

## ДОДАТОК А

## Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ

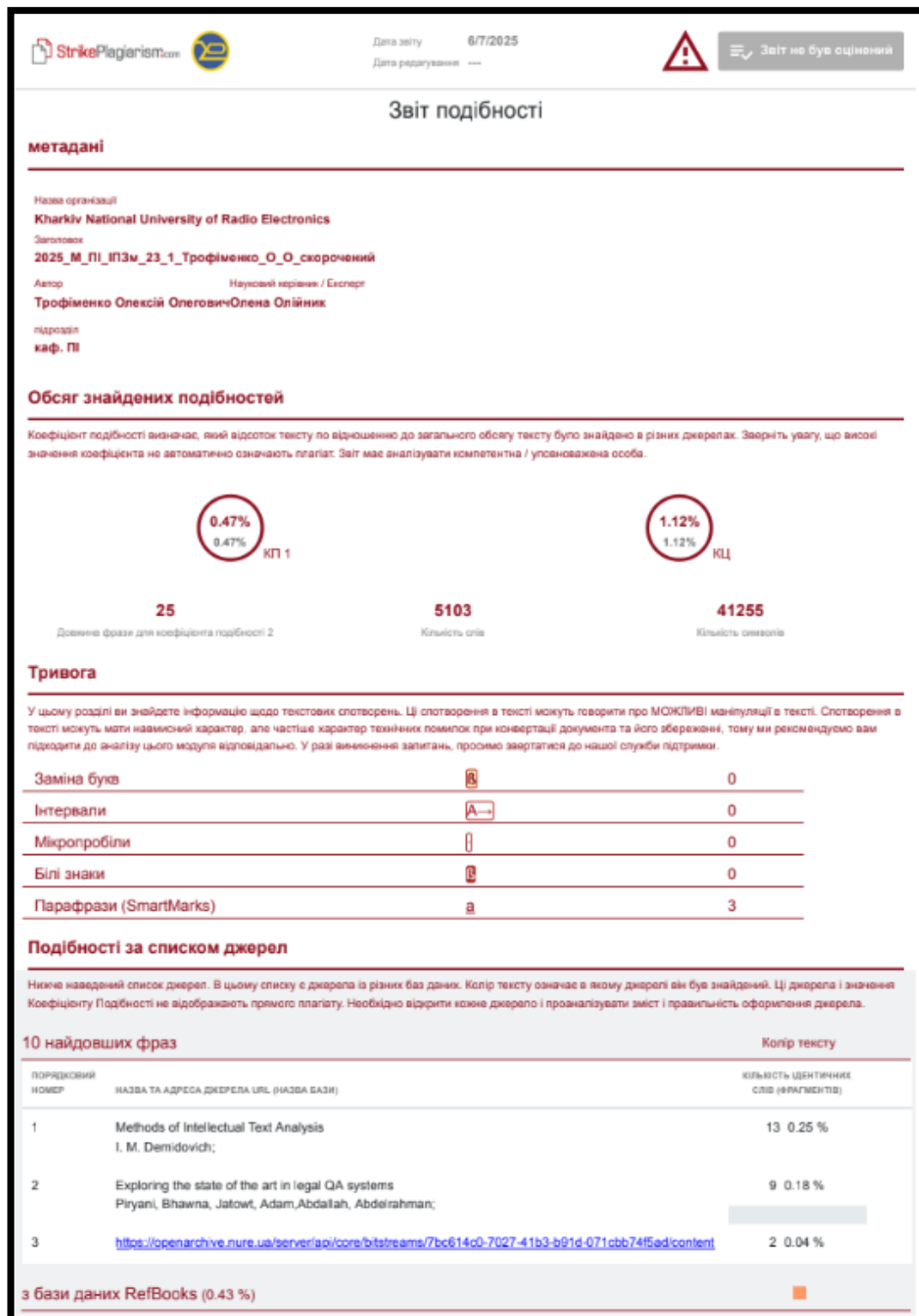




Рисунок А.1 – Звіт з перевірки роботи на унікальність (рисунок створено самостійно)

## ДОДАТОК Б


### Слайди презентації


 МІНІСТЕРСТВО  
ОСВІТИ І НАУКИ  
УКРАЇНИ


 ХАРКІВСЬКИЙ  
НАЦІОНАЛЬНИЙ  
УНІВЕРСИТЕТ  
РАДІОЕЛЕКТРОНІКИ

# Дослідження методів налаштування великих мовних моделей для перекладу художніх текстів

Трофіменко О.О., ІПЗм-23-1  
 Науковий керівник: доц. Чуприна А.С.


