

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерних наук (або центр післядипломної освіти, або навчально-науковий центр заочної форми навчання)
(повна назва)

Кафедра _____ програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти _____ другий (магістерський)

Дослідження методів налаштування великих мовних моделей
для перекладу художніх текстів
(тема)

Виконав:
здобувач _____ 2 _____ року навчання
групи _____ ПЗМ-23-1

Олексій ТРОФІМЕНКО
(Власне ім'я, ПРІЗВИЩЕ)

Спеціальність _____ 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми _____ освітньо-наукова

Керівник _____ доц. Анастасія ЧУПРИНА
(посада, Власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри

(підпис)

Кирило СМЕЛЯКОВ
(Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки
 Факультет _____ комп'ютерних наук _____
 Кафедра _____ програмної інженерії _____
 Рівень вищої освіти _____ другий (магістерський) _____
 Спеціальність _____ 121 – Інженерія програмного забезпечення _____
 Тип програми _____ освітньо-наукова програма _____
 Освітня програма _____ Інженерія програмного забезпечення _____
 (шифр і назва)

ЗАТВЕРДЖУЮ:
 Зав. кафедри _____
 (підпис)

« » _____ 2025 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Трофіменко Олексію Олеговичу _____
 (прізвище імя, по батькові)

1. Тема роботи «Дослідження методів налаштування великих мовних моделей для перекладу художнього тексту»

Затверджена наказом по університету від 15.04.2025 р. №290 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 13.06.2025


3. Вихідні дані до роботи: попередньо навчені великі мовні моделі (LLM) з відкритим доступом та інструменти для роботи з ними. Паралельні корпуси художніх текстів для тренування й тестування моделей, мова програмування Python, фреймворк Flask, середовище розробки VS code, провайдер Together.

4. Перелік питань, які потрібно опрацювати в роботі:
вступ, аналіз предметної галузі, теоритичне дослідження, проведення експериментальних досліджень, висновки, перелік джерел посилання, додатки.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	16.04.2025	виконано
2	Аналіз предметної галузі і постановка задачі	20.04.2025	виконано
3	Огляд і аналіз літературних, наукових джерел	30.04.2025	виконано
4	Постановка задачі	29.05.2025	виконано
5	Теоретичне дослідження	01.05.2025	виконано
6	Практичне дослідження	10.05.2025	виконано
7	Підготовка до апробації результатів дослідження. Публікація матеріалів.	24.05.2025	виконано
8	Підготовка пояснювальної записки	30.05.2025	виконано
9	Підготовка презентації та доповіді	31.05.2025	виконано
10	Перевірка на плагіат	07.06.2025	виконано
11	Нормоконтроль	07.06.2025	виконано
12	Рецензування	08.06.2025	виконано
13	Попередній захист	09.06.2025	виконано
14	Занесення диплома в електронний архів	12.06.2025	виконано
15	Допуск до захисту у зав. кафедри	13.06.2025	виконано

Дата видачі завдання 15 квітня 2025р.

Студент 

 (підпис)

Олексій ТРОФІМЕНКО

Керівник роботи _____
 (підпис)

доц. Анастасія ЧУПРИНА
 (посада, Власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 57 с., 18 рис., 4 додатки, 14 джерел.

ВЕЛИКІ МОВНІ МОДЕЛІ, КОНТЕКСТУАЛЬНИЙ ПЕРЕКЛАД, LORA, ПЕРЕКЛАД ХУДОЖНІХ ТЕКСТІВ, FINE-TUNING.

Об'єктом дослідження є автоматизований переклад художніх текстів із використанням великих мовних моделей.

Метою роботи є розробка підходів до налаштування великих мовних моделей для забезпечення якісного перекладу художніх текстів із врахуванням стилістики та контексту.

Методами дослідження є аналіз існуючих підходів до перекладу, вивчення методів адаптації мовних моделей, таких як LoRA та fine-tuning, і оцінка їх ефективності для вирішення поставленої задачі

У результаті роботи було проведено дослідження методів налаштування великих мовних моделей, яке продемонструвало можливості покращення точності й стилістичної відповідності перекладів художніх текстів, забезпечуючи врахування культурного контексту та узгодженість термінології.

LARGE LANGUAGE MODELS, CONTEXTUAL TRANSLATION, LORA, ARTISTIC TEXT TRANSLATION, FINE-TUNING.

The object of research is the automated translation of artistic texts using large language models.

The purpose of the work is to develop approaches to customizing large language models for ensuring high-quality translation of artistic texts while maintaining style and context

The research methods include analyzing existing translation approaches, studying adaptation methods for language models such as LoRA and fine-tuning, and evaluating their effectiveness in solving the outlined problem.

As a result of the work, a study of large language model adaptation methods was conducted, demonstrating the potential to improve the accuracy and stylistic coherence of artistic text translations while considering cultural context and terminology consistency.

Заява щодо самостійного виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу EIArKhNURE.

Я, Трофіменко Олексій Олегович, студент гр. ПЗм-21-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів налаштування великих мовних моделей для перекладу художніх текстів», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата

Підпис

ЗМІСТ

Вступ.....	9
1 Аналіз предметної галузі і постановка задачі	10
1.1 Аналіз предметної галузі.....	10
1.2 Постановка задачі.....	13
2 Огляд й аналіз літературних, наукових джерел	14
2.1 Огляд основних джерел.....	14
2.2 Аналіз літератури	15
2.3 Оцінка актуальності та новизни	16
2.4 Висновки з огляду літератури.....	17
3 Постановка задачі.....	18
3.1 Постановка задачі.....	18
3.2 Необхідні ресурси	19
3.3 Можливі обмеження	20
4 Теоретичне дослідження	21
4.1 Архітектура та проектування програмного забезпечення.....	21
4.2 Use Case діаграма	22
4.3 Методи та алгоритми.....	24
4.4 Розробка інтерфейсу	27
5 ПРАКТИЧНЕ ДОСЛІДЖЕННЯ	29
5.1 Опис експериментальних даних	29
5.2 Опис реалізованих методів	29
5.3 Опис реалізованих метрик	31
5.4 Результати експериментів	32
Висновки	37
Перелік джерел посилання	39
Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	41
Додаток А Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ	42

Додаток Б Слайди презентації	43
Додаток В Апробація результатів роботи.....	52
Додаток Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам дсту 3008:2015.....	57

ВСТУП

Переклад художніх текстів є важливим завданням, яке потребує не лише точності, але й збереження стилю та контексту. Традиційні інструменти, такі як Google Translate, часто демонструють обмеження у перекладі складних текстів, зокрема втрату стилістичної єдності та помилки у врахуванні граматичних особливостей мов. Це особливо помітно мовах, де граматичні ознаки залежать від роду та числа.

Великі мовні моделі (LLM) відкривають нові можливості для автоматизації перекладу. Завдяки здатності зберігати контекст та адаптуватися до стилістики тексту, вони потенційно здатні забезпечити якісніший переклад, зберігаючи узгодженість термінів і враховуючи культурні особливості. Це дозволяє уникнути багатьох обмежень традиційних підходів.

Мета роботи полягає у дослідженні підходів до налаштування LLM для перекладу художніх текстів, які забезпечують точність, стилістичну відповідність і узгодженість. Для цього важливо дослідити сучасні методи адаптації LLM, такі як налаштування параметрів моделі, і створити рішення, які дозволять передавати авторський стиль та зберігати структуру тексту.

Об'єктом дослідження є процес автоматичного перекладу текстів, предметом – методи адаптації мовних моделей для забезпечення якісного перекладу. У результаті роботи очікується розробка підходів, які сприятимуть покращенню автоматизованого перекладу і забезпечать його застосування для художньої літератури. Результати можуть бути корисними як для дослідників у сфері обробки природної мови, так і для розробників інструментів автоматичного перекладу.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ І ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз предметної галузі

Переклад художніх творів (прози, поезії, драматургії) є складним завданням, яке потребує не лише точної передачі змісту, а й збереження стилістичних, культурних та емоційних складових оригіналу. На відміну від технічних або наукових текстів, де основною вимогою є точна передача фактів та термінології, художні тексти містять образи, метафори, алюзії, діалектизми, гру слів, специфічний ритм, стиль та тональність. Передача цих аспектів при машинному перекладі часто ускладнюється обмеженими можливостями алгоритмів адекватно інтерпретувати та генерувати художні структури, що виходять за рамки формальної логіки.

Перше широко розповсюджене покоління машинного перекладу базувалося на статистичних моделях (див. рис. 1), що вивчали частотні співвідношення між фразами у паралельних корпусах.

