливим завданням, яке може призвести до покращення доступності та якості перекладів, а також сприяти співпраці між країнами та культурами.

1 ОГЛЯД ОСНОВНИХ МЕТОДІВ СТВОРЕННЯ СИСТЕМИ БАГАТОМОВНОГО ПЕРЕКЛАДУ

1.1 Розвиток великомовної моделі

Дослідження методів створення систем багатомовного перекладу є важливою областю машинного перекладу і комп'ютерної лінгвістики. Такі системи дозволяють автоматично перекладати текст з однієї мови на іншу, іноді використовуючи декілька проміжних мов для досягнення кращої якості перекладу. Ось декілька основних методів і підходів до створення систем багатомовного перекладу [7]:

– методи статистичного машинного перекладу: цей підхід базується на аналізі великої кількості паралельних текстів у різних мовах для навчання моделей перекладу. Статистичні моделі, такі як модель фразового перекладу та модель машинного перекладу з нейронними мережами (NMT), можуть бути розширені для багатомовного перекладу;

– методи на основі правил: цей підхід використовує правила, граматичні структури та лінгвістичні знання для перекладу тексту. Системи багатомовного перекладу, побудовані на основі правил, можуть бути складними і вимагати значного напрацювання;

– методи глибокого навчання: з впровадженням нейронних мереж і глибокого навчання досягли великого успіху в машинному перекладі. Моделі NMT можуть бути розширені для багатомовного перекладу, використовуючи багатоенкодера та багатодекодери для обробки тексту в багатьох мовах;

– методи з використанням міжмовного перетину: цей підхід використовує проміжні мови для покращення якості перекладу між парою мов, для яких недостатньо паралельних даних. Можливі варіації такого методу, де можуть використовуватися кілька проміжних мов або знаходити оптимальний маршрут перекладу;

– методи з використанням обмеженого ресурсу: для малоресурсних мов або мов, для яких мало доступних даних для навчання моделей, можуть використовуватися методи, що використовують обмежений ресурс або сукупність знань для створення багатомовних систем перекладу.

Дослідження у цій області триває, і нові методи та підходи постійно розвиваються, щоб поліпшити якість багатомовного перекладу. Важливо також враховувати особливості кожної конкретної мови та мовного парку, оскільки це впливає на вибір і розробку підходу до багатомовного перекладу.

Класична проблема штучного інтелекту – пошук машини, здатної розуміти людську мову.

Наприклад, при пошуку «близьких італійських ресторанів» у вашій улюбленій пошуковій системі алгоритм має аналізувати кожне слово у вашому запиті та виводити відповідні результати. Пристойний застосунок для перекладу має розуміти контекст певного слова англійською мовою та якимось чином враховувати розбіжності у граматиці між мовами.

Усі ці та багато інших завдань підпадають під область інформатики, відому як Обробка природної мови або НЛП. Досягнення NLP призвели до появи широкого спектра практичних програм від віртуальних помічників, таких як Amazon Alexa, до спам-фільтрів, що виявляють шкідливу електронну пошту.

Найостаннішим проривом в НЛП є ідея велика мовна модель або LLM. LLM, такі як GPT-3, стали настільки потужними, що вони, здається, досягли успіху майже в будь-якій задачі NLP або сценарії використання.

За своєю суттю мовна модель – це просто алгоритм, який знає, наскільки можливо, що послідовність слів є правильною пропозицією.

Якщо замінити відносно невеликий набір даних масивним набором даних, взятим з Інтернету, суть почне наближатися до ідеї великої мовної моделі [12].

В нейронній мережі дослідники можуть навчати LLM на велику кількість текстових даних. Через велику кількість текстових даних, які побачила модель, LLM дуже добре пророкує наступне слово в послідовності.

Модель стає настільки складною, що може виконувати багато завдань НЛП. Ці завдання включають узагальнення тексту, створення нового контенту і навіть моделювання людської розмови [7].

Наприклад, дуже популярна мовна модель GPT-3 навчається з більш ніж 175 мільярдами параметрів і на сьогоднішній день вважається найпросунутішою мовною моделлю [9].

1.2 Основні тези дослідження великомовної моделі

За результатами дослідження великомовної моделі для багатомовного з використанням штучного інтелекту були виділені основні тези [13, 14]:

- великомовні моделі демонструють високу точність перекладу. Дослідження показали, що моделі «Encoder-Decoder» можуть досягати високого рівня точності при перекладі текстів з української мови на англійську, що робить їх ефективними інструментами для машинного перекладу;

- метрики якості підтверджують ефективність. Використання метрик, таких як BLEU, METEOR та ROUGE, свідчить про високу якість та співставність перекладу з референсними перекладами;

- потенціал розширення на інші мови та галузі. Великомовні моделі можуть бути застосовані для перекладу на інші мови та в різних галузях, забезпечуючи універсальність і корисність [13];

- адаптація до специфічних завдань. Можливість налаштування та навчання моделі на специфічних даних дозволяє досягти кращої якості перекладу для конкретних завдань та термінології [14];

– вимоги до оптимізації та швидкості виконання. Оптимізація часу виконання стає важливим аспектом для застосування великомовних моделей в реальному часі, особливо в онлайн-комунікації [14, 15];

– подальші можливості вдосконалення та розвитку. Дослідження підкреслюють перспективи розвитку технологій машинного перекладу, включаючи покращення точності, широкомовність та доступність;

– важливість етичних та соціальних аспектів. Розвиток систем машинного перекладу повинен супроводжуватися уважним врахуванням етичних питань та соціальних наслідків, зокрема, впливу на мовну культуру та професіональних перекладачів [16].

Ці тези підкреслюють важливість та потенціал великомовних моделей для полегшення мовного спілкування та забезпечення доступу до інформації в різних мовних середовищах.

1.3 Основні методи дослідження великомовної моделі

Створення багатомовної системи перекладу є складним завданням, яке включає різні кроки та методи. Ось кілька методів і підходів, які слід враховувати під час створення багатомовної системи перекладу:

– збір і підготовка даних: збір великих і різноманітних наборів даних із текстом кількома мовами. Цей набір даних має включати паралельні корпуси, які є наборами текстів різними мовами, які мають доступні переклади. Попередня обробка даних шляхом маркування, очищення та нормалізації тексту. Цей крок необхідний для ефективного навчання та тестування вашої системи;

– нейронний машинний переклад (NMT): моделі NMT, такі як моделі послідовності до послідовності з механізмами уваги (наприклад, Transformer), стали найсучаснішими для машинного перекладу. Ці моделі можуть працювати з кількома мовами, навчаючись на паралельних корпусах;

– архітектура моделі: використання однієї багатомовної моделі або набору моделей, що стосуються певної мови. Багатомовні моделі, як-от mBERT (Multilingual BERT), можуть працювати з кількома мовами в одній моделі, тоді як моделі для окремих мов можуть запропонувати кращу якість перекладу для окремих мов;

– навчання: навчання моделі на підготовленому наборі даних. Використання такі методи, як зворотний переклад, форсування вчителя та відсікання градієнта, щоб стабілізувати та покращити процес навчання. Треба експериментувати з різними гіперпараметрами, такими як швидкість навчання, розміри пакетів і архітектури моделей, щоб оптимізувати якість перекладу;

– точне налаштування: налаштування моделі на даних, що стосуються окремих мов, щоб покращити якість перекладу для окремих мов. Це може передбачати використання одномовних наборів даних або даних, що стосуються домену;

– оцінка та показники: використання показників оцінювання, як-от BLEU, METEOR або TER, щоб оцінити якість перекладів. Проведення як автоматичне, так і людське оцінювання, щоб удосконалити та порівняти свою систему;

– постредагування: впровадження зручного інтерфейсу, який дозволяє людям-перекладачам переглядати та редагувати створені машиною переклади для покращення якості;

– постійне вдосконалення: регулярне оновлення та перенавчання моделі новими даними для адаптації до мов і мовних стилів, що розвиваються;

– масштабованість і розгортання: створення своєї системи так, щоб вона була масштабованою, оскільки ви можете додати більше мов або вдосконалити існуючі. Розгляд можливості розгортання системи як хмарної служби або API для легкого доступу;

– етичні міркування: під час створення та розгортання багатомовної системи перекладу треба пам'ятати про етичні міркування, такі як упередженість у навчальних даних, проблеми конфіденційності та культурні особливості;

– відгуки користувачів: треба заохочувати відгуки користувачів, щоб постійно покращувати якість перекладів і відповідати конкретним потребам користувачів.

Головною метою є досягнення високої якості автоматичного перекладу, який буде корисним для різних сфер, включаючи бізнес, науку, освіту та культурний обмін.

Вибір напрямку дослідження методів формування моделі.

Дослідження спрямоване на вивчення та аналіз сучасних методів у галузі машинного перекладу, включаючи глибоке навчання та великомовні моделі. Основний акцент зроблено на розробці та оптимізації моделі «Encoder-Decoder» з метою покращення якості перекладу української мови на інші мови та навпаки.

Дослідження також включає аналіз методів збору, підготовки та аугментації даних для навчання моделі, а також вивчення етичних аспектів та конфіденційності у використанні великомовних моделей.

Загальна мета полягає в створенні високоякісної, швидкої та ефективної системи автоматичного перекладу, яка буде користуватися попитом у різних сферах життя та сприятиме подальшому розвитку машинного перекладу в Україні та за її межами.

1.4 Аналіз літературних джерел

У джерелах [1, 2] досліджено, як застосовується штучний інтелект у медицині. Об'єктом дослідження є методологія, методи, моделі та інформаційна технологія створення підсистеми збору, обробки,

перетворення, відображення та передачі інформації в спеціалізовану протиковідну інформаційно-аналітичну систему в умовах COVID пандемії.

Джерело [3] пояснює специфіку використання комп'ютерного моделювання застосування у мобільних пристроях охорони. Використання GSM модуля в пристроях сигналізації дозволяє створити надійну систему, яка може передавати повідомлення про вторгнення або аварійні ситуації на мобільний телефон або до центру моніторингу. У цьому огляді розглянуто склад охоронної системи і пристрій сигналізації на базі GSM модуля, а також переваги цієї розробки для різних груп користувачів.