 13 червня 2025

1

Рисунок Б.1 – Слайд 1 (рисунок створено самостійно)

# Аналіз предметої галузі

 Speak of the devil

 Speak of Cao Cao and Cao Cao arrives





## Особливості художнього тексту

- Вигадана термінологія, стиль, культурні та мовні аспекти
- Складність збереження цих елементів у машинному перекладі

2

Рисунок Б.2 – Слайд 2 (рисунок створено самостійно)

## Аналіз предметої галузі

### Еволюція машинного перекладу

- Статистичні моделі
- Нейронні мережі
- Трансформери - LLM

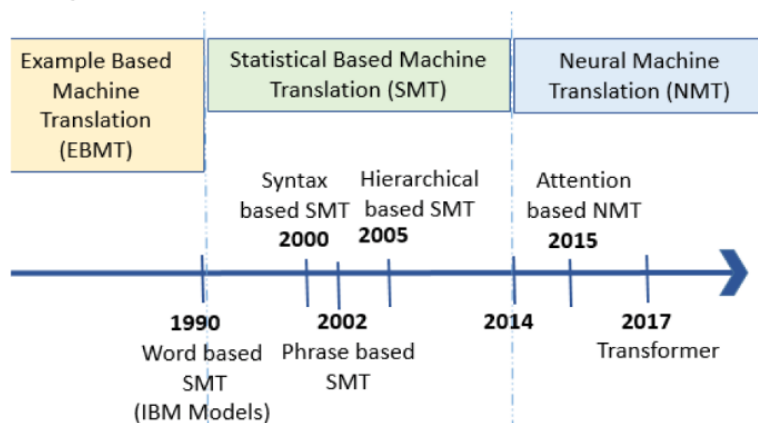


Рисунок Б.3 – Слайд 3 (рисунок створено самостійно)

## Огляд й аналіз літературних, наукових джерел

### Критерії відбору

- Авторитетність, актуальність
- Відкриті дані
- Відсутність російськомовних публікацій

IEEE Xplore®

ResearchGate

### Напрями досліджень

- Загальні підходи до машинного перекладу
- Адаптація великих мовних моделей
- Художній переклад
- Метрики оцінки

Рисунок Б.4 – Слайд 4 (рисунок створено самостійно)

## Огляд й аналіз літературних, наукових джерел

### Аналіз

- LLM мають потенціал для вирішення задачі перекладу
- Метрики BLEU, METEOR не враховують семантичну подібність

### Актуальність

- Потреба у швидкому та якісному художньому перекладі
- Застосування LLM для літератури поки що обмежене

### Висновки

- Необхідне рішення для термінологічної відповідності та розуміння контексту
- Поєднання автоматичних і ручних оцінок

Рисунок Б.5 – Слайд 5 (рисунок створено самостійно)

## Постановка задачі

### Hugging Face

LLM адаптована під задачі перекладу та підтримку інструкцій



Дослідити різні стратегії покращення якості перекладу та оцінити якість.



Порівняти отриманий переклад з людським.



Зробити висновки щодо практичного застосування.

Рисунок Б.6 – Слайд 6 (рисунок створено самостійно)

# Теоретичне дослідження

## Fine-tuning

- Загальна адаптація LLM на нових текстових корпусах
- Навчання на необов'язково анотованих даних
- Зберігає загальні знання моделі

## Supervised fine-tuning + Reinforcement Learning with Human Feedback

- Багатоступенева доадаптація моделі з урахуванням людських переваг
- Навчання на прикладах корисних відповідей
- Уточнення через оцінки якості людиною
- Покращує корисність, безпечність і узгодженість відповіді.



7

Рисунок Б.7 – Слайд 7 (рисунок створено самостійно)

# Теоретичне дослідження

## BLUE-метрика

- Порівнює n-грами перекладеного тексту з референсним перекладом.
- Підходить для швидкого автоматичного аналізу якості перекладу.

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

## Експертна оцінка

- Аналіз стилістики, культурної адекватності, контексту.
- Виявляє недоліки, що не враховуються метрикою BLEU.

**English**  
No one had expected that **Obama** who had been struck by two heavy blows would suddenly bore through the earth to launch a **sneak attack** on Lei Qianshang.

**ML Translate**  
Ніхто не очікував, що **Обама**, який отримав два важких удари, раптово прорветься крізь землю і почне **таємну** атаку на Лей Цяньшана

**Human Translate**  
Ніхто не очікував, що Обамний, який переніс два сильних удари, раптово прорветься крізь землю, щоб здійснити приховану атаку на Лей Цяньшана.

**Chinese original**  
平江主修为不算太高, 却同样是狄九任命的太版界五大庭柱之一。

**Chinese pinyin**  
Pingjiang shengxiu wei bu suan shi tai gao, que tongyang shi dijiu renming de taijiu jie wuda tingzhu zhizhi yizhi.

**Ukrainian ML**  
Рівень самовдосконалення Піна Цзяншена не надто високий, але він також є одним із п'яти стовпів світу **тайцзи**, призначених Ді Цяю.



8

Рисунок Б.8 – Слайд 8 (рисунок створено самостійно)

## Практичне дослідження. Методологія

Для оцінки якості художнього перекладу було використано велику мовну модель LLaMA 3.3-70B-Instruct з трьома цільовими мовами:

- Мови перекладу: французька, німецька, українська
- Вхідні дані: 2 вручну відібрані художні уривки
- Сегментація тексту:
  - 6 фрагментів наративу (описова проза)
  - 18 фрагментів діалогів (пряма мова та внутрішні думки персонажів)
- Кожен сегмент перекладався окремо залежно від підходу
- Оцінка якості включала семантичну зв'язність, відповідність глосарію



Рисунок Б.9 – Слайд 9 (рисунок створено самостійно)

## Практичне дослідження. Методологія



Рисунок Б.10 – Слайд 10 (рисунок створено самостійно)