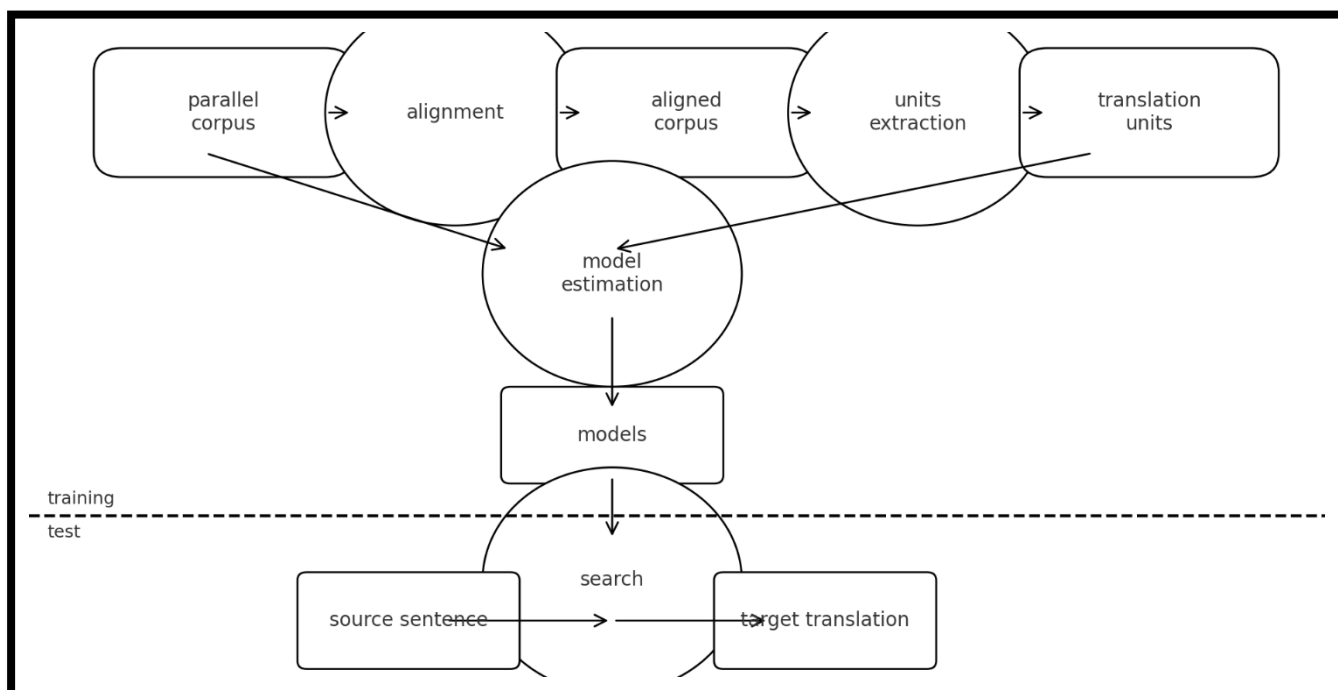


Рисунок 1 – Принцип роботи статистичної моделі перекладу (SMT) (рисунок створено самостійно)

Такий підхід давав задовільні результати для структурованих текстів, однак був слабким у відтворенні стилістики та контекстних нюансів. Відсутність у статистичних моделей розуміння глибинних семантичних зв'язків часто призводила до неточних чи надто буквальних перекладів, позбавлених художньої цінності.

Запровадження нейронних мереж (див. рис. 2) для перекладу тексту стало кроком уперед: моделі змогли враховувати контекст речення краще [1], ніж статистичні аналоги.

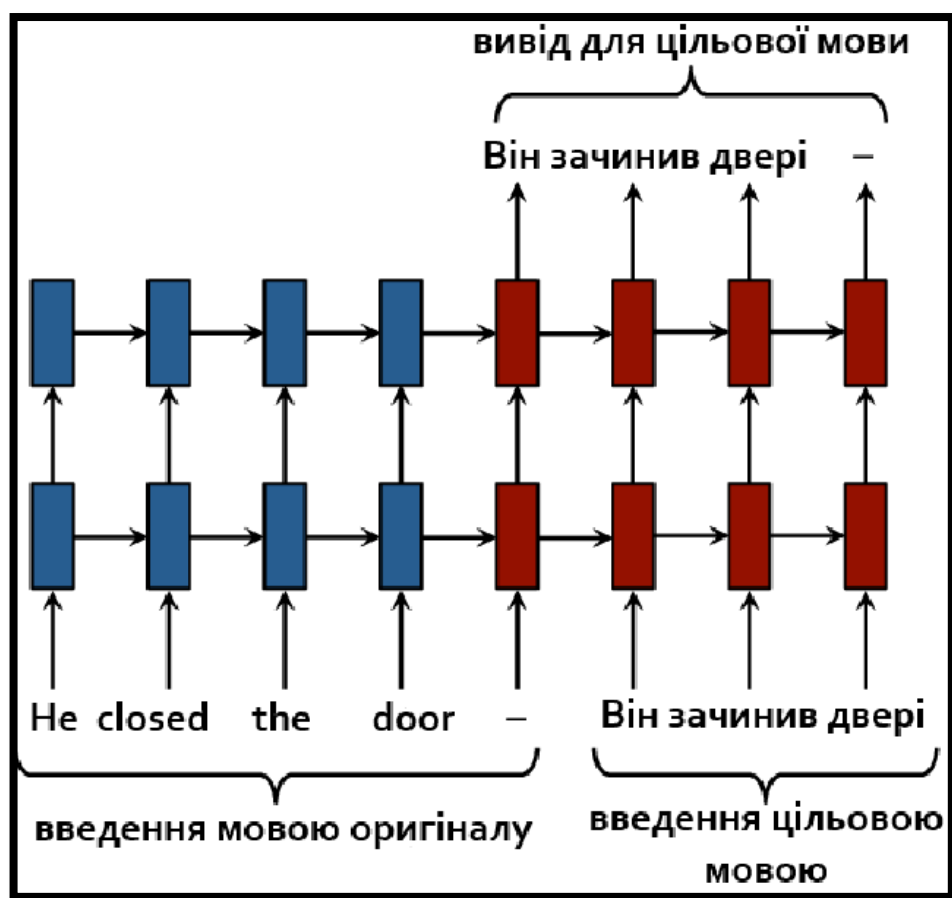


Рисунок 2 – Діаграма роботи NMT моделі (рисунок створено самостійно)

Проте обмежений контекст (кілька речень або абзаців) та недостатня увага до стилістичних деталей залишалися проблемою. Хоча нейронні мережі цього покоління навчилися краще передавати зміст, естетичні і культурні аспекти все ще втрачалися.

Архітектура трансформерів дозволила опрацювати довший контекст та краще моделювати залежності між словами. Цей підхід покращив якість перекладу [2], однак базові трансформерні моделі без спеціальної адаптації все ще далекі від ідеального відтворення стилістики художніх творів. Введення великих мовних моделей (LLM) із мільярдами параметрів надало змогу моделювати більш тонкі смислові зв'язки, але питання адаптації до специфіки літературного тексту залишилося актуальним.

Сучасні підходи передбачають використання готових LLM, донавчених (fine-tuned) на специфічних корпусах художніх текстів. Це відкриває шлях до більш тонкого контролю за стилістикою і тональністю, може підвищити якість при роботі з іншими мовами [3] на специфічних доменах [4]. Легковагові адаптери, такі як LoRA, дозволяють мінімізувати ресурси на перенавчання [5] та швидко «підлаштувати» модель під конкретний жанр, автора або конкретний твір. Таким чином, з'являється можливість ефективно переймати стилістичні риси оригіналу, уникаючи проблем перекладу мови.

Сучасна практика перекладу художніх текстів часто базується на ручній праці професійних перекладачів, які мають не лише глибоке знання мов, але й чуття стилю, літературні здібності, культурний та історичний контекст. Однак, з огляду на збільшення обсягів текстів (книги, сценарії, література для кіно та ігрової індустрії, літературні бази даних), виникає потреба у автоматизованих або напівавтоматизованих рішеннях, здатних пришвидшити роботу перекладача, надавати йому чорновий матеріал для подальшої редактури або навіть виконувати якісний переклад повністю автоматично.

Масштаб проблеми полягає у тому, що багато літературних творів потенційно потребують перекладу або адаптації для різних мовних аудиторій. Поточні рішення пропонують певний рівень автоматизації, проте для отримання перекладів, наближених до людської якості, необхідні подальші дослідження. Поки що жодна зі стандартних систем, що працюють без додаткових налаштувань, не гарантує достатньо високого рівня художньої цінності перекладу.

1.2 Постановка задачі

Виходячи з аналізу предметної галузі, основною проблемою є покращення якості та точності автоматизованого перекладу художніх текстів за допомогою великих мовних моделей, зокрема шляхом їх адаптації. Метою дослідження стає розробка підходів, які забезпечуватимуть високу адекватність перекладу, збереження стилістичної цілісності, культурної адаптації та послідовності наративу.

Повинні бути пройдені наступні кроки для дослідження:

- потрібно проаналізувати існуючі техніки донавчання великих мовних моделей (fine-tuning) на спеціалізованих корпусах художньої літератури;
- потрібно оцінити ефективність легковагових адаптерів для налаштування моделі без повного перенавчання;
- потрібно оцінити можливості комбінування підходів для покращення стилістичної чутливості перекладу.

Очікуваним результатом дослідження є розробка таких підходів до адаптації LLM, які сприятимуть покращенню якості автоматизованого перекладу художньої літератури.

2 ОГЛЯД Й АНАЛІЗ ЛІТЕРАТУРНИХ, НАУКОВИХ ДЖЕРЕЛ

2.1 Огляд основних джерел

Для формування бази джерел було враховано такі критерії як: авторитетність, актуальність, об'єктивність, достовірність, країна надходження та мова видання.

Пріоритет надавався публікаціям у відомих наукових журналах і конференціях, роботам відомих дослідницьких груп та університетських наукових бібліотек.

Перевага надавалася матеріалам, опублікованим за останні 5–10 років, оскільки галузь машинного перекладу швидко розвивається, і старіші дослідження можуть бути менш релевантними.