Дослідження застосування штучного інтелекту під час роботи з кримінальними елементами [4]. Створення інтелектуальних систем для аналізу відвідувачів супермаркету з метою виявлення злочинних елементів є складним завданням, яке породжує кілька етичних і технічних проблем.

Джерело [7] досліджує моделювання повторюваних понять у незбалансованих потоках даних. Це задача машинного навчання, яка полягає в класифікації даних, коли кількість прикладів кожного класу неоднакова. Незбалансованість даних може виникнути, наприклад, коли один з класів дуже рідкісний або коли деякі приклади були відкинуті через помилки даних або неповноту. Для моделювання повторюваних понять у незбалансованих потоках даних можна використовувати різні методи, залежно від конкретної задачі і властивостей даних.

У дослідженні [9] представлено спрощений підхід до досягнення мети за допомогою деяких базових пакетів машинного навчання, наприклад TensorFlow, Keras, OpenCV і Scikit-Learn. Запропонований аналізує у людське обличчя на зображенні, а потім визначає, чи є на ньому маска чи ні.

У джерелі [10] досліджено, як штучний інтелект впливає на перекладацьку індустрію. ШІ – це дійсно потужна технологія, що постійно розвивається та трансформує перекладацьку індустрію. Разом з тим 43% респондентів дослідження Європейської комісії, Інституту мовознавців та

Інституту перекладу (Велика Британія) не погоджуються з тим, що розвиток технологій зробить людський переклад менш важливим в майбутньому.

У джерелі [11] досліджено, що автоматизований переклад зазвичай дослівний, без урахування особливостей тієї чи іншої мови та часто з помилками. Тому такий спосіб спілкування, до прикладу, можна порадити тому, хто зовсім не знає мови, яка йому потрібна. Але там, де III Міжнародна науково-практична конференція молодих учених та студентів «Філософські виміри техніки» (PDT-2022) 68 використовуються шаблони (юридичні, урядові документи), ці інструменти забезпечують досить якісний результат.

Джерело [12] стверджує, що з'являються нові види перекладу, такі як аудіовізуальний переклад, аудіодискрипція та транскреція. Для прикладу, послуги аудіовізуального перекладу і, зокрема, перекладу відео, користуються величезним попитом, цей попит продовжує зростати разом з непомірним попитом на відеоконтент.

Результати досліджень джерела [13]: в умовах активного використання засобів автоматизації перекладу доцільно говорити про новий етап дидактики перекладу: у змісті та формах навчання перекладу враховуються нові вимоги перекладацької галузі та значна частка машинного перекладу, водночас актуальними залишаються всі види немашинного перекладу.

Джерело [14]: висвітлюється проблема подання знань на основі природної мови у освітніх системах штучного інтелекту.

У джерелі [15] розроблено метод синтаксичного аналізу з елементами семантики природномовних конструкцій, результатом якого є семантична функція, що визначає синтаксичну й семантичну структуру. Розроблено підхід до побудови фрагмента системи інтелектуального контролю знань при електронному навчанні. Запропоновано метод використання семантичних функцій у системах машинного перекладу, де послівний переклад здійснюється в контексті речення.

Також, джерело [14] пояснює специфіку використання комп'ютерного моделювання у машинному перекладі.

У джерелі [15] аналізується сучасна диференціація мови, що зумовлено автоматичним аналізом і синтезом наукових текстів, логіко-лінгвістичних моделей у системах керування та виникненням інженерної лінгвістики. Проілюстровано приклади відтворювальних моделей, що використовуються в системах машинного анотування, реферування-перекладу і в машинних діалогових системах штучного інтелекту.

У джерелі [16] досліджені проблеми становлення українського термінотворення в галузі штучного інтелекту. Українське термінознавство перебуває в процесі свого становлення. Багато проблем залишаються поки що невирішеними. Термінологія штучного інтелекту не є винятком із правил. Контактуючи з різними мовами, українська так чи інакше запозичує нові слова. Тому нерідко постає проблема точного перекладу.

Метою джерела [17] є розробка інтелектуальної системи виправлення граматичних помилок для української мови на основі глибинних рекурентних нейронних мереж з використанням анотованих та синтетичних наборів даних.

Джерело [18] розповідає про рішення на основі глибокого навчання, які допомагають генерувати текстові виводи та скорочують час оцифрування документів Modi. Більшість роботи відбувалася лише на рівні символів у Modi OCR (оптичне розпізнавання символів).

В джерелах [19–24] представлено роботи щодо багатомовного нейронного машинного перекладу (MNMT), яке набуло значної популярності в останні роки. MNMT був корисним для покращення якості перекладу в результаті передачі знань про переклад (трансферне навчання). MNMT є більш перспективним і цікавим, ніж його аналог статистичного машинного перекладу, тому що наскрізне моделювання та розподілені представлення відкривають нові шляхи для досліджень машинного перекладу. Було запропоновано багато підходів для використання багатомовних паралельних корпусів для покращення якості перекладу. Однак відсутність комплексного

дослідження ускладнює визначення того, які підходи є перспективними і, отже, заслуговують на подальше дослідження.

Джерела [26–34] надають опис технології прийняття рішень в інформаційних системах, методи інтелектуального аналізу та оброблення даних, аналіз багатовимірних даних. Розглядаються особливості та перспективи використання інструментів штучного інтелекту (ШІ) в галузі письмового перекладу та навчання перекладу у закладах вищої освіти (ЗВО) України. Відповідно до результатів наукових розвідок світових та українських дослідників застосування ШІ в перекладацькій діяльності ще не набуло значного поширення, не зважаючи на стрімкий розвиток цього інструменту за останні 10 років. Окрім того, актуальним залишається питання підвищення технічної компетентності українських перекладачів.

У джерелах [35–44] досліджено сучасні системи штучного інтелекту, перспективи та проблеми новітніх технологій, вибір моделі перекладу. Виділені перспективні напрямки розвитку галузі, окреслені переваги та недоліки розглянутих методів у задачах «розуміння» природної мови, перекладу такого тексту та відповіді на питання.

Взагалі, технології та алгоритми для багатомовного перекладу з використанням штучного інтелекту включають в себе різноманітні підходи та методи, спрямовані на оптимізацію процесу перекладу. Ось деякі ключові аспекти, які можна розглянути в рамках такого дослідження:

– моделі машинного навчання:

1) перенавчання моделей (Transfer Learning): використання передньо навчених моделей для покращення якості перекладу та ефективності навчання нових пар мов;

2) Автокодувальні моделі (Autoencoder-based Models): використання автокодувальних моделей для отримання векторних представлень речень та їх подальше використання в перекладі;

– системи на основі Transformer:

1) моделі з архітектурою Transformer: використання Transformer-моделей, таких як BERT, GPT, або власні варіації, для здійснення багатомовного перекладу;

2) моделі з архітектурою Encoder-Decoder: використання моделей типу Encoder-Decoder для ефективного вирішення задачі перекладу між багатьма мовами;

– синтез мови та генерація тексту:

1) генерація тексту за допомогою GAN (Generative Adversarial Networks): використання алгоритмів GAN для синтезування тексту, що може допомогти у покращенні якості перекладу;

2) генерація тексту з використанням механізмів звертання (Attention Mechanisms): використання механізмів звертання для збору та врахування контексту при генерації перекладу;

– системи на основі Reinforcement Learning: навчання з підсиленням для оптимізації перекладу Використання методів навчання з підсиленням для покращення вибору слів та фраз у процесі перекладу;

– автоматизоване оцінювання та покращення якості:

1) метрики оцінювання:- BLEU, METEOR, ROUGE та інші метрики для автоматизованого оцінювання якості перекладу;

2) самооцінювання систем: використання механізмів самооцінювання для вдосконалення моделей на основі зворотного зв'язку;

– багатомовна семантична аналітика: семантичне врядування. Використання технік семантичного врядування для кращого врахування смислового змісту при перекладі між багатьма мовами;

– оптимізація інфраструктури та обчислень:

1) розподілені обчислення: Використання технік розподілених обчислень для прискорення процесу навчання та інференсу;

2) ефективне використання ресурсів хмарних сервісів: Можливе використання хмарних сервісів для вирішення проблеми обчислювальних ресурсів.

Ці технології та алгоритми представляють лише частину можливих підходів. У реальній розробці системи багатомовного перекладу може бути важливим експериментування з різними методами та їх комбінаціями для досягнення оптимального результату в конкретному контексті або завданні.

1.5 Постановка задачі дослідження

Метою роботи є дослідження технологій та алгоритмів для багатомовного перекладу з використанням штучного інтелекту, які дозволяють оптимізувати процес перекладу.

Досягнення визначеної мети передбачає вирішення таких завдань:

- дослідити методи для систем багатомовного перекладу з використанням штучного інтелекту;
- описати типову модель багатомовного перекладу;
- провести аналіз відомих технологій перекладу;
- проаналізувати сценарій ефективного використання методів багатомовного перекладу з використанням штучного інтелекту;
- дати оцінку ефективності їх реалізації та навести статистичні дані щодо наслідків імплементації.

Об'єктом дослідження є методи для багатомовного перекладу з використанням штучного інтелекту.

Предмет дослідження – технології та алгоритми багатомовного перекладу, інформаційні системи, сервіси та продукти даного напрямку.

2 МОДЕЛЬ «ENCODER-DECODER». ДОСЛІДЖЕННЯ МЕТОДІВ

2.1 Модель «Encoder-Decoder»

Модель «Encoder-Decoder» – це архітектурний підхід в машинному перекладі, який використовується для створення систем багатомовного перекладу. Цей підхід базується на використанні двох основних компонентів: «енкодера» і «декодера». В недавні роки моделі «Encoder-Decoder» стали популярними завдяки успіхам глибокого навчання та нейронних мереж у машинному перекладі.