## Практичне дослідження. Отримання глосарію

- Мета: забезпечити узгоджений переклад ключових термінів
- Витяг ключових термінів (імена, титули тощо) за допомогою аналізу BookNLP
- Машинний аналіз для видалення зайвого
- Вказання бажаного перекладу

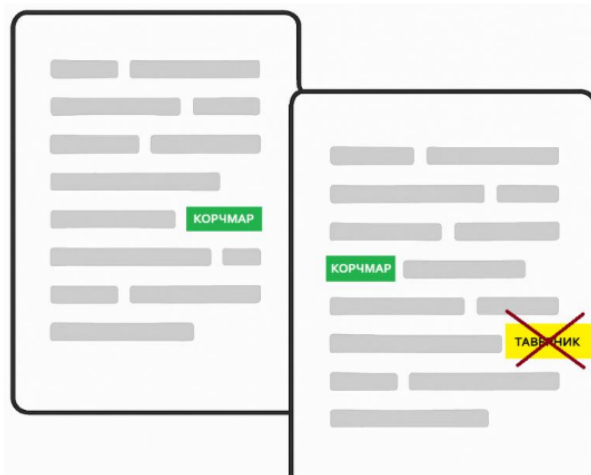
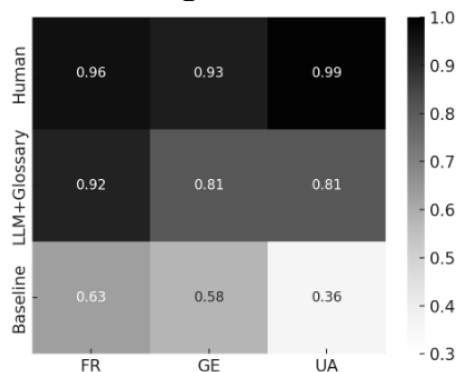
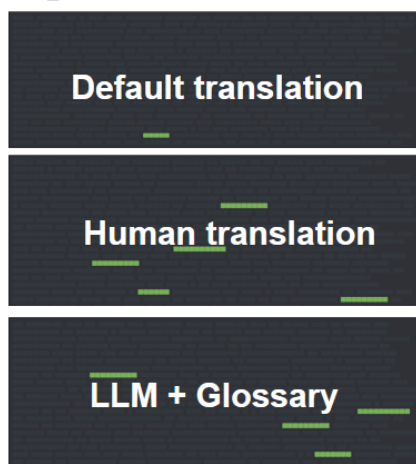


Рисунок Б.11 – Слайд 11 (рисунок створено самостійно)

## Практичне дослідження. Інтеграція глосарію



**+70.2%**

покращення правильного використання термінів порівняно з базовим перекладом

Рисунок Б.12 – Слайд 12 (рисунок створено самостійно)

## Практичне дослідження. Progressive context distillation

### Мета:

Поступово оновлювати знання про контекст (ситуацію, локацію) під час обробки тексту, щоб покращити переклад наступних фрагментів.

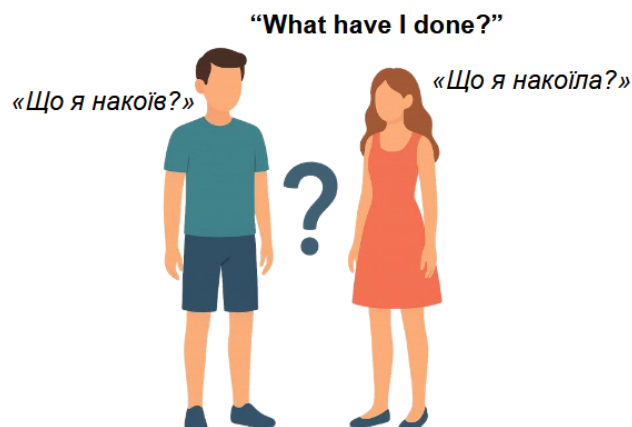
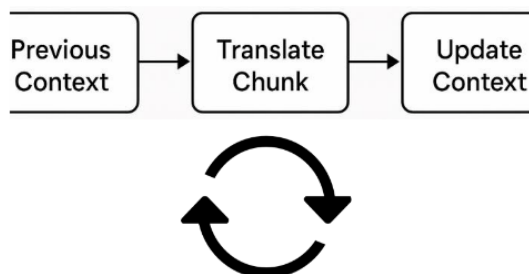
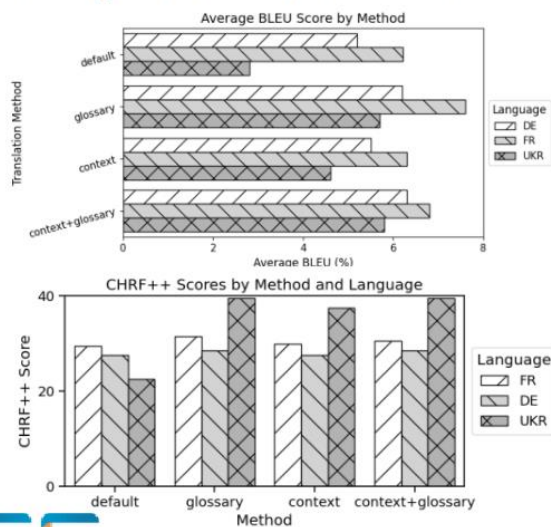


Рисунок Б.13 – Слайд 13 (рисунок створено самостійно)

## Метрики оцінювання: BLEU & CHRF++



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'яченого 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'яченого 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, накинутий через плечі. Він привертав до себе увагу.