Обиралися статті з прозорою методологією та зрозумілим описом експериментів, бажано з відкритими наборами даних.

Не використовувалися джерела з країни-агресора та російськомовні видання.

Перевагу надано рецензованим публікаціям, роботам із цитуваннями, а також документам, які мають підтвердження у вигляді верифікованих результатів експериментів.

Було виділено наступні згруповані теми джерел:

- загальні підходи до машинного перекладу;
- адаптація LLM та спеціалізовані методи налаштування;
- дослідження у сфері художнього перекладу;
- метрики оцінки якості перекладу.

Згідно джерел виявлено потребу у пошуку статей, що описують еволюцію від SMT до NMT, а також перехід до трансформерних архітектур, робот, присвячених fine-tuning, LoRA, та іншим підходам оптимізації великих мовних моделей для доменно-специфічних задач, працям, де розглядаються проблеми стилю, культурної адаптації, передачі метафор та поетичних засобів, літератури, що пропонує кількісні та якісні методи вимірювання точності, стилістичної відповідності та ефективності перекладу.

2.2 Аналіз літератури

У базових працях з машинного перекладу, таких як «Evaluation of SMT and NMT Systems», висвітлено еволюцію від статистичних до нейронних методів. Стаття аналізує порівняльну продуктивність SMT і NMT систем за допомогою метрик BLEU, TER та METEOR, демонструючи переваги NMT у точності передачі значення та стилістичній відповідності.

У контексті великих мовних моделей (LLM), дослідження у публікації «Multilingual Machine Translation with Large Language Models» аналізує мультилінгвальний машинний переклад за допомогою LLM. Незважаючи на високі показники BLEU при перекладі на англійську, ці моделі поступаються супервізованим підходам, як-от NLLB, особливо для малоресурсних мов. Автори також відзначають важливість підбору контекстних прикладів для навчання, що підкреслює складність інтерактивного налаштування.

Методи адаптації великих мовних моделей розглянуті у публікації «LoRA vs Full Fine-tuning: An Illusion of Equivalence» пропонують ефективне рішення для зниження обчислювальних витрат при донавчанні. LoRA використовує низькорангові матриці для модифікації параметрів моделей, що дозволяє досягати продуктивності, близької до повного fine-tuning.

Дослідження художнього перекладу висвітлюють унікальні виклики, пов'язані з передачею стилю, культурних елементів та метафор. Наприклад, у публікації «A Survey on Evaluation Metrics for Machine Translation» [6] наголошується на важливості інтеграції баз знань для інтерпретації культурно маркованих елементів. Використання паралельних корпусів літературних текстів з анотаціями стилістичних особливостей значно підвищує якість перекладу таких текстів.

У сфері оцінки якості перекладу ключовими залишаються автоматичні метрики, як-от BLEU, METEOR, та новіші, зокрема COMET і BLEURT. Ці метрики показують вищу кореляцію з людською оцінкою завдяки використанню контекстуальних векторів. Проте навіть вони не завжди здатні врахувати стилістичну та емоційну складову текстів, що є важливим для художнього

перекладу. Автори підкреслюють необхідність залучення експертів для ручної оцінки перекладів, щоб забезпечити збереження авторського стилю та емоційного забарвлення.

Загалом, сучасні дослідження демонструють тенденцію до поєднання автоматизованих підходів із людською експертизою для оцінки та вдосконалення перекладу. Особливу увагу приділяють розробці адаптивних методів налаштування моделей (Prefix-Tuning, Prompt-Tuning), які дозволяють оптимізувати їх для специфічних завдань. Важливим напрямком залишається вдосконалення метрик, що здатні оцінювати не лише точність, а й стилістичну відповідність перекладів.

2.3 Оцінка актуальності та новизни

Актуальність теми обумовлена стрімким зростанням попиту на переклад літератури для міжнародних ринків. Потреба у швидкому та недорогому перекладі постійно збільшується, тому що кількість публікованих книг у жанрі художньої літератури стрімко зростає (див. рис. 3).

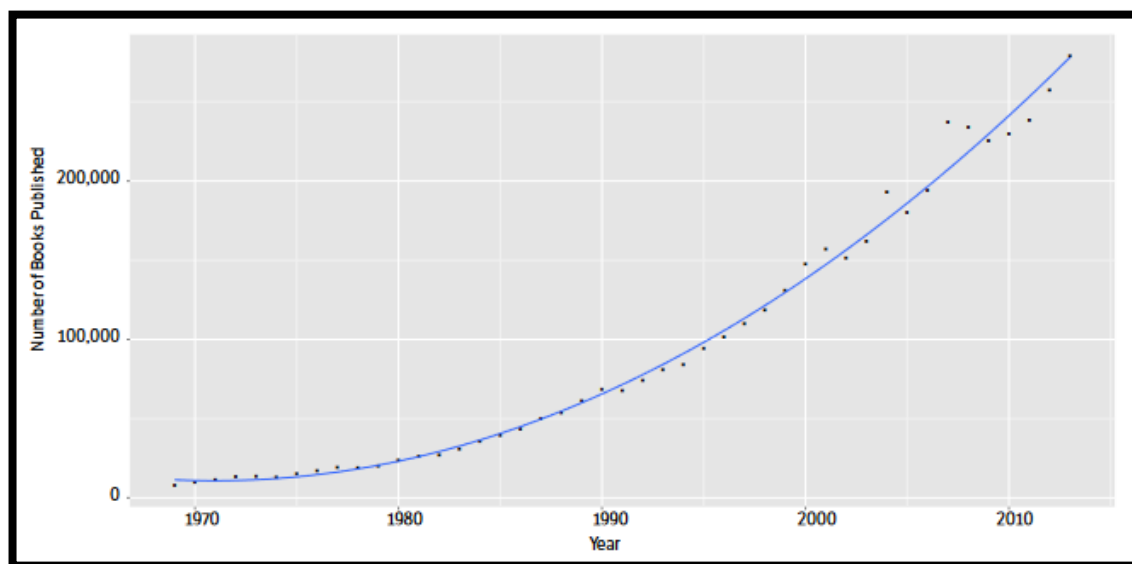


Рисунок 3 – Кількість опублікованих англійських книг за роками (за даними [7])

Однак якісні переклади – все ще рідкість. Новизна полягає у застосуванні найсучасніших мовних моделей (LLM) та методів їхньої адаптації до художніх

текстів. Більшість робіт, опублікованих останнім часом, ще не досягли рівня масового застосування у видавничій сфері, що свідчить про високий потенціал дослідження.

Сучасний стан проблеми відображає необхідність пошуку нових методів оцінки якості: література також висвітлює ідеї створення комбінації автоматизованих та експертних методів оцінки. Дослідження все більше наголошують увагу про необхідність додаткової людської валідації.

2.4 Висновки з огляду літератури

LLM мають достатній інтелектуальний запас, щоб розуміти складні контексти та семантичні зв'язки. Однак без адаптації вони не гарантують якісного художнього перекладу. Fine-tuning та легковагові адаптери (LoRA) є перспективними методами, що дозволяють підвищити стилістичну відповідність перекладу, зберігаючи водночас гнучкість моделі.

Література наголошує на потребі розуміння культурних контекстів та жанрових особливостей. Створення спеціалізованих корпусів, словників, інтеграція з базами знань – ключ до точнішого перекладу.

Класичні метрики не завжди адекватно оцінюють літературні аспекти. Потрібен комплексний підхід, який передбачає оцінку людиною. Бракує рішень, які б масово впроваджувалися на практиці.

Немає усталених підходів, які б гарантували відтворення унікального стилю автора чи складних поетичних форм. Залишається відкритим питання, як масштабувати й автоматизувати адаптацію LLM для різних жанрів, історичних періодів, культур.

Отже, аналіз наукових джерел підтверджує наявність складної, актуальної та перспективної проблематики. Результати огляду сформуvalи основу для уточнення завдань дослідження, зокрема націлившись на розробку підходів до адаптації LLM, які забезпечать стилістичну, культурну та контекстуальну адекватність перекладу художніх текстів.

3 ПОСТАНОВКА ЗАДАЧІ

3.1 Постановка задачі

Основна задача дослідження полягає в розробці підходів та методів адаптації LLM, які дозволять підвищити якість перекладу, зберігаючи стилістичну відповідність, культурну та контекстуальну адекватність твору.

Потрібно розробити та обґрунтувати ефективний підхід до адаптації великих мовних моделей для задачі художнього перекладу, який забезпечить достатню адекватність передачі змісту та мінімізує редагування людиною в порівнянні з машинним перекладом.

Потрібно створити стратегію, яка дозволить підлаштувати модель під переклад текстів, зокрема через введення стилістичних інструкцій та роботи з текстом.

У межах дослідження планується підготувати корпус сторінок книги, а також штучні фрагменти тексту. Для перевірки роботи з контекстом фентезі-текстів, де часто зустрічаються вигадані назви, імена персонажів чи унікальні топоніми, передбачається додатковий аналіз того, як модель відтворює та узгоджує такі елементи перекладу.

У рамках дослідження виконуватиметься запити до обраної моделі, а також експерименти по передачі контекстної інформації для перевірки, чи зможе такий підхід точніше зберегти узгодженість термінів.