Основні принципи моделі «Encoder-Decoder» в контексті багатомовного перекладу такі:

– Енкодер (Encoder): Енкодер призначений для перетворення вхідного тексту однієї мови на внутрішнє представлення, яке містить інформацію про зміст тексту. Це представлення може бути вектором або послідовністю векторів, що включає в себе інформацію про слова, фрази та контекст тексту;

– Декодер (Decoder): Декодер призначений для генерації вихідного тексту у цільовій мові на основі внутрішнього представлення, отриманого від енкодера. Декодер використовує це представлення для послідовного генерування перекладу, звертаючи увагу на контекст та раніше згенеровані слова;

– Механізм уваги (Attention): для покращення якості перекладу і здатності моделі до роботи з різними довжинами речень, у моделі «Encoder-Decoder» може використовуватися механізм уваги. Цей механізм дозволяє декодеру звертати увагу на різні частини вхідного тексту під час генерації вихідного тексту;

– Багатомовний переклад: модель «Encoder-Decoder» може бути розширена для багатомовного перекладу шляхом навчання моделі на паралельних даних, які включають текст у декількох мовах. У цьому

випадку, енкодер і декодер пристосовані для роботи з декількома мовами, і модель може здійснювати переклад між ними.

Моделі «Encoder-Decoder» зазвичай базуються на глибокому навчанні та використовуються для багатомовного перекладу, оскільки вони здатні враховувати складний контекст та взаємозв'язки між мовами. Цей підхід дозволяє покращити якість перекладу і зробити систему більш гнучкою для роботи з різними мовами та завданнями перекладу.

2.2 Збір і підготовка даних (Data Collection and Preparation)

Метод «Data Collection and Preparation» є одним з важливих кроків при побудові моделі для перекладу української мови з використанням штучного інтелекту. Цей процес включає в себе збір та підготовку даних для тренування та тестування моделі. Ось ключові аспекти цього методу [11]:

– збір даних (Data Collection):

1) корпуси текстів: перше завдання – зібрати велику кількість текстів на українській мові та їх відповідні переклади на інші мови. Це може включати в себе текстові дані з різних джерел, таких як вебсторінки, тексти книг, новини та інше;

2) різноманітність джерел: важливо збирати дані з різних джерел та різних стилів текстів, щоб модель навчалася розуміти та перекладати різні типи тексту;

3) збір паралельних текстів: якщо можливо, важливо збирати паралельні тексти, де кожен текст на українській мові має відповідний текст на іншій мові. Це допоможе моделі навчатися перекладу;

– підготовка даних (Data Preparation):

1) токенизація: дані повинні бути токенизовані на окремі слова або підрядки (токени), щоб модель могла працювати з ними на рівні слів;

- 2) видалення низькочастотних слів: деякі слова можуть бути рідкісними і нести мало інформації. Такі слова можна видалити з даних;
- 3) нормалізація тексту: важливо нормалізувати тексти, включаючи перетворення всіх літер у нижній регістр, видалення розділових знаків, чисел і т.д;
- 4) розділення на тренувальні та тестові дані: дані повинні бути розділені на навчальний та тестовий набори для оцінки якості моделі;
- 5) розмітка даних: для кожної пари український – цільовий текст повинні бути створені вектори або індекси, щоб модель могла їх обробляти;
- 6) аугментація даних: в деяких випадках може бути вигідно створювати додаткові дані шляхом аугментації, наприклад, додавання синонімів, перестановок слів, або зміни структури речень.

2.3 Нейронний машинний переклад (NMT)

Метод «Neural Machine Translation (NMT)» є одним з ключових підходів при побудові моделей для автоматичного перекладу, включаючи переклад української мови. Neural Machine Translation (NMT) включає в себе наступні аспекти [12]:

– нейронні мережі: у NMT використовуються нейронні мережі для моделювання процесу перекладу. Зазвичай це включає в себе використання глибоких нейронних мереж, таких як рекурентні нейронні мережі (RNN), Long Short-Term Memory (LSTM), або модель Transformer;

– Encoder-Decoder архітектура: основна структура NMT – це архітектура Encoder-Decoder, де Encoder обробляє вхідний текст, а Decoder генерує вихідний переклад. Encoder витягує семантичну інформацію з вхідного тексту, а Decoder генерує послідовність слів на цільовій мові;

– механізм уваги (Attention Mechanism): механізм уваги використовується для того, щоб Decoder міг фокусуватися на певних

частинах вхідного тексту, коли генерує кожне слово перекладу. Це допомагає покращити якість перекладу та забезпечує більшу контекстну відповідність;

– тренування: модель навчається за допомогою паралельних текстових корпусів, де кожен текст на початку навчання пов'язаний з відповідним перекладом. Під час тренування модель вивчає, як краще перекладати різні фрази та вирази;

– гіперпараметри та архітектура моделі: підбір гіперпараметрів та архітектури моделі є важливою частиною успішного впровадження NMT. Це включає в себе вибір кількості шарів, розміру прихованих шарів, швидкості навчання та інше.

Переваги NMT:

– висока якість перекладу: моделі NMT зазвичай забезпечують кращу якість перекладу в порівнянні зі старішими методами машинного перекладу;

– здатність вивчати контекст: завдяки механізму уваги, NMT здатні краще розуміти та відтворювати контекст в перекладах;

– здатність до роботи з довільними мовами: моделі NMT можуть використовуватися для перекладу між будь-якими парами мов.

Недоліки NMT:

– вимоги до обчислювальних ресурсів: навчання та використання NMT вимагає значних обчислювальних ресурсів, особливо для великих масштабних моделей;

– необхідність великої кількості даних: для досягнення високої якості перекладу потрібно велику кількість паралельних текстів для тренування моделі;

– проблеми з рідкісними мовами: для деяких рідкісних мов може бути важко зібрати достатньо даних для навчання ефективної моделі.

2.4 Архітектура моделі (Model Architecture)

Важливі аспекти дослідження методу «Model Architecture» [13]:

– тип нейронної мережі:

1) RNN (Recurrent Neural Network): рекурентні нейронні мережі були використані на ранніх стадіях розвитку машинного перекладу, але їхні обмеження (проблема зі зниклими градієнтами) призвели до використання більш сучасних архітектур;

2) LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit): ці рекурентні архітектури здатні краще керувати довгостроковою залежністю і вирішують проблему зі зниклими градієнтами;

3) Transformer: модель Transformer, яка базується на механізмах уваги, стала домінуючою архітектурою для задач машинного перекладу. Вона забезпечує паралельну обробку послідовностей та дозволяє досягати високої якості перекладу;

– кількість та розмір шарів: важливо визначити, скільки шарів буде в архітектурі Encoder та Decoder. Зазвичай використовуються багат шарові моделі для досягнення високої якості перекладу;

– Механізм уваги (Attention Mechanism): Механізм уваги дозволяє моделі фокусуватися на певних частинах вхідного тексту під час генерації перекладу. Дослідження різних типів механізмів уваги та їх підходів до реалізації є важливою частиною дослідження архітектури;

– функції активації та регуляризація: вибір функцій активації (наприклад, ReLU, Leaky ReLU, або Sigmoid) та методів регуляризації (наприклад, dropout) може впливати на здатність моделі генерувати високоякісний переклад;

– розмір ембедінгів (Embeddings): розмір та якість векторних представлень слова (ембедінгів) грають важливу роль у властивостях моделі. Великий розмір ембедінгів може покращити якість, але вимагає більше обчислювальних ресурсів;

– функція втрат (Loss Function): вибір функції втрат для навчання моделі є важливим аспектом архітектури. Зазвичай використовується функція втрат Cross-Entropy;

– застосування ансамблів (Ensemble): деякі дослідники використовують ансамблі моделей, що складаються з кількох нейронних мереж, для покращення якості перекладу;

– параметри навчання та оптимізатори: вибір параметрів навчання (швидкість навчання, розмір пакету тощо) і оптимізаторів (наприклад, Adam) важливий для зближення моделі до оптимальних параметрів.

Переваги архітектури моделі:

- здатність до роботи з послідовностями будь-якої довжини;
- здатність вивчати складні залежності та контекст у тексті;
- можливість досягати високої якості перекладу за наявності великих обчислювальних ресурсів і даних.

Недоліки архітектури моделі:

- вимоги до обчислювальних ресурсів: великі моделі можуть бути вимогливими до обчислювальних ресурсів для навчання та інференсу;
- потреба в великих обсягах даних: для досягнення високої якості перекладу потрібна велика кількість паралельних текстів для тренування.

2.5 Навчання (Training)

Метод «Training» є однією з ключових складових при побудові моделі. Ось деякі важливі аспекти методу «Training» [14];

- тренувальні дані: вибір відповідних тренувальних даних є критичним аспектом. Тренувальний корпус має містити достатньо текстів на українській мові та їхніх перекладів на цільову мову. Чим більше та різноманітніше тренувальних даних, тим краще модель може вивчити мовні особливості;

– розмір міні-пакетів (Batch Size): вибір розміру міні-пакетів впливає на швидкість навчання та стабільність процесу. Зазвичай використовують різні розміри міні-пакетів для навчання та валідації;

– функція втрат (Loss Function): вибір підходящої функції втрат є важливим аспектом. Зазвичай для задачі машинного перекладу використовується функція втрат Cross-Entropy;

– швидкість навчання (Learning Rate): швидкість навчання визначає темп оновлення параметрів моделі. Невірно налаштована швидкість навчання може спричинити проблеми з збіжністю моделі.

Переваги методу «Training»:

– дозволяє моделі вчити мовні особливості та адаптуватися до конкретного завдання перекладу;

– забезпечує можливість досягти високої якості перекладу за наявності великих тренувальних даних та обчислювальних ресурсів.

Недоліки методу «Training»:

– вимагає значних обчислювальних ресурсів та часу, особливо для навчання великих моделей;

– потребує великої кількості тренувальних даних, щоб досягти високої якості перекладу;

– вимагає уважної настройки гіперпараметрів та процесу навчання.