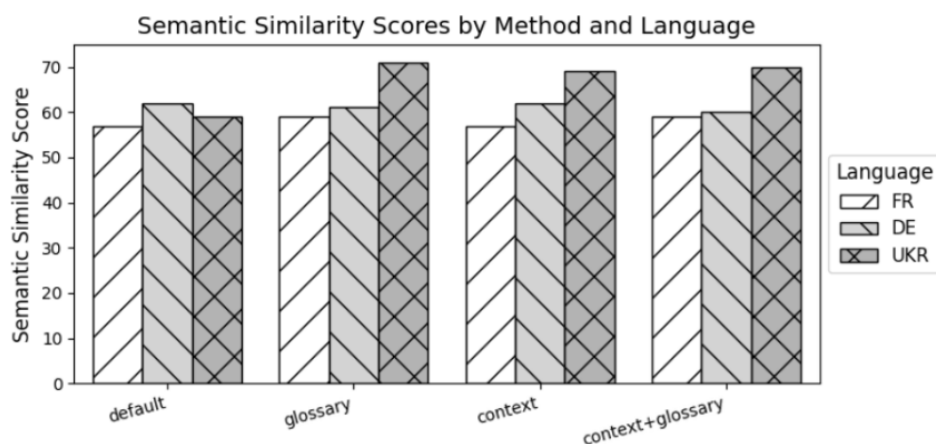
*BLEU штрафує переклади через зміни порядку слів та лексичні варіації, незважаючи на семантичну правильність.*

Рисунок Б.14 – Слайд 14 (рисунок створено самостійно)

## Метрика оцінювання: Semantic Similarity

- Передає зміст, а не лише дослівні формулювання
- Тolerує стилістичні варіації
- Підходить для парафразованого контенту
- Відображає людиноподібне розуміння
- Краще відповідає цілям художнього перекладу

Усі методи перевершують базовий переклад



15

Рисунок Б.15 – Слайд 15 (рисунок створено самостійно)



## Публікація результатів

Результати дослідження були представлені на  
**2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)**  
 та доступні в архіві IEEE Xplore

Exploring Strategies for Literary Translation Using Large Language Models

Publisher: IEEE

Oleksii Trofimenko; Serhiy Smelyakov; Anastasiya Chupryna; Zoltan Ducsar

**Abstract:**  
 Large language models (LLMs) have made significant progress in processing and generating text across multiple languages. However, translating long literary works remains challenging due to the need for consistent character interactions, specialized vocabulary, and coherence across chapters. This paper explores these difficulties and examines methods to achieve decontextualized LLM-based translation of literary texts. Various approaches are considered, including techniques for improving contextual awareness and integrating domain-specific vocabulary to reduce inconsistencies. The analysis highlights both the strengths and limitations of current methods, suggesting that targeted context management and fine-tuning strategies have the potential to improve translation accuracy in certain cases. These insights contribute to the development of more effective translation systems for literary texts and multilingual content.

Published in: 2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)

Authors

Date of Conference: 24-24 April 2025

DOI: 10.1109/eStream69908.2025.11918963

Figures

Date Added to IEEE Xplore: 02 June 2025

Publisher: IEEE

References

► ISBN Information:

Conference Location: Vilnius, Lithuania



16

Рисунок Б.16 – Слайд 16 (рисунок створено самостійно)

## Висновки

- Інтеграція глосарію забезпечує сталість термінології
- Дистиляція контексту покращує наративну зв'язність
- Комбінований підхід перевершує базові методи перекладу за допомогою LLM
- BLEU та CHRF++ не є достатніми для оцінки художньої якості
- Оцінка людиною залишається необхідною

Рисунок Б.17 – Слайд 17 (рисунок створено самостійно)

## ДОДАТОК В

### Апробація результатів роботи

Conferences > 2025 IEEE Open Conference of ... ?

## Exploring Strategies for Literary Translation Using Large Language Models

**Publisher:** IEEE Cite This PDF

Oleksii Trofimenko ; Serhii Smelyakov ; Anastasiya Chupryna ; Zoia Dudar [All Authors](#)