Потрібно провести серію експериментів: спочатку без передачі контексту (як базовий контрольний варіант), а потім з різними стратегіями передачі контексту.

Здійснити порівняння результатів з результатами роботи базової моделі без передачі додаткового контексту, для оцінки переваг та недоліків запропонованого підходу. Оцінити вплив параметрів на якість та ресурсомісткість.

Окрім того, буде проведено збір статистики під час виконання дослідження: вимірювання часу.

На основі отриманих результатів запропонувати практичні рекомендації щодо адаптації LLM для художнього перекладу.

3.2 Необхідні ресурси

Як базу для моделей планується використати відкриті варіанти з Hugging Face [8] (див. рис. 4), оскільки вони доступні під ліцензіями, що дозволяють локальний запуск і дослідження.

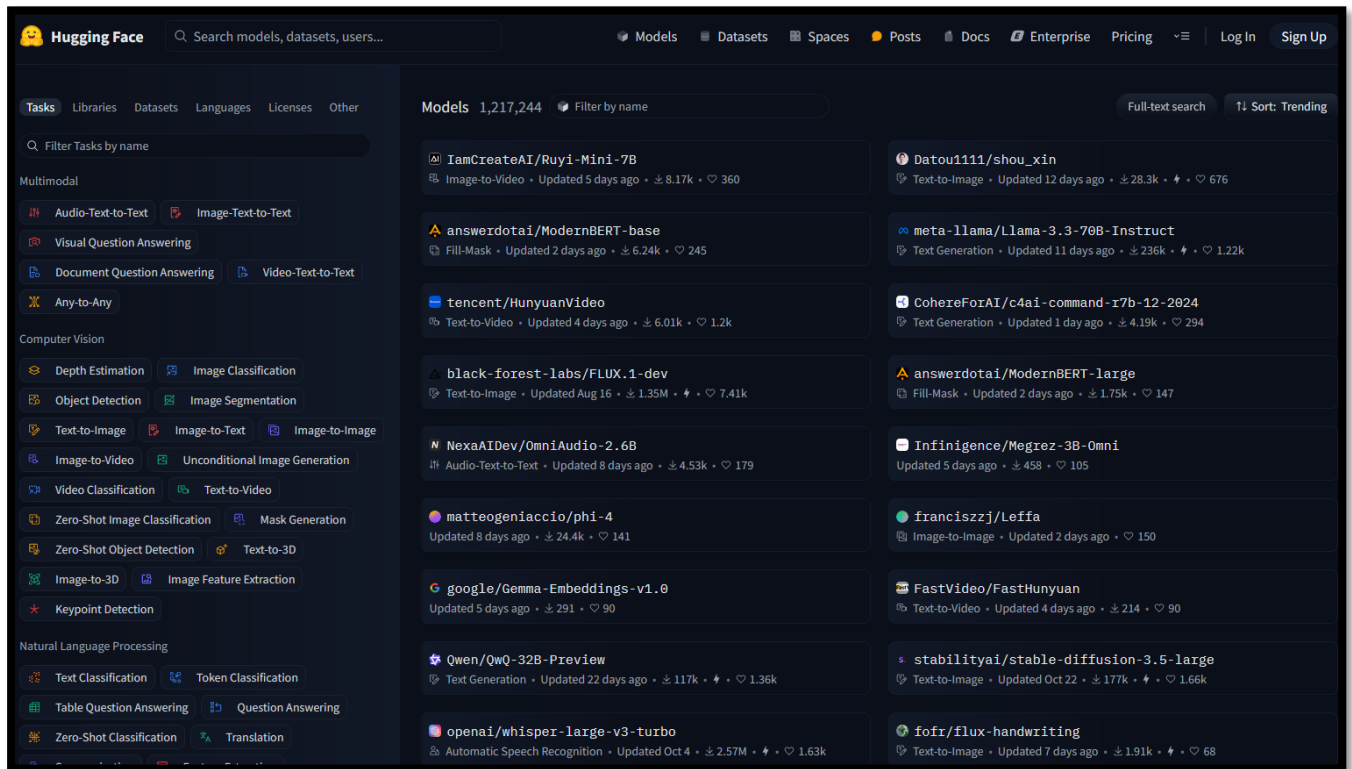


Рисунок 4 – Список доступних моделей на ресурсі Hugging Face (рисунок створено самостійно)

Зокрема LLaMA чи Mistral як основу для експериментів. Для оцінки якості перекладу передбачено застосувати метрику BLEU та також додаткову ручну перевірку результатів, щоб з'ясувати, наскільки модель зберігає контекст.

Для отримання англomовних джерел творів, які перебувають у суспільному надбанні та не порушують авторські права, буде використовуватися ресурс Project Gutenberg [9] або доступні для перегляду фрагменти Google Books. На цьому сайті викладені тексти, термін дії авторських прав на які вже сплив, і їх можна вільно використовувати для навчальних чи експериментальних цілей.

3.3 Можливі обмеження

Якість відповідей моделі залежить від обсягу та якості інформації стосовно потрібної мови, українська мова представлена не так сильно як інші. Потрібно знайти адекватно великі та якісні моделі, які повинні підтримувати декілька мов та бути інструктованими, що може бути складно.

Стилістична адекватність та культурна відповідність не завжди піддаються формалізації. Що призводить до того, що експертні оцінки можуть різнитися, що ускладнює об'єктивне вимірювання успіху.

Повний fine-tuning LLM на великих корпусах є ресурсомістким, тому доцільно використовувати вже навчені моделі Hugging face. Хоча використання LoRA та інших легковагових адаптерів знижує вимоги, провайдери працюють загалом з популярними моделями і не дозволяють безкоштовне навчання. Через високе навантаження на систему при роботі з LLM є потреба у використанні облачних інструментів Hugging Face Hub для проведення досліджень у купі з провайдерами.

4 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

4.1 Архітектура та проектування програмного забезпечення

Передбачається розробка комплексної системи, яка включатиме наступні компоненти (див. Рис. 5) :

- модуль лінгвістичної підготовки / BookNLP-аналіз;
- модуль моделі;
- модуль перекладу;
- модуль оцінки якості;
- інтерфейс користувача.

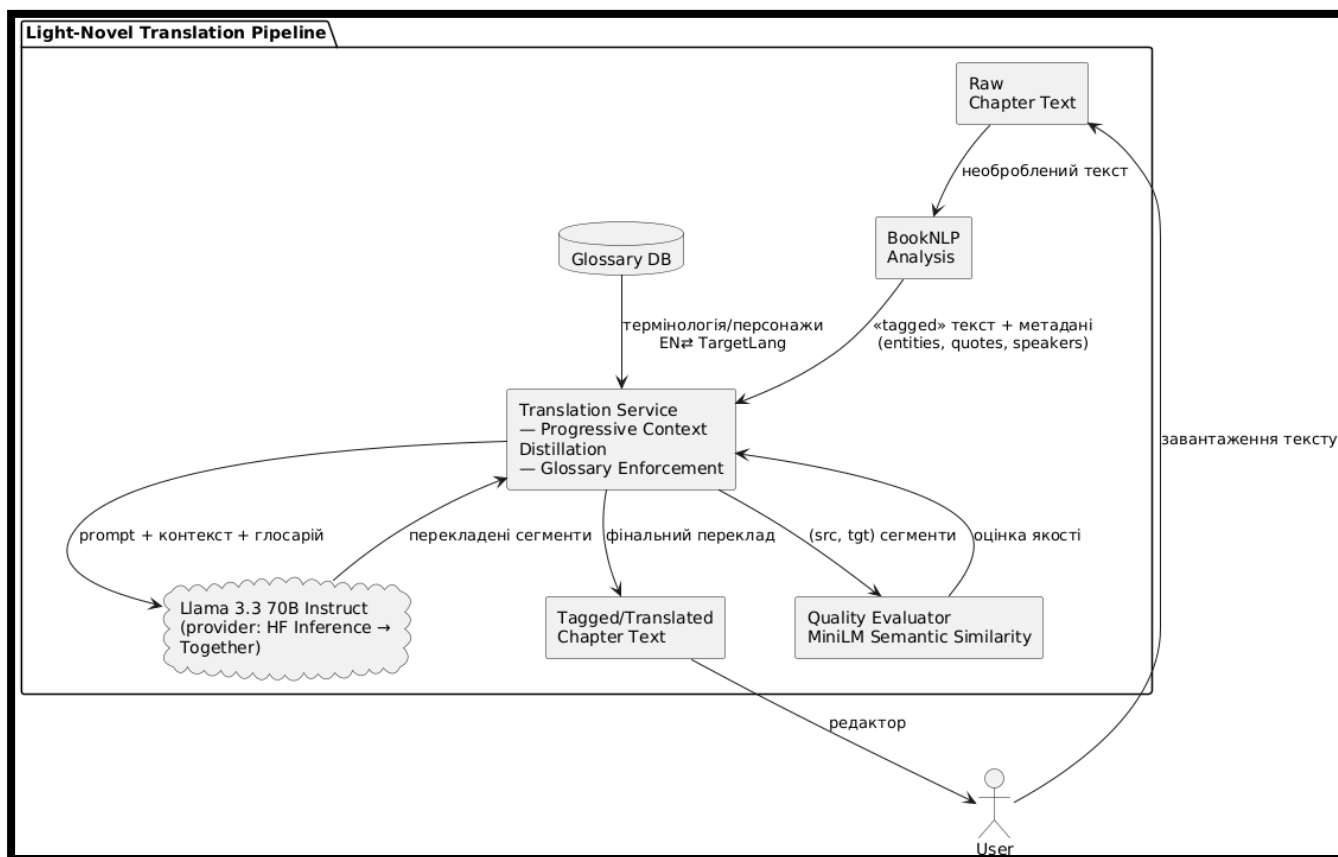


Рисунок 5 – Діаграма архітектури (рисунок створено самостійно)

Діаграма відображає потік даних від початкового завантаження корпусів до процесу оцінювання перекладу.