2.6 Точне налаштування (Fine-Tuning)

Метод «Fine-Tuning» або «Точне налаштування» є важливим етапом при побудові моделі і має деякі важливі аспекти дослідження [14]:

– вибір базової моделі: першим кроком є вибір попередньо навченої моделі, яка має високу якість або придатність для задачі перекладу. Такі моделі, як BERT, GPT, або Transformer, можуть служити в якості базової моделі для подальшого налаштування;

– вибір завдання та даних: для точного налаштування моделі потрібно визначити конкретне завдання перекладу української мови та зібрати або підготувати відповідні дані для навчання. Дані мають містити пари текстів на українській мові та їхніх перекладів на цільову мову.

Переваги методу «Fine-Tuning»:

– ефективний спосіб досягнути високої продуктивності моделі на конкретній задачі перекладу без необхідності навчати модель з нуля;

– використання попередньо навчених ваг дозволяє використовувати вже існуючі знання, отримані на великих наборах даних.

Недоліки методу «Fine-Tuning»:

– потребує наявності попередньо навченої моделі, яка підходить для задачі перекладу;

– може вимагати велику кількість тренувальних даних для досягнення високої якості перекладу;

– налаштування параметрів моделі та процесу навчання є важким завданням і може вимагати дослідження.

2.7 Оцінка та показники (Evaluation and Metrics)

Важливі аспекти дослідження методу «Evaluation and Metrics» [15]:

– метрики якості перекладу: для визначення якості перекладу використовуються різні метрики, такі як BLEU (Bilingual Evaluation Understudy), METEOR, TER (Translation Edit Rate), та інші. Важливо вибрати відповідну метрику або комбінацію метрик, яка відображає задовільність користувача;

– валідація на відкладеному наборі даних: для об'єктивної оцінки якості моделі важливо використовувати відкладений набір даних, який не використовувався під час навчання. Це дозволяє уникнути перенавчання та перевірити загальну здатність моделі до узагальнення;

– семантичні метрики: окрім традиційних метрик, деякі дослідження включають в себе використання семантичних метрик, які оцінюють відповідність перекладу до оригінального тексту на рівні семантики. Це може бути важливим для задач перекладу з точністю до значень слів та фраз;

– оцінка людськими експертами: для отримання більш об'єктивної оцінки якості перекладу можна залучити людських експертів. Це може включати оцінку на етапі розробки та після впровадження моделі в реальні умови;

– метрики продуктивності: поза метриками якості перекладу, важливо враховувати продуктивність моделі, таку як час перекладу, кількість ресурсів та швидкість реакції. Це особливо важливо для використання моделі в реальних застосунок.

Переваги методу «Evaluation and Metrics»:

– дозволяє об'єктивно оцінювати якість перекладу та порівнювати різні моделі;

– допомагає вдосконалювати модель шляхом вимірювання результатів та визначення напрямків покращень.

Недоліки методу «Evaluation and Metrics»:

– метрики можуть бути обмежені та не завжди відображати реальну задовільність користувача;

– вимагає наявності належних відкладених наборів даних та експертів для оцінки.

2.8 Постредагування (Post-Editing)

Метод «Post-Editing» передбачає процес післяперекладу, під час якого людський редактор втручається для поліпшення якості перекладу, виправлення помилок та адаптації перекладу до конкретного контексту або

вимог замовника. Ось деякі важливі аспекти дослідження методу «Post-Editing»:

– роль редакторів: визначення ролі та кваліфікації редакторів, які будуть втручатися у процес перекладу. Редактори повинні бути досвідченими у використанні машинного перекладу та розуміти вимоги та стандарти перекладу;

– інтерфейс та інструменти: розробка зручного інтерфейсу для редакторів, який дозволить ефективно вносити зміни до перекладу. Інструменти автоматичної підказки та аналізу тексту можуть полегшити процес редагування;

– вимоги до якості: визначення стандартів та вимог до якості, які редактори повинні враховувати при редагуванні перекладу. Це може включати граматичну правильність, лексичну відповідність, стильову узгодженість та інші критерії;

– збір інформації для покращення моделі: під час процесу постредагування можна збирати інформацію щодо помилок та покращень, які вносять редактори. Ця інформація може бути використана для навчання моделі та покращення її продуктивності в майбутньому.

Переваги методу «Post-Editing»:

– забезпечує високу якість перекладу, оскільки люди можуть краще розуміти контекст та вимоги завдання;

– дозволяє адаптувати переклад до конкретних потреб клієнта або особливостей тексту.

Недоліки методу «Post-Editing»:

– збільшує час та витрати на переклад у порівнянні з повністю автоматизованим перекладом;

– вимагає наявності досвідчених редакторів та координації їх роботи.

2.9 Постійне вдосконалення (Continuous Improvement)

Метод «Continuous Improvement» передбачає постійне оновлення та покращення моделі на основі зібраних даних та відгуків користувачів. Ось деякі важливі аспекти дослідження методу «Continuous Improvement» [16]:

– збір інформації: постійний збір інформації про якість перекладу, помилки, вимоги користувачів і відгуки. Це може бути здійснювано через відгуки в реальному часі, анкети, аналіз користувачських запитів та інші джерела;

– аналіз та вдосконалення: проведення аналізу зібраної інформації для виявлення шаблонів помилок та можливостей покращення. Це може включати розробку нових правил перекладу, корекцію алгоритмів та інші модифікації;

– навчання на нових даних: використання нових навчальних даних, які відображають актуальність і зміни в мові та контексті. Це дозволяє моделі адаптуватися до змін у мові та збільшувати точність перекладу;

– автоматизація процесу: впровадження автоматизованих процесів оновлення та вдосконалення моделі, щоб забезпечити ефективність та швидкість вдосконалення;

– заходи якості: забезпечення контролю якості під час процесу вдосконалення, включаючи валідацію та тестування оновлених моделей перед їхнім впровадженням.

Переваги методу «Continuous Improvement»:

– дозволяє підтримувати високу якість перекладу в умовах змінюючоїся мовної діна, нових вимог користувачів та інших факторів;

– сприяє постійному вдосконаленню та адаптації моделі до конкретних потреб користувачів.

Недоліки методу «Continuous Improvement»:

- вимагає великих зусиль та ресурсів на підтримку та розвиток системи перекладу;
- потребує постійного моніторингу та аналізу великої кількості даних.

2.10 Масштабованість і розгортання (Scalability and Deployment)

Метод «Scalability and Deployment» включає в себе планування, як модель може бути масштабована та розгорнута в різних вимірах та умовах. Ось деякі ключові аспекти дослідження методу «Scalability and Deployment» [17]:

- інфраструктура та ресурси: визначення необхідних обчислювальних ресурсів, таких як сервери, облака, кластери, для масштабування та розгортання моделі. Важливо обирати оптимальну інфраструктуру для завдань перекладу масштабного обсягу;

- автоматизація розгортання: розробка процесів автоматичного розгортання, які дозволяють легко встановлювати та оновлювати модель на різних серверах або обчислювальних вузлах. Це спрощує процес розгортання та забезпечує єдність середовища;

- масштабованість та завантаження: вивчення та планування того, як модель може бути масштабована, щоб відповідати високому завантаженню та забезпечити швидку відповідь на запити користувачів. Це може включати горизонтальне та вертикальне масштабування;

- балансування навантаження: розробка механізмів балансування навантаження для розподілу запитів між різними екземплярами моделі та забезпечення надійності та доступності;

- моніторинг та управління: впровадження систем моніторингу та управління, які дозволяють відстежувати працездатність моделі та реагувати на проблеми у реальному часі.

Переваги методу «Scalability and Deployment»:

- забезпечує можливість розгортання моделі в різних середовищах та масштабування для обробки великої кількості запитів;
- забезпечує високу доступність та надійність системи перекладу.

Недоліки методу «Scalability and Deployment»:

- вимагає складної інфраструктури та ресурсів для масштабування та розгортання;
- вимагає постійного моніторингу та адміністрування для забезпечення ефективності та надійності системи.

2.11 Етичні міркування (Ethical Considerations)

Метод «Ethical Considerations» або «Етичні міркування» включає в себе розгляд питань, пов'язаних із впливом технології на суспільство, прозорістю та відповідальністю за рішення, прийняті моделлю. Ось деякі ключові аспекти дослідження методу «Ethical Considerations»:

- забезпечення справедливості та багатокультурності: розгляд можливих впливів моделі на різні культури та групи людей. Важливо забезпечити, щоб переклад був справедливим, не дискримінував жодну групу та відображав різноманіття мови;

- контроль за біасами: аналіз та зменшення можливих біасів у перекладах, які можуть бути результатом навчання моделі на великих обсягах тексту. Важливо враховувати, що біаси можуть впливати на сприйняття інших культур і груп;

- прозорість та відповідальність: розробка механізмів для пояснення рішень, прийнятих моделлю, та визначення відповідальних осіб або організацій за роботу моделі;

- конфіденційність та захист даних: забезпечення конфіденційності користувачів та захисту їхніх особистих даних під час використання моделі;

– запитання авторських прав та власності: вирішення питань авторських прав на дані та модель, які використовуються в системі перекладу.

Переваги методу «Ethical Considerations»:

- забезпечує відповідність етичним нормам та стандартам у розробці та використанні технології;
- допомагає зберегти довіру користувачів та забезпечити позитивний вплив технології на суспільство.

Недоліки методу «Ethical Considerations»:

- вимагає додаткових ресурсів та зусиль для впровадження етичних принципів у розробку та використання моделі;
- може виникнути необхідність в постійному моніторингу та оновленні етичних принципів у зв'язку з змінами у суспільстві та технології.