---

<p><b>Abstract</b></p> <hr/> <p>Document Sections</p> <ul style="list-style-type: none"> <li>I. Introduction</li> <li>II. Data Description</li> <li>III. Methods</li> <li>IV. Evaluation Metrics</li> <li>V. Experiments and Results</li> </ul> <p><a href="#">Show Full Outline</a> ▾</p> <hr/> <p><a href="#">Authors</a></p> <hr/> <p><a href="#">Figures</a></p> <hr/> <p><a href="#">References</a></p> <hr/> <p><a href="#">Keywords</a></p>	<p><b>Abstract:</b></p> <p>Large language models (LLMs) have made significant progress in processing and generating text across multiple languages. However, translating long literary works remains challenging due to the need for consistent character interactions, specialized vocabulary, and coherence across chapters. This paper explores these difficulties and examines methods to achieve decent-quality LLM-based translation of literary texts. Various approaches are considered, including techniques for improving contextual awareness and integrating domain-specific vocabulary to reduce inconsistencies. The analysis highlights both the strengths and limitations of current methods, suggesting that targeted context management and fine-tuning strategies have the potential to improve translation accuracy in certain cases. These insights contribute to the development of more effective translation systems for literary texts and multilingual content.</p> <p><b>Published in:</b> <a href="#">2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)</a></p> <p><b>Date of Conference:</b> 24-24 April 2025 <span style="float: right;"><b>DOI:</b> <a href="#">10.1109/eStream66938.2025.11016863</a></span></p> <p><b>Date Added to IEEE Xplore:</b> 02 June 2025 <span style="float: right;"><b>Publisher:</b> IEEE</span></p> <p><b>► ISBN Information:</b> <span style="float: right;"><b>Conference Location:</b> Vilnius, Lithuania</span></p> <p><b>▼ ISSN Information:</b></p>
--	---

Рисунок В.1 – Робота на сайті IEEE (рисунок створено самостійно)

# Exploring Strategies for Literary Translation Using Large Language Models

Oleksii Trofimenko <i>Software Engineering Department</i> Kharkiv National University <i>of Radio Electronics</i> Kharkiv, Ukraine oleksii.trofimenko@nurc.ua	Serhii Smelyakov <i>Software Engineering Department</i> Kharkiv National University <i>of Radio Electronics</i> Kharkiv, Ukraine serhii.smeliakov@nurc.ua	Anastasiya Chupryna <i>Software Engineering Department</i> Kharkiv National University of <i>of Radio Electronics</i> Kharkiv, Ukraine anastasiya.chupryna@nurc.ua	Zoia Dudar <i>Software Engineering Department</i> Kharkiv National University <i>of Radio Electronics</i> Kharkiv, Ukraine zoia.dudar@nurc.ua
--	--	---	--

**Abstract**—Large language models (LLMs) have made significant progress in processing and generating text across multiple languages. However, translating long literary works remains challenging due to the need for consistent character interactions, specialized vocabulary, and coherence across chapters. This paper explores these difficulties and examines methods to achieve decent-quality LLM-based translation of literary texts. Various approaches are considered, including techniques for improving contextual awareness and integrating domain-specific vocabulary to reduce inconsistencies. The analysis highlights both the strengths and limitations of current methods, suggesting that targeted context management and fine-tuning strategies have the potential to improve translation accuracy in certain cases. These insights contribute to the development of more effective translation systems for literary texts and multilingual content.

**Index Terms**—Large Language Models, translation of literary texts, contextual awareness, domain-specific vocabulary.

## I. INTRODUCTION

Large language models (LLMs) have significantly improved multilingual text generation and understanding. However, translating literary texts remains difficult due to the need for stylistic fidelity, coherent character interaction, and consistent narrative structure across long spans of text [1]. Unlike general texts, literary works often include figurative expressions, invented terminology, and shifting narrative perspectives. Standard neural machine translation (NMT) approaches typically process sentences in isolation, leading to inconsistencies in terminology, style, and reference tracking [2], [3]. This is especially problematic in genres like fantasy, where maintaining consistent renderings of names and fictional terms is critical.

Recent research has addressed these issues through context-aware architectures and document-level modeling. Hierarchical NMT systems and LLM prompt engineering techniques have shown promise in preserving intratextual coherence by maintaining evolving context across translation units [4]. In parallel, integrating glossaries or domain-specific constraints into the translation pipeline has improved consistency for specialized vocabulary [5].

Some languages, particularly Ukrainian, can introduce additional challenges [6] due to their limited representation in the training data for large language models [7].

Furthermore, evaluating the quality of literary translation requires more than BLEU or CHRF++. Studies have proposed using sentence embeddings, LLM-based evaluators, or human frameworks such as MQM to assess semantic preservation, fluency, and cultural adequacy [8], [9].

The study examines how Glossary Integration and Progressive Context Distillation can enhance the quality of LLM-based literary translation.

## II. DATA DESCRIPTION

The experiments conducted in this research used excerpts from a fantasy literary work written in English. The fantasy genre was specifically chosen because it frequently contains fictional locations, characters, and specialized terminology. This choice allowed for a thorough evaluation of translation consistency, particularly for challenging domain-specific vocabulary.

Two distinct excerpts were selected for the experiments: one predominantly narrative-oriented with a limited number of characters and another dialogue-heavy with multiple interacting characters. Both excerpts were translated from English into three languages: French, German, and Ukrainian.

Human-made translations into these languages were collected and used as reference translations for evaluation purposes.

For automated translation, the baseline approach used a locally deployable large language model (LLM) Meta-Llama 3.3-70B-Instruct [10]. This multilingual model employs an optimized Transformer architecture with supervised fine-tuning (SFT) and reinforcement learning with human feedback (RLHF), designed specifically to handle multilingual generation tasks effectively. The choice of a locally deployable model was deliberate, enabling independence from external API services and providing greater control over the translation process. Meta-Llama 3.3 supports multiple languages, including French and German, which are well represented in the model's training dataset. Ukrainian is not officially listed among the supported languages and is less represented in the training data.