Модуль модуль лінгвістичної підготовки виконує збирання, фільтрацію та попередню обробку тексту через бібліотеку BookNLP, яка дозволяє знайти іменовані сутності та зв'язки між ними, дослідити які звертання використовуються до кожної сутності. Він включає в себе засоби для нормалізації тексту, виділення іменованих сутностей, зберігання інформації у json та можливість виправлення помилок.

Модуль моделі застосовує провайдер Together та HF inference API для роботи з моделю через видалений сервер. Надає можливість обирати адаптовані LLM, додає до існуючої моделі можливість працювати з контекстом, без потреби у повному перенавчанні моделі, зберігає та завантажує налаштовані версії моделей при наявності достатньої обчислювальної техніки.

Модуль перекладу отримує на вхід текст оригіналу й формує запити до адаптованої LLM, враховуючі інструкції, підказки і контекст. Генерує переклад, забезпечуючі збереження контекстних елементів.

Модуль оцінки якості вимірює метрики та автоматизовані показники для оцінки, а саме показник Semantic Similarity, який дозволяє перевірити семантичну схожість з людським перекладом.

Інтерфейс користувача надає можливості завантаження оригінального тексту, запуску процесу перекладу та перегляду результатів. Забезпечує інтерфейси для аналізу, порівняння різних версій перекладу та експорту глосарію.

4.2 Use Case діаграма

Use Case діаграма (див. рис. 6) відображає взаємодію між різними акторами (користувачами системи) та основними функціональними можливостями (варіантами використання) у процесі використання великих мовних моделей (LLM) для перекладу художніх текстів.



Рисунок 6 – Use case діаграма (рисунок створено самостійно)

Дослідник – це основний користувач, який готує корпуси текстів, виконує налаштування моделі, обирає різні моделі, аналізує результати якості за може зберігати і завантажувати готові моделі.

Редактор/Перекладач – це користувач, який запускає переклад в інтерфейсі користувача, перевіряє якість отриманого перекладу, надає зворотній зв'язок.

Система – це сутність, яка виступає внутрішнім сервісом, який дозволяє змінювати базові налаштування системи та локальні параметри.

4.3 Методи та алгоритми

Fine-tuning – це процес донавчання великої моделі, вже попередньо натренованої на загальних корпусах, на конкретному меншому корпусі художніх текстів (див. рис. 7). Він дозволяє моделі засвоїти специфічні патерни, лексику, стилі та культурні особливості, притаманні цільовій літературі. У більшості сучасних моделей (наприклад, LLaMA 3.3 70B Instruct) fine-tuning включає два етапи: supervised fine-tuning (SFT), де модель навчається на парі “вхід – бажаний вихід”, та reinforcement learning with human feedback (RLHF), де підкріпленням виступають людські оцінки якості відповіді. Це дозволяє краще узгодити поведінку моделі з людськими очікуваннями щодо корисності та безпечності.

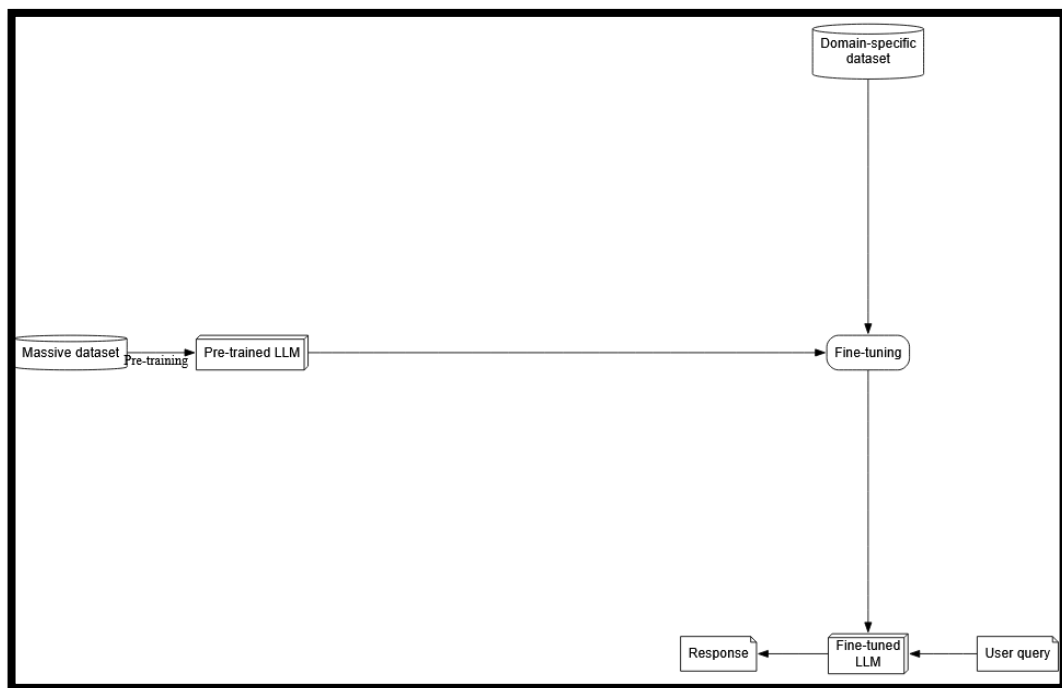


Рисунок 7 – Принцип роботи fine-tuning (рисунок створено самостійно)

Під час донавчання модель поступово пристосовується до нової задачі, зберігаючи при цьому загальні знання про мову.

Low-rank adapters – цей підхід дозволяє додати до фіксованої моделі невеликий набір параметрів. Ідея полягає в тому, що модель вже містить багатий набір загальних знань, і нам потрібно лише м’яко її підштовхнути до потрібного

стилю чи жанру. LoRA розкладає ваги шарів на низькорозмірні матриці, які потім коригуються, мінімізуючи зміну основних параметрів моделі.

Використовується принцип додавання низькорангових матриць до заморожених попередньо натренованих ваг W (див. рис. 8). Ці ваги залишаються незмінними, а матриці (A та B) додаються до них, формуючі оновлені ваги, одна з матриць ініціалізується випадковими значеннями з нормального розподілу для уникнення упередженості в напрямі адаптації на ранніх етапах навчання, інша є нульовою на початку навчання, що гарантує, що внесок адаптаційних матриць починається лише після початку навчання, зменшуючи ризик втрати корисної інформації у вагах.

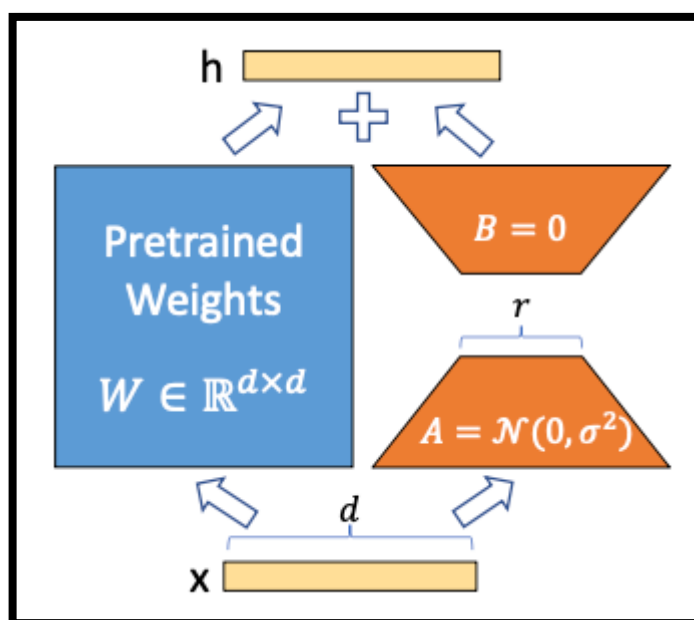


Рисунок 8 – Принцип роботи LoRA-адаптеру (за даними [10])

Це дозволяє економити пам'ять та обчислювальні ресурси [10], швидко адаптуватися до нових стилістичних вимог, перемикатися між різними жанрами, маючи кілька LoRA-адаптацій для однієї й тієї ж базової моделі.

Одним із найпоширеніших способів кількісної оцінки якості машинного перекладу є метрика BLEU (Bilingual Evaluation Understudy), що розраховується за формулою 5.1:

$$BLEU = BP * \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (5.1)$$

де BP (Brevity Penalty) – це штраф за занадто короткі переклади,

p_n – це частота збігу n -грам між перекладом і референсом,

w_n – це вагові коефіцієнти для n -грам,

N – максимальна довжина n -грам.

Її сутність полягає у порівнянні n -грам перекладеного фрагмента з відповідними n -грамами в еталонному (людському) перекладі. Висока частота збігів свідчить про ближчий до оригіналу зміст та лексичне наповнення. Перевагою BLEU є можливість отримати об'єктивний і швидкий показник якості перекладу, що дає змогу проводити масштабні тести без безпосередньої участі експертів.