2.12 Відгуки користувачів (User Feedback)

Метод «User Feedback» або «Відгуки користувачів» є важливим аспектом при побудові моделі для перекладу української мови з використанням штучного інтелекту. Відгуки користувачів надають важливу інформацію щодо якості та ефективності системи перекладу. Деякі ключові аспекти дослідження методу «User Feedback»:

– збір та аналіз відгуків: розробка механізмів для збору відгуків користувачів про якість та точність перекладів. Це може включати в себе оцінку відгуків, аналіз текстових коментарів і створення додаткових метрик для оцінки результатів;

– запитання та опитування: проведення опитувань та анкет серед користувачів для отримання конкретних відгуків та визначення їхніх потреб;

– аналіз та вдосконалення: використання зібраних відгуків для покращення моделі та алгоритмів перекладу. Аналіз відгуків може допомогти виявити шаблони помилок та проблем, які потребують виправлення;

– забезпечення відкритості та прозорості: забезпечення доступу користувачів до механізмів зворотного зв'язку та інформації про те, як їхні відгуки використовуються для покращення моделі.

Переваги методу «User Feedback»:

– дозволяє здійснювати постійне вдосконалення моделі на основі реальних потреб користувачів;

– забезпечує співробітництво зі спільнотою користувачів та підвищує їхню задоволеність.

Недоліки методу «User Feedback»:

– може вимагати значних ресурсів на обробку та аналіз великої кількості відгуків;

– може бути проблемами з впровадженням, якщо не вирішити питання конфіденційності та безпеки даних користувачів.

2.13 Обмеження великомовної моделі

Великомовної моделі здатні виконувати велику кількість варіантів використання навіть без тонкого налаштування. Тим не менш, ці моделі, як і раніше, мають свої обмеження:

– «Відсутність Семантичного Розуміння Світу». Здається, що великомовна модель демонструють інтелект. Однак ці моделі працюють не так, як людський мозок робить. Модель покладаються виключно на статистичні розрахунки для отримання результатів. Вони не здатні самостійно розмірковувати про ідеї та концепції. Через це LLM може видавати безглузді відповіді просто оскільки слова здаються «правильними»

чи «статистично ймовірними», коли вони розташовані у цьому конкретному порядку;

– «Галюцинації». Такі моделі, як GPT-3 також страждають від неточних відповідей. LLM можуть страждати від явища, відомого як галюцинація, де моделі видають фактично невірну відповідь, не усвідомлюючи, що відповідь не має під собою реальної основи. Наприклад, користувач може попросити модель пояснити думки Стіва Джобса про останній iPhone. Модель може генерувати цитату з повітря на основі своїх навчальних даних;

– «Упередження і Обмежені знання». Як і багато інших алгоритмів, великі мовні моделі схильні успадковувати усунення, присутні у навчальних даних. Оскільки процеси починають більше покладатися на LLM для отримання інформації, розробники цих моделей повинні знайти способи пом'якшити потенційно шкідливі наслідки упереджених відповідей. Так само сліпі зони навчальних даних моделі також заважатимуть самій моделі. Нині навчання великих мовних моделей йдуть місяці. Ці моделі покладаються на набори даних, обсяг яких обмежений. Ось чому ChatGPT має обмежені знання про події після 2021 року.

2.14 Переваги та недоліки розглянутих методів

Аналіз якості перекладу показав як переваги, так і недоліки розглянутих методів [45]. Коротко, позитивні і негативні моменти даних методів вказані в таблиці 2.1.

Таблиця 2.1– Позитивні і негативні моменти методів

Метод	Позитив	Негатив
1	2	3
Збір і підготовка даних	<ul style="list-style-type: none"> – Розширення корпусу текстів – Доступність великої кількості ресурсів – Застосування методів очищення та попередньої обробки – Можливість адаптації до конкретного завдання 	<ul style="list-style-type: none"> – Залежність від якості даних – Затрати часу та ресурсів – Авторські права та ліцензії
Нейронний машинний переклад (NMT) та Архітектура моделі	<ul style="list-style-type: none"> – Висока якість перекладу – Здатність до врахування контексту – Універсальність – Зручна інтеграція 	<ul style="list-style-type: none"> – Велика обчислювальна складність – Складність розробки та підтримки – Потреба в великому обсязі даних – Можливість перевантаження
Навчання	<ul style="list-style-type: none"> – Здатність до адаптації – Використання різних джерел даних 	<ul style="list-style-type: none"> – Наявність шуму в даних – Складність оптимізації

Продовження таблиці 2.1

1	2	3
Оцінка та показники	<ul style="list-style-type: none"> – Зрозумілість перекладу – Порівняння з іншими моделями – Покращення в процесі часу 	<ul style="list-style-type: none"> – Суб'єктивність оцінки – Залежність від якості тестових наборів – Необхідність ресурсів
Постредагування	<ul style="list-style-type: none"> – Коригування неточностей – Забезпечення відповідності контексту – Адаптація до специфічних завдань 	<ul style="list-style-type: none"> – Людський фактор – Можливість втрати структури та стилю
Постійне вдосконалення	<ul style="list-style-type: none"> – Адаптація до змін – Урахування відгуків користувачів – Досягнення кращої ефективності 	<ul style="list-style-type: none"> – Час та ресурси – Ризик помилок – Постійна оновленість
Масштабованість і розгортання	<ul style="list-style-type: none"> – Збільшення обсягу перекладу – Глобальний доступ – Спільне використання ресурсів – Зменшення витрат 	<ul style="list-style-type: none"> – Збільшення обчислювальних вимог – Проблеми з конфіденційністю

Продовження таблиці 2.1

1	2	3
Етичні міркування	<ul style="list-style-type: none"> – Захист конфіденційності – Визначення меж використання – Зменшення ризику негативних наслідків 	<ul style="list-style-type: none"> – Складність регулювання – Суб'єктивність оцінки – Обмеження для розвитку
Відгуки користувачів	<ul style="list-style-type: none"> – Поліпшення якості перекладу – Адаптація до потреб користувачів – Залучення користувачів до розробки – Виявлення негайних помилок 	<ul style="list-style-type: none"> – Суб'єктивність відгуків – Обмежена репрезентативність – Велика кількість відгуків – Недостатнє контролювання даними

Усі ці аспекти підкреслюють важливість дбайливого та обачного підходу до збору та підготовки даних для великомовних моделей для перекладу.

Створення надійної багатомовної системи перекладу є безперервним процесом, який вимагає поєднання лінгвістичного досвіду, методів машинного навчання та предметних знань. Треба мати на увазі, що конкретні методи та інструменти, які використовуються, можуть відрізнятися залежно від цілей і мов, які потрібно підтримувати.

3 КОМП'ЮТЕРНА МОДЕЛЬ ДЛЯ БАГАТОМОВНОГО ПЕРЕКЛАДУ

3.1 Обґрунтування вибору середовища програмної реалізації

Дослідження методів створення системи багатомовного перекладу може бути досить важким і часомістким завданням, і вибір середовища програмної реалізації грає важливу роль у процесі дослідження. Ось декілька обґрунтувань, які можуть вплинути на вибір середовища програмування для створення системи багатомовного перекладу:

– мови програмування та бібліотеки: вибір мови програмування та бібліотек для розробки системи багатомовного перекладу залежить від ваших навичок, досвіду та завдань дослідження. Python є популярним вибором для машинного навчання та обробки природної мови завдяки багатьом бібліотекам, таким як TensorFlow, PyTorch і scikit-learn. Вибір мови може також вплинути на доступність ресурсів та спільноту розробників;

– обсяг даних: великий обсяг паралельних даних для навчання є ключовим фактором в успішному створенні системи багатомовного перекладу. Вибір середовища має включати можливість роботи з великими обсягами даних, зокрема зберігання, обробку та навчання моделей на них;

– глибоке навчання та нейронні мережі: багатомовний переклад часто базується на глибокому навчанні та нейронних мережах, таких як моделі «Encoder-Decoder» або моделі трансформера. Вибір середовища повинен дозволяти розроблювати і навчати такі моделі;

– ресурси і обчислювальна потужність: потреби в обчислювальній потужності для навчання глибоких нейронних мереж можуть бути значними. Вибір середовища повинен враховувати доступність обчислювальних ресурсів, таких як GPU або TPU, які можуть значно прискорити навчання моделей;

– зручність розробки і налагодження: зручність розробки та налагодження коду також важлива. Вибір середовища, яке надає ефективні

інструменти для налагодження, профілювання та тестування, може спростити процес дослідження;

– спільнота та підтримка: існування активної спільноти розробників і наявність підтримки у вигляді документації, форумів та інших ресурсів можуть значно полегшити процес розробки та дослідження.

Обсяг даних також є важливим фактором в розробці системи багатомовного перекладу, і проблеми з ним можуть виникнути в різних аспектах процесу. Ось декілька типових проблем із обсягом даних та способи їх вирішення:

– недостатньо даних для навчання моделі може призвести до поганої якості моделі перекладу. Вирішення:

1) розглядати можливість використання передньо навчених моделей (pre-trained models), які можуть бути доопрацьовані на вашому обмеженому датасеті;

2) збільшувати обсяг доступних даних шляхом обробки інших джерел, використовуючи техніки паралельного корпусування та використання відкритих даних;

3) використовувати методи зменшення розміру моделі, які можуть потребувати менше навчальних даних;

– дефіцит даних для рідкомовних мов може бути складно знайти достатню кількість навчальних даних для якісного перекладу. Вирішення:

1) розглядати можливість залучення спільноти користувачів для збору та анотації даних;

2) використовувати методи аугментації даних для створення більше прикладів з обмеженого обсягу;

– нерівномірність даних між мовами. Обсяг навчальних даних може бути нерівномірним між мовами, що призводить до нерівномірності якості перекладу. Вирішення:

1) зберігати баланс між мовами у датасетах, збираючи однакову кількість прикладів для кожної мови;

2) використовувати техніки зважування даних для збалансування впливу різних мов на навчання моделі;

– помилки та шум в даних. Дані можуть містити помилки, неточності та шум, що може впливати на якість перекладу. Вирішення:

1) використовуйте методи фільтрації та обробки даних для видалення шуму та корекції помилок;

2) використовуйте експертну анотацію та ревізію даних для поліпшення якості датасету.