Рисунок В.2 – Сторінка 1 (рисунок створено самостійно)

All source texts and translations were stored in plain text (.txt) format. The glossaries were stored in JSON format.

### III. METHODS

This section describes the proposed translation methods aimed at improving consistency in translations of literary texts. The research introduces two complementary techniques: a glossary-based translation approach and a Progressive Context Distillation method, together with their combined implementation.

#### A. Glossary Integration Technique

To achieve terminological consistency, this technique involves a two-stage process. Initially, the language model (Meta-Llama 3.3-70B-Instruct) automatically identifies key domain-specific terms within the source text. This includes recognizing fictional terms, character names along with their gender information, and important locations, accompanied by brief descriptive annotations. Subsequently, these terms receive precise translations based on human-generated reference translations. The resulting structured glossary, which contains verified translations, is stored separately for each target language.

When translating, the glossary serves as a contextual constraint explicitly provided to the language model. Text is segmented through semantic chunking, grouping entire paragraphs to minimize context disruption. The glossary terms and corresponding translations from the prepared JSON glossary guide the translation process, promoting an accurate and consistent rendering of specialized terminology throughout the text.

#### B. Progressive Context Distillation Method

Maintaining narrative coherence in literary translation, particularly with regard to locations and ongoing events, is critical. To address this, the proposed Progressive Context Distillation method (1) systematically extracts and updates contextual information throughout the translation process (2). Context extraction and updating occur iteratively, guiding the translation of subsequent text chunks by ensuring continuity and consistency of narrative information.

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

$$T_i = \text{Translate}(C_i, CTX_i) \quad (2)$$

where  $C_i$  represents the current text chunk being translated, while  $CTX_{i-1}$  denotes the distilled context of the previous chunk. The function `UpdateContext` combines this prior context with the current chunk to produce an updated context  $CTX_i$ . This new context is then passed along with  $C_i$  to the function `Translate`, resulting in the translated output  $T_i$ .

To ensure computational efficiency and remain within the model limitations, the length of  $CTX_i$  was capped at 500 tokens.

### IV. EVALUATION METRICS

To objectively evaluate translation quality, this study employs several automatic metrics: BLEU, CHRF++, Semantic Similarity, and a novel Context Consistency (CoCon) metric.

#### A. Translation quality assessment methods and their limitations

BLEU (3) is a widely used metric to evaluate the quality of machine translation by calculating precision based on the overlap of n-grams between candidate translation and reference translations.

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (3)$$

where  $BP$  is a brevity penalty,  $p_n$  is precision of n-grams,  $w_n$  are uniform weights summing up to 1.

Despite its popularity, BLEU scores are heavily dependent on exact matches and strict word order, making it less suitable for literary translations where stylistic variations are common.

CHRF++ evaluates translation quality based on character n-gram F-scores between candidate and reference texts. Although CHRF++ accommodates flexible word ordering better than BLEU, it still faces challenges due to literary paraphrasing and stylistic variation.

Like BLEU, CHRF++ can undervalue semantically correct translations that differ structurally from the reference.

Semantic Similarity metrics based on sentence embeddings provide a meaningful semantic assessment of translations beyond structural matching. In this research, sentence embeddings were computed using Sentence-BERT (model: paraphrase-multilingual-MiniLM-L12-v2) [11]. Cosine similarity was applied to measure semantic closeness between the translated and reference texts.

Semantic Similarity effectively captures nuanced meaning, making it the most suitable automated metric among those considered for literary translation.

#### B. Context Consistency Metric

To quantitatively measure the consistency of glossary terms throughout the translated text, the Context Consistency metric (CoCon) (4) is introduced.

$$CoCon = \frac{1}{N} \sum_{j=1}^N F1(term_j) \quad (4)$$

where  $N$  is the total number of unique glossary terms found in the source text. By integrating fuzzy matching into the F1 (5) computation, the CoCon metric robustly handles translation variations, penalizing both omissions and inconsistencies. This quantification of terminological stability is particularly crucial in literary works, where domain-specific vocabulary must be consistently rendered.

$$F1(term_j) = \frac{2 \times \text{Precision}(term_j) \times \text{Recall}(term_j)}{\text{Precision}(term_j) + \text{Recall}(term_j)} \quad (5)$$

Рисунок В.3 – Сторінка 2 (рисунок створено самостійно)

where for each glossary term  $term_j$ , precision and recall are computed by comparing the reference (human) translation with the machine translation using fuzzy matching, which detects approximate matches even in the presence of inflectional changes or typographical differences. Specifically, precision is defined as the ratio of correct fuzzy matches to the total number of possible matches in the machine translation, while recall is the ratio of correct fuzzy matches to the total occurrences in the human reference.

This metric quantifies terminological stability and consistency in translations between narrative segments, crucial for literary works containing domain-specific vocabulary.

All combined metrics provide a comprehensive evaluation framework. However, despite advances in automated metrics, human analysis remains essential for accurately evaluating the quality of literary translation due to subtle stylistic and semantic nuances that are difficult to quantify automatically.