Однак у контексті художніх творів навіть високий показник BLEU не завжди гарантує збереження якісного перекладу тексту. Тому додатково застосовується експертна оцінка (див. рис. 9), суть якої проаналізувати, наскільки повно і точно збережено контекст ситуації та культурні нюанси перекладу. Такий підхід дозволяє виявити специфічні помилки або недоліки, що не відображаються у статистичних метриках, але мають вагоме значення для адекватного сприйняття художнього тексту.

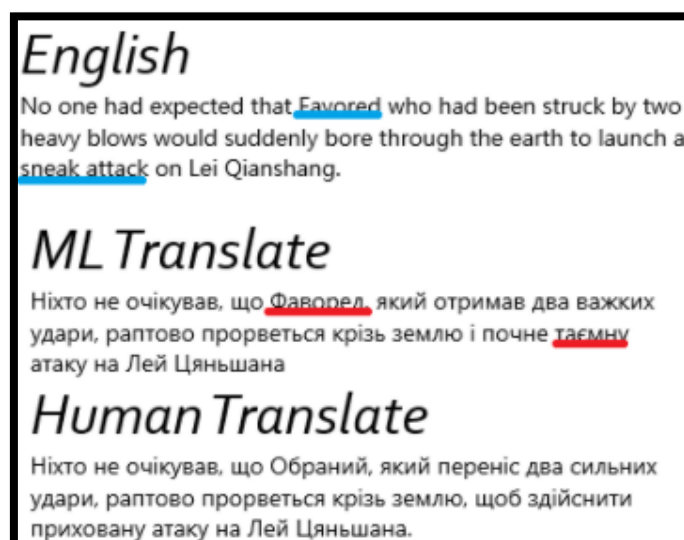


Рисунок 9 – Приклад аналізу тексту (рисунок створено самостійно)

Машинний переклад Google Translate залишив «Фаворед», правильний переклад адаптує назву титулу для кращої передачі тексту, у контексті атаки краще використовувати «прихована», а не «таємна». Експертна оцінка також краще враховує особливості, які можуть наблюдатися при перекладі специфічних термінів (див. рис. 10).



Рисунок 10 – Приклад аналізу при перекладі специфічного терміну (рисунок створено самостійно)

Машинний переклад Google Translate перекладає «тайцзи» одним словом, як найпоширеніший варіант перекладу, що не є помилкою. Проте певна література [11] може вимагати роздільного перекладу «Тай Дзи» для передачі акценту на етимології. Експертний підхід у аналізі фрагментів на культурну адекватність є суб'єктивним, проте важливим для перекладу художніх текстів.

4.4 Розробка інтерфейсу

При розробці користувацького інтерфейсу потрібно забезпечити зручний та інтуїтивно зрозумілий дизайн інтерфейсу.

Інтерфейс головної сторінки (див. рис. 11) повинен бути простим, користувач повинен мати змогу завантажити вихідний текст, обрати мову оригіналу та цільову мову перекладу, вказати стилістичні вимоги.

The wireframe shows a rectangular layout with a dark blue header bar. The header is divided into three sections: 'Title' on the left, 'NLP Analysis Result' in the center, and 'Stats' on the right. Below the 'Title' section is a white rectangular input field. Below the 'Text' section is another white rectangular input field. The 'NLP Analysis Result' and 'Stats' sections are currently empty white areas.

Рисунок 11 – Макет головної сторінки (рисунок створено самостійно)

На головній сторінці системи передбачається форма заповнення інформації про книгу, форма аналізу результатів, яка буде показуватися після виконання обчислень результатів з можливістю виправити інформацію або зберегти глосарій.

Аналогічно потрібно створити сторінку що дозволить порівняти 2 тексти та отримати результат їх схожості. На сторінки порівняння повинно бути присутньо 2 поля для введення тексту, кнопка початку аналізу, та малюватися графік після виконання аналізу.

Потрібно також створити сторінку перекладу тексту, на якій можливо буде вказувати потрібну велику мовну модель, мову перекладу, вводити текст та додавати глосарій, отриманий у попередньому аналізі. Повинна бути функція інтелектуального дослідження глосарію та виправлення за допомогою мовної моделі і функції штучного редагування.

Повинна бути присутня кнопка початку перекладу та чекбокс Context Distillation, при наявності якого після перекладу буде виведено додаткову інформацію про контекст та ситуацію на кожному блоці аналізованого тексту.

5 ПРАКТИЧНЕ ДОСЛІДЖЕННЯ

5.1 Опис експериментальних даних

Для експериментів, проведених у цьому дослідженні, були використані уривки з фентезійного літературного твору, написаного англійською мовою. Жанр фентезі було обрано свідомо, оскільки він часто включає вигадані локації, персонажів та спеціалізовану термінологію. Такий вибір дав змогу здійснити ґрунтовну оцінку узгодженості перекладу, особливо у випадках складної термінології, характерної для конкретного домену.

Було обрано два різні уривки: один мав переважно наративний характер із невеликою кількістю персонажів, а інший містив велику кількість діалогів між багатьма персонажами. Обидва уривки були перекладені з англійської мови трьома мовами: французькою, німецькою та українською.

Переклади, створені людиною цими мовами, були зібрані та використані як референтні переклади для цілей оцінювання.

Для автоматичного перекладу було використано базовий підхід, заснований на локально розгорнутій великій мовній моделі (LLM) Meta-Llama 3.3-70B-Instruct [12]. Ця багатомовна модель використовує оптимізовану трансформерну архітектуру з навчанням під наглядом (SFT) і навчанням з підкріпленням на основі зворотного зв'язку від людини (RLHF), що спеціально призначена для ефективної роботи з багатомовними задачами генерації тексту. Локальне розгортання моделі було свідомим вибором, що дозволяє уникнути залежності від зовнішніх API-сервісів і забезпечує більший контроль над процесом перекладу. Meta-Llama 3.3 підтримує кілька мов, зокрема французьку та німецьку, які добре представлені в навчальному наборі моделі. Українська мова не входить до переліку офіційно підтримуваних і менш представлена в навчальних даних.

Усі вихідні тексти та переклади зберігалися у форматі простого тексту (.txt). Глосарії зберігалися у форматі JSON.

5.2 Опис реалізованих методів

У цьому підрозділі описуються запропоновані методи перекладу, спрямовані на покращення узгодженості в перекладах художніх текстів. Дослідження пропонує дві взаємодоповнюючі техніки: підхід із використанням глосаріїв та метод поступової дистиляції контексту (Progressive Context Distillation), а також їхнє комбіноване застосування.

Техніка інтеграції глосарію передбачає двоетапний процес для досягнення термінологічної узгодженості. На першому етапі велика мовна модель (Meta-Llama 3.3-70B-Instruct) автоматично ідентифікує ключові терміни, що є специфічними для обраного домену, у тексті оригіналу. Це охоплює розпізнавання вигаданих термінів, імен персонажів із зазначенням їхньої статі, а також назв важливих локацій із короткими описовими примітками. Після цього ці терміни отримують точні переклади на основі людських референтних перекладів.

Структурований глосарій із перевіреними перекладами створюється окремо для кожної цільової мови та зберігається у форматі JSON.

Під час перекладу глосарій використовується як контекстуальне доповнення, що явно надається мовній моделі. Текст поділяється шляхом семантичного групування, де об'єднуються цілі абзаци, аби мінімізувати втрату контексту. Терміни з підготовленого глосарію та їхні відповідники у цільовій мові спрямовують процес перекладу, забезпечуючи точну та узгоджену передачу спеціалізованої термінології упродовж усього тексту.

Щоб зберегти наративну цілісність у перекладі художнього тексту, особливо щодо місць дії та подій, було запропоновано метод поступової дистиляції контексту.

Цей метод передбачає систематичне вилучення та оновлення контекстної інформації протягом усього процесу перекладу. Контекст оновлюється і використовується ітеративно, що дає змогу зберігати безперервність і узгодженість у передаванні наративних елементів.

Метод розраховується за формулами 6.1 та 6.2:

$$CTX_i = \text{UpdateContext}(CTX_{i-1}, C_i) \quad (6.1)$$

$$T_i = \text{Translate}(C_i, CTX_i) \quad (6.2)$$

де C_i – поточний текстовий фрагмент, що перекладається, а CTX_{i-1} , – це контекст, дистильований із попереднього фрагмента. Функція `UpdateContext` поєднує попередній контекст із новим фрагментом, створюючи оновлений контекст CTX_i . Цей новий контекст разом із C_i передається до функції `Translate`, результатом якої є перекладений фрагмент T_i .

5.3 Опис реалізованих метрик

Для об'єктивної оцінки якості перекладу в цьому дослідженні використовуються кілька автоматичних метрик: BLEU, CHRF++, семантична подібність і нова метрика контекстної узгодженості (`Context Consistency`, або `CoCon`).

BLEU – це одна з найпоширеніших метрик оцінювання машинного перекладу, яка вимірює точність на основі збігу n-грамів між перекладеним текстом і референтом. Незважаючи на свою популярність, BLEU сильно залежить від точного збігу слів та порядку, що знижує його ефективність у контексті художніх перекладів, де стилістичні варіації є нормою.