Обчислювальні ресурси можуть бути обмеженими, і під час розробки системи багатомовного перекладу можуть виникати проблеми, пов'язані з їх недостатністю. Ось декілька загальних проблем, які можуть виникнути, та способи їх вирішення:

– недостатньо потужності GPU/TPU: великі моделі глибокого навчання, такі як моделі перекладу, можуть вимагати значної обчислювальної потужності для навчання та інференсу. Якщо у вас обмежений доступ до графічних процесорів (GPU) або спеціалізованих обчислювальних пристроїв (TPU), це може бути проблемою. Вирішення:

1) збільшити обсяг доступних ресурсів, якщо це можливо;

2) використовувати менші моделі з меншою кількістю параметрів, якщо це не впливає на якість перекладу;

3) розглянути використання обчислювальних хмарних сервісів, де можна орендувати додаткові ресурси при необхідності;

– помилки при роботі з пам'яттю: обробка великих датасетів та моделей може призвести до вичерпання обсягу доступної оперативної пам'яті, що може спричинити помилки та вимушені зупинки. Вирішення:

1) використовувати оптимізовані бібліотеки та методи для управління пам'яттю, такі як фрагментація пам'яті, пакетна обробка даних, робота з даними в невеликих пакетах тощо;

2) використовувати техніки мінімізації споживання пам'яті, які надаються бібліотеками глибокого навчання;

– довгий час навчання та інференсу: обчислювально витратні завдання, такі як навчання великих моделей або стекованих перекладачів, можуть вимагати довгих часових витрат. Вирішення:

1) розглядати можливість використання розподіленого обчислення для прискорення навчання та інференсу;

2) збільшити кількість ресурсів, якщо це можливо;

– дорогі моделі та дані: використання дорогих моделей та дорогоцінних даних може вимагати значних фінансових ресурсів. Вирішення:

1) розглядати використання безкоштовних або менш дорогих альтернатив, якщо це можливо;

2) використати алгоритми для перенавчання моделей на менших датасетах або зменшення розміру моделі без втрати якості перекладу;

– керування ресурсами на платформі: деякі платформи обмежують ресурси, доступні для обчислень. Вирішення:

1) дотримуватися обмежень платформи та вивчати їхні можливості для оптимізації роботи з ресурсами;

2) розглядати можливість використання інших платформ або апаратного забезпечення.

Обчислювальні проблеми можуть бути важливими при розробці систем багатомовного перекладу, і їх вирішення вимагає ретельного планування та оптимізації.

Вибір середовища програмування повинен враховувати ці фактори і відповідати конкретним потребам вашого дослідження методів створення системи багатомовного перекладу. Також важливо зазначити, що комбінування різних інструментів і мов програмування може бути вигідним для різних аспектів проєкту.

Розробка багатомовної системи перекладу є складним дослідницьким проєктом, який складається з кількох етапів. Нижче надано загальний огляд цих кроків разом із деякими прикладами коду Python. Слід зазначити, що це спрощена демонстрація, і створення повноцінної дослідницької системи

потребуватиме більш глибокої роботи та, можливо, групи дослідників та інженерів.

У рамках кваліфікаційної роботи були розроблені такі кроки:

- збір і підготовка даних,
- модель нейронного машинного перекладу (NMT),
- навчання,
- оцінювання,
- тонка настройка та оптимізація,
- розгортання.

3.2 Програмна реалізація

Цей скрипт створює систему багатомовного перекладу на Python з використанням моделей з бібліотеки Hugging Face Transformers та оцінює якість перекладу за допомогою метрики BLEU. Ось опис реалізації скрипта:

– імпортується необхідні бібліотеки та класи для роботи з мовними моделями, датасетами та обчислення метрики BLEU;

```
from datasets import load_dataset
import evaluate
from progressbar import progressbar
from transformers import MarianMTModel, MarianTokenizer
```

– визначається інтерфейс ‘TranslatorInterface’, який містить абстрактний метод ‘translate’. Цей інтерфейс служить базовим класом для всіх перекладачів;

```
class TranslatorInterface:
    def translate(self, text: str) -> str:
        pass
```

– клас ‘Translator’ реалізує інтерфейс ‘TranslatorInterface’. У цьому класі ініціалізуються модель та токенизатор для конкретної пари мов, а також визначено метод ‘translate’, який приймає вхідний текст, токенизує його, перекладає та декодує результат;

```
class Translator(TranslatorInterface):
    def __init__(self) -> None:
        super().__init__()
        self.model = None
        self.tokenizer = None

    def translate(self, text: str) -> str:
        # Tokenize and convert the text to IDs
        input_ids = self.tokenizer.encode(text, return_tensors='pt')

        # Translate the text to the target language
        translated = self.model.generate(
            input_ids,
            max_length=100,
            num_return_sequences=1,
            early_stopping=True
        )

        # Decode the translation
        return self.tokenizer.decode(translated[0], skip_special_tokens=True)
```

– дочірні класи ‘UaToEnTranslator’, ‘UaToEsTranslator’ та ‘EsToEnTranslator’ успадковують функціональність класу ‘Translator’ та

ініціалізують моделі та токенизатори для конкретних мовних пар. Наприклад, ‘UaToEnTranslator’ відповідає за переклад з української на англійську;

```
class UaToEnTranslator(Translator):
    def __init__(self) -> None:
        super().__init__()
        # Load the pretrained Ukrainian-to-English translation model
        self.model_name = 'Helsinki-NLP/opus-mt-uk-en'
        self.model = MarianMTModel.from_pretrained(self.model_name)
        self.tokenizer = MarianTokenizer.from_pretrained(self.model_name)
```

```
class UaToEsTranslator(Translator):
    def __init__(self) -> None:
        super().__init__()
        # Load the pretrained Ukrainian-to-Spanish translation model
        self.model_name = 'Helsinki-NLP/opus-mt-uk-es'
        self.model = MarianMTModel.from_pretrained(self.model_name)
        self.tokenizer = MarianTokenizer.from_pretrained(self.model_name)
```

```
class EsToEnTranslator(Translator):
    def __init__(self) -> None:
        super().__init__()
        # Load the pretrained Spanish-to-English translation model
        self.model_name = 'Helsinki-NLP/opus-mt-es-en'
        self.model = MarianMTModel.from_pretrained(self.model_name)
        self.tokenizer = MarianTokenizer.from_pretrained(self.model_name)
```

– клас ‘StackedTranslator’ реалізує інтерфейс ‘TranslatorInterface’ і приймає список перекладачів. В методі ‘translate’ текст послідовно перекладається кожним з перекладачів у заданому порядку;

```

class StackedTranslator(TranslatorInterface):
    def __init__(self, translators: list[TranslatorInterface]) -> None:
        super().__init__()
        self.translators = translators

    def translate(self, text: str) -> str:
        result = text
        for translator in self.translators:
            result = translator.translate(result)
        return result

```

– у функції ‘main’ завантажуються моделі та датасет для оцінки якості перекладу. Здійснюється переклад та розрахунок метрики BLEU для прямого перекладу (з української на англійську) та стекованого перекладу (з української на іспанську, а потім на англійську);

```

def main():
    print('Load models')
    ua_en_translator = UaToEnTranslator()
    ua_es_translator = UaToEsTranslator()
    es_en_translator = EsToEnTranslator()
    ua_es_en_translator = StackedTranslator([ua_es_translator,
                                             es_en_translator])

    print('Load test dataset')
    corpus = load_dataset('trondizzy/uk_en_combined_sets')['train']
    sample_size = 10000
    sample = corpus[:sample_size]
    source, target = sample['UK'], sample['EN']

```

```

print('Translate')
direct_translation, stacked_translation = [], []
for text in progressbar(source):
    direct_translation.append(ua_en_translator.translate(text))
    stacked_translation.append(ua_es_en_translator.translate(text))

print('Calculate BLEU metric')
metric = evaluate.load("sacrebleu")
direct_score = metric.compute(predictions=direct_translation,
    references=target)
stacked_score = metric.compute(predictions=stacked_translation,
    references=target)

print(f"UK -> EN score:\t\t{direct_score['score']}")
print(f"UK -> ES -> EN score:\t\t{stacked_score['score']}")

```

– результати оцінки виводяться на екран.

Цей скрипт демонструє, як створити систему багатомовного перекладу з використанням моделей Hugging Face Transformers та оцінити якість перекладу за допомогою метрики BLEU.

3.3 Інструкція користувача

Для роботи користувача в програмному застосунку розроблені вимоги до оточення та інструкція користування:

– Вимоги:

- 1) python 3.8+
- 2) 6 ГБ вільного місця

– Встановлення:

- 1) `python3 -m venv venv # створити віртуальне оточення`
- 2) `source venv/bin/activate # активувати віртуальне оточення`
- 3) `pip install -r requirements.txt # встановити залежності`

– Використання:

- 1) якщо віртуальне оточення не активоване, то
`source venv/bin/activate`
- 2) `python main.py # или ./main.py`

3.4 Тестування та аналіз розробленої моделі

Аналіз результатів виконання даного скрипту включає оцінку якості перекладу та порівняння результатів прямого перекладу та стекованого перекладу. Було розглянуто кожен частину аналізу окремо:

– оцінка якості перекладу. У скрипті проводиться оцінка якості перекладу з української на англійську мову двома способами: прямий переклад та стекований переклад (з української на іспанську, а потім на англійську):

1) ‘direct_score’: це якість перекладу з української на англійську за допомогою моделі ‘ua_en_translator’;

2) ‘stacked_score’: це якість перекладу з української на англійську через проміжну іспанську мову за допомогою моделей ‘ua_es_translator’ та ‘es_en_translator’.