V. EXPERIMENTS AND RESULTS

Experiments were conducted using Meta-Llama 3.3-70B-Instruct through the Hugging Face Inference API [12] and the Together AI infrastructure, leveraging their server resources to perform translations. In total, translations were generated for two literary text excerpts (one narrative-oriented and one dialogue-heavy), resulting in 6 chunks for the narrative excerpt and 18 chunks for the dialogue excerpt. Each excerpt was translated into three languages (French, German, Ukrainian) using four translation methods:

- Baseline LLM translation (Default)
- LLM translation with Glossary (Glossary)
- LLM translation with Progressive Context Distillation (Context)
- LLM translation integrating both Glossary and Context (Context+Glossary)

To objectively evaluate the translation methods, the averaged results for each method were computed across narrative and dialogue excerpts. As expected, the standard automated metrics BLEU (Fig. 1) and CHRF++ (Fig. 2) demonstrated relatively low scores for all translation methods due to their inherent limitations in evaluating semantically accurate but structurally varied literary translations.

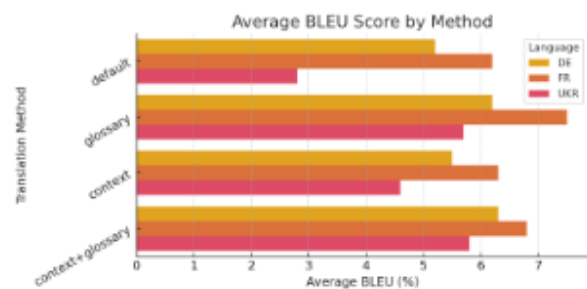


Fig. 1. Average BLEU score by method.



Fig. 2. Average CHRF++ score by method.

Fig. 3 provides an illustrative example clearly showing why BLEU scores remain low despite accurate semantic translation.

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 Брану Лінварів. Йшов пішки, а навічючого коня він за повід. Стояло пізно пообіддя, і крам лінварів та римарів було замкнено, а вуличка була порожньою. Було тепло, але чоловік той мав напнутого на плечі чорного плаща. Чим привертає увагу.
Machine	10.17	Півночі казали, що чоловік прийшов з півночі, з Брани Лінварів. Він прийшов пішки, ведучи свого в'юнкого коня за повід. Було пізно післяобіддя, і стійки ропарів, сідальників та шорників вже були закриті, вулиця була порожня. Було жарко, але чоловік мав чорний плащ, накинаний через плечі. Він привертає до себе увагу.

Fig. 3. Example of a BLEU sensitivity.

As illustrated, BLEU penalizes translations due to word-order changes and lexical variations despite semantic correctness.

Semantic Similarity scores (Fig. 4) provided a more accurate reflection of translation quality.

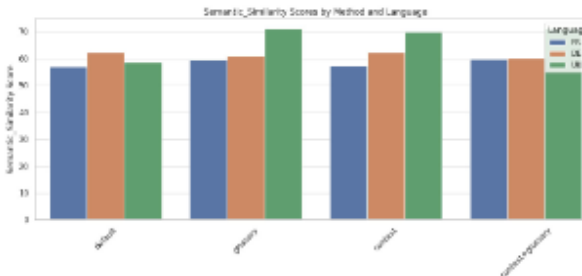


Fig. 4. Semantic Similarity scores.

The results indicate that all methods (Glossary, Context, Context+Glossary) outperform the baseline LLM translation. Notably, the combined Context+Glossary method consistently achieved decent Semantic Similarity scores across languages.

The heatmap in Fig. 5 illustrates that the methods incorporating glossary integration, namely Glossary and Context

Рисунок В.4 – Сторінка 3 (рисунок створено самостійно)

and glossary, consistently outperform the default model in all target languages.

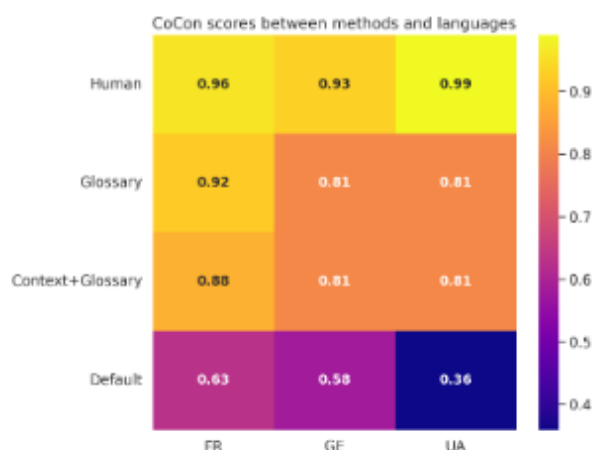


Fig. 5. CoCon scores.

These methods produce CoCon scores that are substantially closer to human reference, highlighting the critical role of the glossary in improving terminological accuracy. This analysis focuses specifically on a segment of narrative content, where consistent rendering of specialized terms is particularly important. In contrast, the default LLM shows significantly lower performance, highlighting the decisive contribution of the glossary to translation quality.

Table I summarizes the evaluation results, listing the average precision, recall, and F1 scores for each language and translation method.

TABLE I  
COMPARATIVE RESULTS FOR EACH METHOD

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

The results presented in Table I demonstrate that translation methods that incorporate glossary information consistently produce higher precision, recall, and average F1 scores compared to the default baseline method. These findings validate the efficacy of the CoCon metric based on fuzzy matching to capture terminological consistency across narrative segments, underscoring its importance in maintaining stability in translations of literary works with specialized vocabulary.