CHRF++ оцінює якість перекладу на основі F-міри для символічних n-грамів між кандидатним і референтним текстами. Ця метрика краще пристосована до гнучкого порядку слів, ніж BLEU, однак також має обмеження, пов'язані з переказами та стилістичними змінами, притаманними літературі. Як і BLEU, вона може недооцінювати семантично правильні переклади, які структурно відрізняються від оригіналу.

Семантична подібність дозволяє оцінювати переклади з точки зору змісту, а не лише формальних збігів. У цьому дослідженні використовувалися результати,

згенеровані за допомогою Sentence-BERT (модель: paraphrase-multilingual-MiniLM-L12-v2) [13]. Для вимірювання семантичної близькості між перекладом і референтом застосовувалася косинусна подібність. Такий підхід дозволяє точно передати нюанси значення і вважається найбільш придатним серед автоматизованих методів оцінювання художнього перекладу.

Для кількісного вимірювання узгодженості вживання глосарних термінів у перекладі запропонована спеціальна метрика Context Consistency (CoCon) що розраховується за формуламию 6.3 та 6.4:

$$CoCon = \frac{1}{N} \sum_{j=1}^N F1(term_j) \quad (6.3)$$

$$F1(term_j) = \frac{2 \times Precision(term_j) \times Recall(term_j)}{Precision(term_j) + Recall(term_j)} \quad (6.4)$$

Для кожного терміна глосарію точність (Precision) визначається як відношення кількості коректних нечітких збігів у машинному перекладі до загальної кількості можливих згадувань терміна в перекладі. Повнота (Recall) визначається як відношення кількості коректних згадувань у перекладі до кількості згадувань у людському перекладі. Для виявлення збігів використовується нечітке зіставлення (fuzzy matching), яке враховує орфографічні варіації, відмінювання й інші лінгвістичні відхилення та правильно передано ключову термінологію між різними частинами тексту.

Усі розглянуті метрики разом формують комплексну систему оцінки якості перекладу. Однак, незважаючи на досягнення в галузі автоматизованих підходів, людський аналіз залишається незамінним для повноцінного оцінювання якості художнього перекладу через тонкі стилістичні й семантичні нюанси, які важко формалізувати.

5.4 Результати експериментів

Експерименти було проведено з використанням моделі Meta-Llama 3.3-70B-Instruct через Hugging Face Inference API та інфраструктуру Together AI, що надала серверні ресурси для виконання перекладів. Загалом було згенеровано переклади для двох уривків художнього тексту: один із них був орієнтований на наратив, інший складався переважно з діалогів. Наративний уривок було розбито на 6 фрагментів, а діалоговий на 18 фрагментів. Кожен із них було перекладено трьома мовами (французькою, німецькою та українською) за допомогою чотирьох різних методів перекладу:

- базовий переклад LLM (Default)
- переклад LLM з використанням глосарію (Glossary)
- переклад LLM із застосуванням дистиляції контексту (Context)
- переклад LLM із поєднанням глосарію та контексту (Context+Glossary)

Для об'єктивного оцінювання результатів було розраховано усереднені показники для кожного методу на основі обох типів уривків нарративного та діалогового. Як і передбачалося, стандартні автоматизовані метрики BLEU показали відносно низькі оцінки (див. рис. 12) для всіх методів перекладу через обмеження цих метрик у випадках семантично точного, але структурно варіативного художнього перекладу.

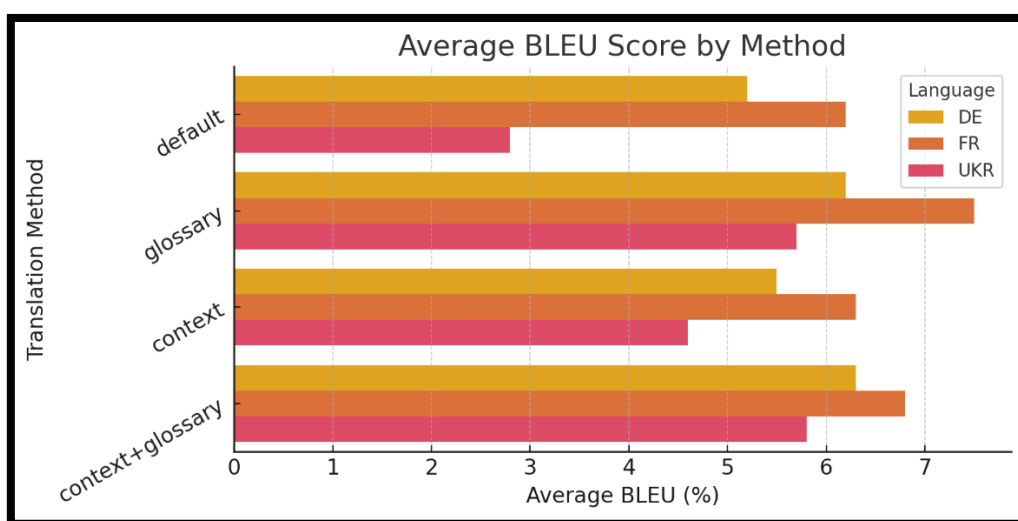


Рисунок 12 – Середні значення BLEU (рисунок створено самостійно)

Кращі результати демонструє метод CHRF++ (див. рис. 13), також є приріст для певних мов при використанні різних методів, але цей результат також не враховує семантичну близькість.

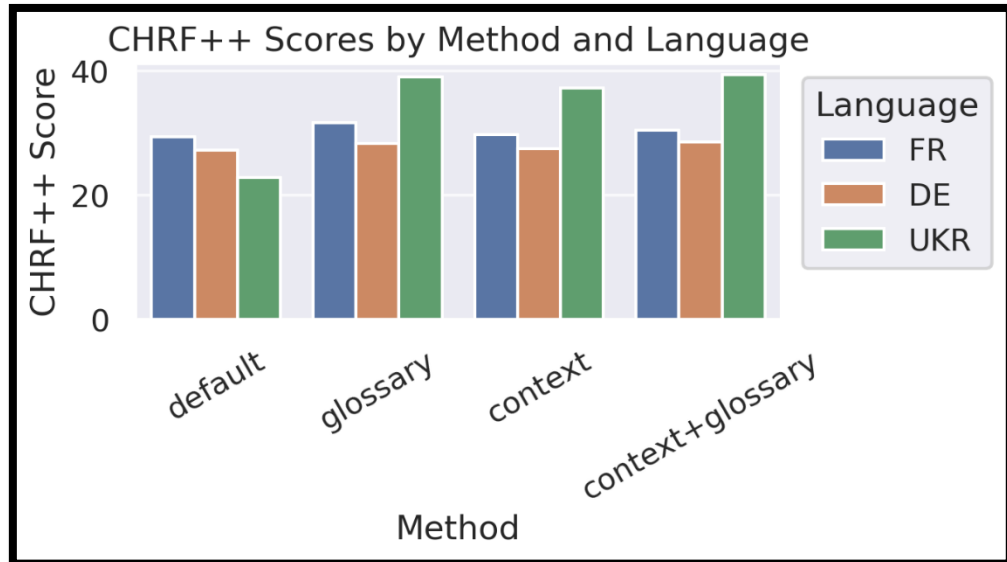


Рисунок 13 – Середні значення CHRF++ (рисунок створено самостійно)

На рисунку 14 наведено приклад, що наочно демонструє, чому стандартні оцінки на прикладі BLEU залишаються низькими навіть за умов семантично точного перекладу.

Sentence 1	BLEU	Text
Source	-	Later, it was said the man came from the north, from Ropers Gate. He came on foot, leading his laden horse by the bridle. It was late afternoon and the ropers', saddlers' and tanners' stalls were already closed, the street empty. It was hot but the man had a black coat thrown over his shoulders. He drew attention to himself.
Human	100.00	Потім казали, що чоловік той прийшов у місто з півночі, через браму Линварів. Йшов пішки, а нав'юченого коня вів за повід. Стояло пізно пообіддя, і крам линварів та римарів було замкнено, а вуличка була порожньою. Було тепло, але чоловік той мав напнутого на плечі чорного плаща. Чим привертав увагу.
Machine	10.17	Пізніше казали, що чоловік прийшов з півночі, з Брами Линварів. Він прийшов пішки, ведучи свого в'яучого коня за повод. Було пізно післяобіддя, і стійки ропарів, сідельників та шорників вже були закриті, вулиця була порожня. Було жарко, але чоловік мав чорний плащ, накинтий через плечі. Він привертав до себе увагу.

Рисунок 14 – Приклад обчислення оцінки BLEU (рисунок створено самостійно)

Як видно з ілюстрації, BLEU карає переклади за зміну порядку слів і лексичні варіації, навіть якщо переклад залишається семантично правильним. Метрика семантичної подібності (див. рис. 15) краще відображає реальну якість перекладу, адже не залежить від точного порядку слів і граматичної структури.