– порівняння результатів. Зазвичай важливо порівнювати результати прямого та стекованого перекладу, щоб визначити, чи можна поліпшити якість перекладу за допомогою такого підходу. Якщо результат стекованого

перекладу кращий або принаймні на рівні прямого перекладу, то цей підхід може бути вигідним;

– вивід результатів. Результати оцінки якості перекладу виводяться на екран у вигляді балів BLEU для обох методів. Результати можуть бути у вигляді числа від 0 до 100, де вище значення вказує на кращу якість перекладу. Це дозволяє оцінити, наскільки кращим є стекований переклад порівняно з прямим;

– інші фактори: результати можуть бути вплинуті наступними факторами: обсягом тренувальних даних, якістю моделей, обсягом тестового корпусу та іншими. Важливо бути свідомим цих факторів під час аналізу.

Зазвичай метрика BLEU є досить грубою оцінкою якості перекладу, і для більш об'єктивної оцінки може бути використана додаткова експертна аналіз, залучення користувачів або інші метрики, такі як METEOR, ROUGE тощо.

Аналіз результатів може виявити, наскільки ефективним є використання стекованого перекладу для покращення якості перекладу і чи варто використовувати цей підхід для конкретних завдань перекладу.

Після виконання скрипта отримано результати (лістинг 3.1).

Лістинг 3.1 Результати виконання скрипту:

```

Load models
Load test dataset
Translate
100% (1000 of 1000)
|#####|
Elapsed Time: 1:05:32 Time: 1:05:32
Calculate BLEU metric
UK -> EN score:      18.967307739140043
UK -> ES -> EN score: 16.93822836490653

```

Це значить, що наявний метод прямого перекладу з української на англійську має оцінку якості перекладу 18,96 за метрикою BLEU, в той час як метод стекованого перекладу (з української на іспанську, а потім на англійську) має оцінку 16,93.

За результатами метрики BLEU, метод прямого перекладу показав кращу якість перекладу порівняно з стекованим перекладом. Різниця між оцінками становить 2 бала, що може бути важливим поліпшенням у деяких випадках.

Оцінка BLEU є грубою метрикою і не завжди відображає реальну користь в конкретних завданнях. Поруч із числовими метриками, слід проводити і експертну оцінку якості перекладу, оскільки якість перекладу може бути суб'єктивною.

Важливо продовжувати дослідження, вдосконалювати моделі та враховувати особливості конкретних завдань перекладу. Результати аналізу можуть вказувати на необхідність подальших досліджень та розробки.

Загалом, аналіз результатів показав покращення якості перекладу завдяки прямому перекладу в даному контексті, але завжди слід бути обережним і враховувати інші фактори, які можуть вплинути на прийняття рішення про використання цього підходу в реальних проєктах.

ВИСНОВКИ

У рамках кваліфікаційної роботи на підставі результатів дослідження великомовної моделі для багатомовного перекладу з використанням штучного інтелекту були зроблені наступні висновки:

– точність перекладу зростає. Великомовні моделі, такі як модель «Encoder-Decoder», здатні досягати високої точності перекладу. Оцінки якості перекладу за допомогою метрик, таких як BLEU, METEOR та ROUGE, показують високі результати, що свідчить про ефективність таких моделей;

– застосування в різних галузях. Великомовні моделі для перекладу можуть бути використані в різних галузях, включаючи локалізацію програмного забезпечення, автоматичний переклад текстів новин, літературний переклад, переклад для електронної комерції та багато інших сфер;

– зручність використання. Великомовні моделі дозволяють створювати системи машинного перекладу, які легко взаємодіють з користувачами. Інтерфейси для введення тексту та отримання перекладу можуть бути інтуїтивно зрозумілими та зручними для користувачів;

– можливості налаштування та покращення. Для підвищення якості перекладу можливість налаштування гіперпараметрів, оптимізації функцій втрат та додаткової обробки даних є важливими аспектами. Відповідна настройка може значно покращити результати перекладу;

– спеціалізовані завдання. Для завдань з особливою термінологією та вимогами до якості перекладу може вимагатися додаткова настройка та навчання моделі на відповідних даних;

– час виконання. Час виконання інференсу (перекладу) може бути значущим фактором, особливо при обробці великих обсягів тексту. Оптимізація швидкості може бути важливою задачею для реалізації великомовних систем.

Існують варіанти майбутньої роботи з покращення МП між іншими офіційними мовами та українською окрім якості, досягнутої в рамках цієї роботи. Один із них полягає в отриманні більш якісних даних. Ще одним потенційним напрямком майбутньої роботи є навчання багатомовних моделей машинного перекладу, що перекладають з багатьох вихідних мов однією цільовою мову. У такій ситуації включення однієї багатої на ресурси мовної пари, такої як англійська-українська, може допомогти за допомогою засобів трансфертного навчання або, принаймні, як форма регуляризації. [45]

У загальному, дослідження великомовної моделі для перекладу української мови з використанням штучного інтелекту показало великий потенціал для покращення якості та зручності машинного перекладу. Такі системи можуть бути корисними для різних завдань та вигідними в різних галузях, але вимагають належної настройки та оптимізації для досягнення найкращих результатів [46-55].

На основі результатів дослідження великомовної моделі для перекладу української мови з використанням штучного інтелекту можна виділити кілька перспектив та напрямків для подальшого розвитку та вдосконалення цієї технології:

- покращення якості перекладу. Продовжувати дослідження та розробку моделей та алгоритмів для покращення якості перекладу. Враховувати в основному різні галузі та сфери застосування для покращення специфічності перекладу;

- багатомовність. Розглядати можливості розширення системи на інші мови та створення мульти- або багатомовних моделей, що може бути корисним для багатьох багатомовних спільнот та компаній;

- адаптація до контексту: розробляти моделі та методи, які можуть краще адаптуватися до контексту тексту. Наприклад, розуміння іменованих сутностей, фраз з використанням ідіом тощо;

– забезпечення прозорості. Робити системи машинного перекладу більш прозорими та інтерпретованими, щоб користувачі могли розуміти, як вони працюють та чому давали той чи інший результат;

– оптимізація часу виконання. Працювати над покращенням швидкості виконання машинного перекладу, особливо в режимі реального часу, що дозволить застосовувати систему в різних галузях, включаючи онлайн-комунікацію та послуги;

– підвищення якості навчання. Розширювати обсяги навчальних даних та покращувати якість даних для навчання моделей, що може призвести до покращення результатів;

– робота з шумними даними. Розробляти алгоритми та методи, які можуть працювати з шумними або поганими даними, оскільки в реальних умовах джерело даних часто може бути неідеальним;

– застосування у вищій освіті та наукових дослідженнях. Впровадження систем машинного перекладу у вищій освіті для полегшення доступу до академічних джерел та публікацій для студентів та дослідників.

Враховуючи ці перспективи, системи машинного перекладу з використанням великомовних моделей можуть стати ще більш важливими та корисними інструментами в різних сферах життя, допомагаючи покращити спілкування та доступ до інформації в міжнародному контексті.

Наукова новизна одержаних результатів: проведено аналіз технічної реалізації технологій багатомовного перекладу з використанням штучного інтелекту.

Практичне значення отриманих результатів. Застосування отриманих результатів дозволяє підвищити якість машинного перекладу.

Результати дослідження апробовано у вигляді тез доповідей під час I Міжнародної науково-практичної конференції «NEW WAYS OF CREATING SCIENTIFIC IDEAS FOR IMPLEMENTATION» [25], та статті у студентському науковому журналі «UNIVERSUM» [56].

Великі мовні моделі можуть справді змінити те, як як відбувається взаємодія з технологіями та нашим світом загалом.

Безліч даних, доступних в Інтернеті, дало дослідникам можливість моделювати складності мови. Однак по ходу справи ці мовні моделі, здається, засвоїли людське розуміння світу таким, яким він є.

У міру того, як громадськість починає довіряти цим мовним моделям у забезпеченні точних результатів, дослідники та розробники вже знаходять способи додати бар'єри, щоб технологія залишалася етичною.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Прокіпець, В., & Кузьомін, О. (2022). МЕТОДИ АНАЛІЗУ ЗОБРАЖЕНЬ ЛЕГЕНІВ ДЛЯ ДІАГНОСТУВАННЯ COVID. *Grail of Science*, (14-15), 356-361.
2. Шустрова, А. Є., & Кузьомін, О. Я. (2023). Спільний автокодер із порогом виявлення аномалії суглоба на виробничих лініях.
3. Меденцев, Д. В., & Кузьомін, О. Я. (2023). *Розробка пристрою з GSM сигналізацією* (Doctoral dissertation).
4. Berkovskyi, D., & Kuzomin, O. (2023). CREATION OF INTELLIGENT SYSTEMS FOR ANALYZING SUPERMARKET VISITORS TO IDENTIFY CRIMINAL ELEMENTS. *Collection of scientific papers «SCIENTIA»*, (May 5, 2023; Sydney, Australia), 113-118.
5. Uchqun o'g'li, B. S., Kuzomin, O., & Lyashenko, V. (2023). Decision support procedures for decision making in a COVID condition.
6. Верколаб, Г. С. (2022). *Розробка та дослідження детектору маски для обличчя з OPENCV, KERAS/TENSORFLOW і глибоким навчання* (Doctoral dissertation).
7. Холодов, С. Є. (2023). Моделювання повторюваних понять у незбалансованих потоках даних.
8. Кузьомін, О. Я., Василенко, О. О., & СВИСТУНОВ, І. (2020). Розробка багатоагентних структур для вирішення проблем медичної системи діагностування. *Радіоелектроніка та інформатика*, 2, 47-54.
9. Кузьомін, О. Я. (2008). *Методи, моделі та інформаційні технології моніторингу і ліквідації наслідків надзвичайних природних ситуацій* (Doctoral dissertation, ОЯ Кузьомін).
10. ЕПОХУ, П., & ІНТЕЛЕКТУ, Ш. (2023). Олена Кулик. *Сайт збірника наукових праць*: <http://ses.journal.in.ua/index.php/ses>, 48.