## VI. CONCLUSIONS

This research successfully achieved its defined objectives, demonstrating the effectiveness of the proposed translation methods. The experimental evaluation confirmed that the Progressive Context Distillation approach effectively enhances Semantic Similarity scores compared to standard translation prompts, indicating its usefulness in preserving narrative coherence across translated text segments.

Integrating glossary information greatly improved terminological consistency, ensuring accurate and consistent use of specialized terms in literary translation.

In general, the results demonstrate that LLMs can produce decent-quality literary translations when augmented with targeted context and terminology management strategies. However, evaluation metrics such as BLEU and CHRF++ struggle to capture their nuanced quality, highlighting the need for human evaluation.

This research shows the potential of context- and glossary-aware methods and lays the groundwork for future studies with larger or fine-tuned models to improve literary translation quality.

## REFERENCES

- [1] Chan V., Tang W. K. "GPT for translation: A systematic literature review" // *SN Computer Science*, 2024, Vol. 5, No. 8. DOI: 10.1007/s42979-024-03340-z
- [2] Palanichamy N., Trojovský P. "Overview and challenges of machine translation for contextually appropriate translations" // *iScience*, 2024, Vol. 27, No. 10. DOI: 10.1016/j.isci.2024.110878
- [3] Corpas Pastor G., Noriega-Santiañez L. "Human versus neural machine translation creativity: A study on manipulated MWEs in literature" // *Information*, 2024, Vol. 15, No. 9, Article 530. DOI: 10.3390/info15090530
- [4] Tan X., Zhang L. Y., Zhou G. D. "Document-level neural machine translation with hierarchical modeling of global context" // *Journal of Computer Science and Technology*, 2022, Vol. 37, pp. 295–308. DOI: 10.1007/s11390-021-0286-3
- [5] Wang, Z.; Chen, Y.; Zhang, J. "Machine translation of electrical terminology constraints" // *Information*, 2023, Vol. 14, No. 9, Article 517. DOI: 10.3390/info14090517
- [6] D. Dashenkov, K. Smelyakov, and O. Turuta, "Methods of multilingual 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
- [7] 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
- [8] Elkins K. "In search of a translator: Using AI to evaluate what's lost in translation" // *Frontiers in Computer Science*, 2024, Vol. 6. DOI: 10.3389/fcomp.2024.1444021
- [9] Al Rousan R., Jaradat R. S., Malkawi M. "ChatGPT translation vs. human translation: An examination of a literary text" // *Cogent Social Sciences*, 2025, Vol. 11, No. 1. DOI: 10.1080/23311886.2025.2472916
- [10] Hugging Face, meta-llama/Llama-3.3-70B-Instruct, Hugging Face. Accessed: Mar. 29, 2025. [Online]. Available: <https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct>
- [11] Sentence Transformers, paraphrase-multilingual-MiniLM-L12-v2, Hugging Face. Accessed: Mar. 29, 2025. [Online]. Available: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2>
- [12] Hugging Face, Inference API Documentation, Hugging Face. Accessed: Mar. 29, 2025. [Online]. Available: <https://huggingface.co/docs/api-inference/en/index>

## ДОДАТОК Г

Експертний висновок результатів перевірки кваліфікаційної роботи на  
відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи		
<u>студент</u> (посада)	<u>програмної інженерії</u> (кафедра)	<u>ІПЗМ-23-1</u> (група)
<u>Олексій ТРОФІМЕНКО</u> <small>(прізвище, ім'я, по батькові)</small>		
Зауваження		
Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Висновки	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	
6.2.2	Якщо додатки є продовженням тексту основної частини звіту, нумерація сторінок додатків — це продовження нумерації сторінок звіту. Кожний додаток повинен мати заголовок, який друкують вгорі малими літерами з першої великої симетрично до тексту сторінки. Над заголовком, але посередині рядка, друкують слово «ДОДАТОК» і відповідну велику літеру української абетки, крім літер Г, Є, З, І, І, Й, О, Ч, Ь, яка позначає додаток. Текст кожного додатка починають з наступної сторінки.	42, далі за текстом.
Методичні вказівки до виконання кваліфікаційної роботи магістра... <b>ЗАТВЕРДЖЕНО</b> кафедрою ІІІ протокол № 5 від 13.11.2023р. 3.2 Оформлення пояснювальної записки згідно з ДСТУ 3008:2015 Звіти у сфері науки і техніки. Структура та правила оформлення. <b>Шаблон</b> затверджений засіданням кафедри №3 від 16.10.2023.	Рисунок повинен розміщуватися одразу після його згадування у тексті, або на наступній сторінці. Під рисунком повинен бути підпис із словом Рисунок, порядковим номером цього рисунку, через тире з великої літери – назва рисунку та в круглих дужках вказується джерело з якого взятий цей рисунок, або то, що його виконано самостійно.	10, далі за текстом
Експерт	_____	Вадим НЕЧВОЛОД
	(підпис)	(прізвище, ініціали)
07.06.2025		

Рисунок Г.1 – Експертний висновок на відповідність вимогам ДСТУ (рисунок створено самостійно)