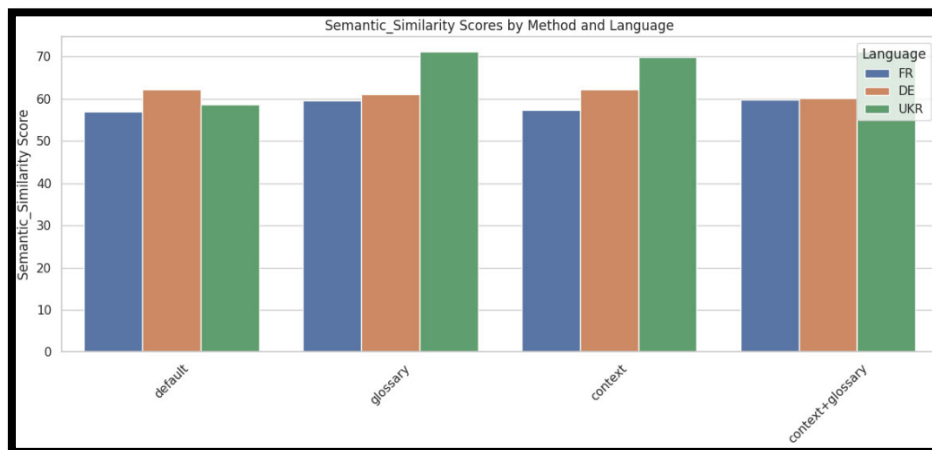


Рисунок 15 – Семантична подібність методів (рисунок створено самостійно)

Результати свідчать, що всі удосконалені методи (Glossary, Context, Context+Glossary) загалом демонструють вищі показники, ніж базовий переклад LLM. Зокрема, поєднання контексту й глосарію стабільно дає стабільні оцінки семантичної подібності для всіх цільових мов.

Теплова карта на рисунку 16 ілюструє, що методи з використанням глосарію, а саме Glossary та Context+Glossary, стабільно перевершують базову модель у всіх мовах при розрахунку метрики CoCon.

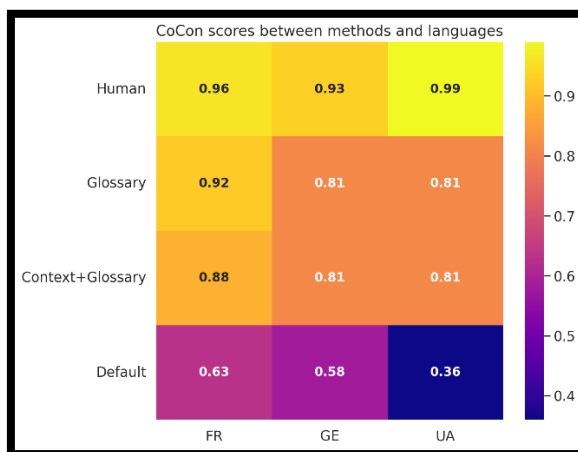


Рисунок 16 – Теплова карта методів (рисунок створено самостійно)

Ці методи демонструють значно вищі оцінки CoCon, наближені до показників людських перекладів, що підтверджує вирішальну роль глосарію у забезпеченні термінологічної точності. Цей аналіз зосереджується на фрагменті наративного типу, де послідовне передання спеціалізованих термінів є особливо важливим. Для порівняння, базовий метод LLM демонструє помітно нижчу якість, що підкреслює важливість інтеграції глосарних даних.

У таблиці, що зображена на рисунку 17 подано порівняльні результати оцінки: середні значення точності (precision), повноти (recall) та F1 для кожної мови й кожного методу перекладу.

Language	Method	Precision	Recall	F1
French	context	0.596	0.623	0.603
French	context+glossary	0.889	0.945	0.878
French	default	0.713	0.705	0.699
French	glossary	0.901	0.957	0.903
German	context	0.434	0.574	0.469
German	context+glossary	0.68	0.863	0.692
German	default	0.478	0.619	0.509
German	glossary	0.713	1	0.798
Ukrainian	context	0.558	0.511	0.509
Ukrainian	context+glossary	0.869	0.958	0.798
Ukrainian	default	0.583	0.506	0.509
Ukrainian	glossary	0.866	0.952	0.877

Рисунок 17 – Результати розрахунків показників (рисунок створено самостійно)

Результати підтверджують, що методи перекладу, які використовують інформацію з глосарію, демонструють вищі значення точності, повноти та середньої F1-міри в порівнянні з базовим методом. Ці висновки також підтверджують ефективність метрики CoCon на основі нечіткого зіставлення, яка здатна виявляти термінологічну узгодженість між частинами тексту, що особливо важливо для перекладу літературних творів зі спеціалізованою лексикою, та свідчать про успішну реалізацію системи для перекладу на основі LLM.

ВИСНОВКИ

У межах цього дослідження було реалізовано та експериментально перевірено підходи до налаштування великих мовних моделей (LLM) для задачі автоматизованого перекладу художніх текстів. Основну увагу зосереджено на збереженні стилістичної цілісності та термінологічної узгодженості перекладу. Запропоновані методи інтеграції глосарію та поступової дистиляції контексту продемонстрували свою ефективність. Результати експериментів підтвердили, що застосування контекстної дистиляції суттєво підвищує оцінки семантичної подібності в порівнянні зі стандартними запитами до мовної моделі, покращуючи нарративну зв'язність між фрагментами перекладеного тексту.

Інтеграція глосарної інформації значно підвищила термінологічну точність, забезпечивши послідовне вживання спеціалізованої лексики упродовж усього тексту. Методика CoCon, заснована на нечіткому зіставленні, дозволила кількісно оцінити цю узгодженість, що є критично важливим для перекладу фентезі та інших стилістично насичених жанрів.

Також було проведено аналіз сучасних підходів до адаптації LLM для художнього перекладу, зокрема технік донавчання (fine-tuning) та легковагових адаптерів (LoRA). Огляд наукових публікацій і експериментальних даних підтвердив ефективність LoRA за наявності достатнього корпусу паралельних текстів із багатим стилістичним варіюванням. На основі цього сформовано набір практичних рекомендацій щодо використання LLM у літературному перекладі.

Окрім цього, були запропоновані підходи для порівняння якості перекладу різних методів, зокрема, шляхом комбінованого використання метрик (BLEU) та експертної оцінки з акцентом на збереження авторського стилю та контекстної цілісності. Також хмарні сервіси для проведення масштабованих обчислень.

Таким чином, дослідження успішно досягло поставленої мети. Було показано, що великі мовні моделі, доповнені контекстними і термінологічно-орієнтованими механізмами, здатні генерувати переклади високої якості з точки зору змісту й стилістичної адекватності. Запропоновані підходи мають потенціал для покращення результатів у порівнянні з універсальними системами машинного

перекладу та закладають фундамент для подальших досліджень із використанням більших або донавчених моделей у сфері літературного перекладу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Statistical and neural approaches have permitted fast improvement in the quality of machine translation, but we are yet to discover how those. URL: https://www.researchgate.net/publication/331873130_Evaluation_of_SMT_and_NMT_Systems_a_study_on_Uses_and_Perceptions (дата звернення: 04.04.2025).
2. Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis. URL: https://www.researchgate.net/publication/369924463_Multilingual_Machine_Translation_with_Large_Language_Models_Empirical_Results_and_Analysis (дата звернення: 04.04.2025).
3. D. Dashenkov, K. Smelyakov and O. Turuta, “Methods of Multilanguage Question Answering,” 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, Ukraine, 2021, pp. 251-255, doi: 10.1109/PICST54195.2021.9772145.
4. D. Maksymenko, N. Saichyshyna, O. Turuta, O. Turuta, A. Yerokhin and A. Babii, “Improving the Machine Translation Model in Specific Domains for the Ukrainian Language,” 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 123-129, doi: 10.1109/CSIT56902.2022.10000529.
5. LoRA vs Full Fine-tuning: An Illusion of Equivalence. URL: https://www.researchgate.net/publication/385318527_LoRA_vs_Full_Fine-tuning_An_Illusion_of_Equivalence (дата звернення: 04.04.2025).
6. A Survey on Evaluation Metrics for Machine Translation. URL: https://www.researchgate.net/publication/368590740_A_Survey_on_Evaluation_Metrics_for_Machine_Translation (дата звернення: 04.04.2025).
7. Between Canon and Corpus. URL: <https://litlab.stanford.edu/LiteraryLabPamphlet8.pdf> (дата звернення: 04.04.2025).
8. Models – Hugging Face. URL: <https://huggingface.co/models> (дата звернення: 04.04.2025).
9. Free eBooks | Project Gutenberg. URL: <https://gutenberg.org/> (дата

звернення: 04.04.2025).

10. Understanding Low-Rank Adaptation (LoRA) for Efficient Fine-Tuning of Large Language Models. URL: <https://ai.plainenglish.io/understanding-low-rank-adaptation-lora-for-efficient-fine-tuning-of-large-language-models-082d223bb6db>

(дата звернення: 04.04.2025).

11. Tai Chi URL: <http://www.shenwu.com/taichi.htm> (дата звернення: 21.12.2024)

12. Hugging Face Meta LLaMA URL: <https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct> (дата звернення: 21.12.2024)

13. Sentence Transformers URL: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (дата звернення: 04.04.2025)

14. GitHub Repository URL: <https://github.com/AlexeyTrofimenko/alcea> (дата звернення: 08.06.2025)

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ
КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ**

3. D. Dashenkov, K. Smelyakov and O. Turuta, “Methods of Multilanguage Question Answering,” 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, Ukraine, 2021, pp. 251-255, doi: 10.1109/PICST54195.2021.9772145.

4. D. Maksymenko, N. Saichyshyna, O. Turuta, O. Turuta, A. Yerokhin and A. Babii, “Improving the Machine Translation Model in Specific Domains for the Ukrainian Language,” 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 123-129, doi: 10.1109/CSIT56902.2022.10000529.