11. Маковійчук, Л. (2022). Штучний інтелект у перекладацькій діяльності. *Збірник тез III Міжнародної науково-практичної конференції молодих учених та студентів „Філософські виміри техніки“*, 67-68.
12. Ігнатенко, В. Д. (2023). Цифровізація та трансформація ринку перекладацьких послуг як відповідь на виклики сьогодення.
13. Головацька, Ю. АКТУАЛЬНІ ТЕНДЕНЦІЇ ПЕРЕКЛАДАЦЬКОЇ ДІЯЛЬНОСТІ ТА ПІДГОТОВКИ МАЙБУТНІХ ПЕРЕКЛАДАЧІВ У МІНЛИВИХ УМОВАХ СУЧАСНОСТІ.
14. Зінченко, О., Краснопольський, В., & Артеменко, Ю. (2023). СПЕЦИФІКА ВИКОРИСТАННЯ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ У МАШИННОМУ ПЕРЕКЛАДІ (НА ПРИКЛАДІ АНГЛІЙСЬКОЇ МОВИ). *Вісник науки та освіти*, (6 (12)).
15. Козак, Л. В. (2016). Сучасна диференціація мови. *Філологічні студії: Науковий вісник Криворізького державного педагогічного університету*, 15, 55-63.
16. Коновалова, В. Б., & Мирошніченко, В. М. (2017). Проблеми становлення українського термінотворення в галузі штучного інтелекту.
17. Яковлева, О. М. (2023). Вирішення задачі виправлення граматичних помилок в текстах з використанням сучасних технологій глибинного навчання.
18. Shah, R., Gupta, M. K., & Kumar, A. (2021, November). Line Level Modi (Heritage script) OCR using Attention based Encoder-Decoder Architecture. In *2021 Sixth International Conference on Image Information Processing (ICIIP)* (Vol. 6, pp. 273-278). IEEE.
19. Dabre, R., Chu, C., & Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5), 1-38.
20. Fu, Z., Lam, W., Yu, Q., So, A. M. C., Hu, S., Liu, Z., & Collier, N. (2023). Decoder-Only or Encoder-Decoder? Interpreting Language Model as a Regularized Encoder-Decoder. *arXiv preprint arXiv:2304.04052*.

21. Lu, L., Zhang, X., & Renais, S. (2016, March). On training the recurrent neural network encoder-decoder for large vocabulary end-to-end speech recognition. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5060-5064). IEEE.

22. Ma, S., Dong, L., Huang, S., Zhang, D., Muzio, A., Singhal, S., ... & Wei, F. (2021). Deltalm: Encoder-decoder pre-training for language generation and translation by augmenting pretrained multilingual encoders. *arXiv preprint arXiv:2106.13736*.

23. Wang, Z., Liu, Z., Wei, W., & Duan, H. (2021). Saled: Saliency prediction with a pithy encoder-decoder architecture sensing local and global information. *Image and Vision Computing, 109*, 104149.

24. Sandhan, J., Singha, R., Rao, N., Samanta, S., Behera, L., & Goyal, P. (2022). TransLIST: A Transformer-Based Linguistically Informed Sanskrit Tokenizer. *arXiv preprint arXiv:2210.11753*.

25. Стебаєв, І. ДОСЛІДЖЕННЯ ВЕЛИКОМОВНОЇ МОДЕЛІ ДЛЯ ПЕРЕКЛАДУ УКРАЇНСЬКОЇ МОВИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ. In *The I International Scientific and Practical Conference «New ways of creating scientific ideas for implementation», September 18-20, 2023, Varna, Bulgaria. 285 p.* (p. 269).

26. Творошенко, І. С. (2021). Технології прийняття рішень в інформаційних системах: навч. посібник. Харків: ХНУРЕ.

27. Гороховатський, В. О., & Творошенко, І. С. (2021). Методи інтелектуального аналізу та оброблення даних: навч. посібник.

28. Гороховатський В.О., Творошенко І.С. (2022) Аналіз багатовимірних даних за описом у формі множини компонент: монографія. Харків: ХНУРЕ, 124 с.

29. Gorokhovatskyi, V.O., Tvoroshenko, I.S., and Peredrii O.O. (2020) Image classification method modification based on model of logic processing of bit description weights vector, *Telecommunications and Radio Engineering, 79*(1), pp. 59-69.

30. Гороховатський В.О., Творошенко І.С., Чмутов Ю.В. (2022) Застосування систем ортогональних функцій для формування простору ознак у методах класифікації зображень, *Сучасні інформаційні системи*, 6(3), С. 5–12.
31. Tvoroshenko I., and Gorokhovatskyi V. (2022) The Application of Hybrid Intelligence Systems for Dynamic Data Analysis, *International Journal of Engineering and Information Systems*, 6(2), pp. 40–48.
32. Pomazan, V., Tvoroshenko, I., & Gorokhovatskyi, V. (2023). Development of an application for recognizing emotions using convolutional neural networks, *International Journal of Academic Information Systems Research*, 7(7), pp. 25-36.
33. Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O., & Vlasenko, N. (2023). Search for visual objects by request in the form of a cluster representation for the structural image description, *Advances in Electrical and Electronic Engineering*, 21(1), pp. 19-27.
34. Красуля, А. В., & Турчина, М. В. (2020). Використання інструментів штучного інтелекту: порівняльний аналіз систем автоматизованого перекладу.
35. Удудяк, Б. О., Берестенко, Д. О., Мелешко, Є. В., & Якименко, М. С. (2023). ДОСЛІДЖЕННЯ СУЧАСНИХ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ З ВІДКРИТИМ КОДОМ ДЛЯ ГЕНЕРАЦІЇ ТЕКСТІВ. *Тези доповідей*, 81.
36. Грабас, С. (2023). Штучний інтелект: перспективи та проблеми новітніх технологій. *Матеріали VI Міжнародної студентської науково-технічної конференції „Природничі та гуманітарні науки. Актуальні питання“*, 136-138.
37. Турчина, М. В. (2020). *Тенденції розвитку технологій штучного інтелекту: перекладацький аспект* (Master's thesis, Сумський державний університет).

38. Мельницька, В. Б. (2023). *Система перетворення мови жестів у текст на основі штучного інтелекту* (Bachelor's thesis, КПІ ім. Ігоря Сікорського).
39. Левчук, А. (2015). Штучний інтелект: лінгвістичні аспекти. *Науковий вісник Східноєвропейського національного університету імені Лесі Українки*, (4), 203-207.
40. Пушик, Н. (2021). КОМП'ЮТЕРНА ЛІНГВІСТИКА ТА «ШТУЧНИЙ ІНТЕЛЕКТ». *Молодий вчений*, (2 (90)), 151-155.
41. Гудкова, Н. М. (2023). Перспективи машинного перекладу та міжкультурна емпатія. In *Діалог культур у Європейському освітньому просторі*. Київський національний університет технологій та дизайну.
42. Матківська, Н. А. (2014). Вибір моделі перекладу при відтворенні ідіолекту. *Наукові записки Національного університету Острозька академія. Серія: Філологічна*, (45), 281-283.
43. Онищенко, К. Г., Данієль, Я., & Каменєв, Р. (2020). Аналіз методів обробки природної мови.
44. Моренцова, А. В. (2018). Використання онтології предметної області для усунення неоднозначностей при комп'ютерному перекладі технічних текстів. *Актуальні питання сучасної науки: III Міжнародна науково-практична інтернет-конференція: тези доповідей, Дніп-ро, 30*, 82-87.
45. Bergmanis, T., & Pinnis, M. (2022). From Zero to Production: Baltic-Ukrainian Machine Translation Systems to Aid Refugees. *arXiv preprint arXiv:2209.14142*.
46. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., and Zeghid M. (2022) Tools for fast metric data search in structural methods for image classification, *IEEE Access*, 10, pp. 124738-124746.
47. Gorokhovatskyi V., Tvoroshenko I., Kobylin O., and Vlasenko N. (2023) Search for visual objects by request in the form of a cluster representation for the structural image description, *Advances in Electrical and Electronic*

Engineering, 21(1), pp. 19-27.

48. Гороховатський В.О., Творошенко І.С., Чмутов Ю.В. (2022) Застосування систем ортогональних функцій для формування простору ознак у методах класифікації зображень, *Сучасні інформаційні системи*, 6(3), С. 5-12.

49. Гороховатський В., Передрій О., Творошенко І., Марков Т. (2023) Матриця відстаней для множини компонентів структурного опису як інструмент для створення класифікатора зображень, *Сучасні інформаційні системи*, 7(1), С. 5-13.

50. Pomazan V., Tvoroshenko I., and Gorokhovatskyi V. (2023) Development of an application for recognizing emotions using convolutional neural networks, *International Journal of Academic Information Systems Research*, 7(7), pp. 25-36.

51. Pomazan V., Tvoroshenko I., and Gorokhovatskyi V. (2023) Handwritten character recognition models based on convolutional neural networks, *International Journal of Academic Engineering Research*, 7(9), pp. 64-72.

52. Tvoroshenko I., Gorokhovatskyi V., Kobylin O., and Tvoroshenko A. (2023) Application of deep learning methods for recognizing and classifying culinary dishes in images, *International Journal of Academic and Applied Research*, 7(9), pp. 57-70.

53. Gorokhovatskyi V., Tvoroshenko I. (2023) Identification of visual objects by the search request. *International scientific symposium «INTELLIGENT SOLUTIONS-S». Computational intelligence (results, problems and perspectives). Decision making theory: proceedings of the international symposium*, September 28, 2023, Kyiv-Uzhorod, Ukraine, pp. 25-27.

54. Yakovleva O., Kovač M., Ardasov V. & Yeremenko I. (2023). Study on adding functionality to the Zoom online conference system for monitoring the participant activities, *Public Administration and Regional Development*, 19(1), pp. 158-184.

55. Alberts, I. L., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?. *European journal of nuclear medicine and molecular imaging*, 50(6), 1549-1552.

56. Стебаєв, І., & Кузьомін, О. (2023). Дослідження великомовної моделі для перекладу української мови з використанням штучного інтелекту. *UNIVERSUM*, (2), 85-94.