

**ПОРІВНЯЛЬНИЙ АНАЛІЗ OCR ТА LLM ПІДХОДІВ
ДЛЯ ПЕРЕКЛАДУ ТЕКСТУ З ЗОБРАЖЕННЯ СТОРІНКИ МАНГИ**

Остапчук В.О.

Науковий керівник – канд. техн. наук, доц. Яковлева О.В.

e-mail: maliuzhenko.valeriia@nure.ua

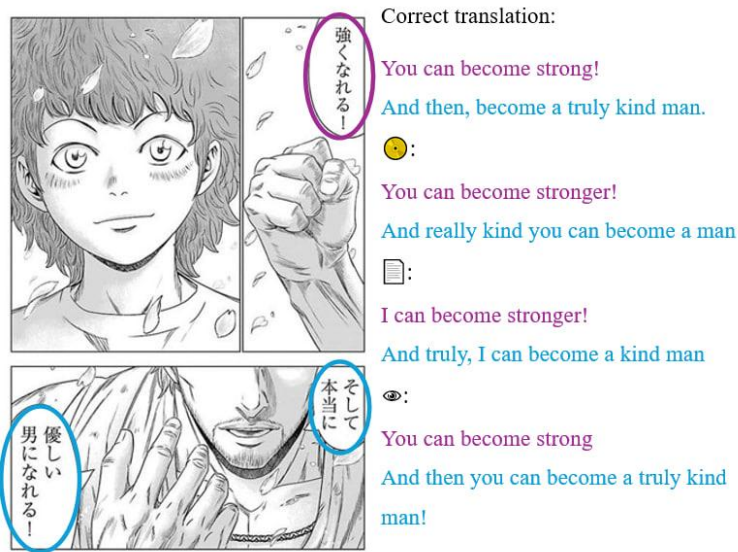
Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

This paper is devoted to the study of using MLLM technology in contextual manga translating field to address the limitation of traditional methods. Two distinct approaches were compared using a custom-made dataset, consisting of original pages and their official English translation. Three models were tested based on translational quality, processing speed and text recognition accuracy. The obtained results showed that MLLM-based approaches can make the process of translation and distribution significantly more efficient. Future work involves developing an application for easy localization.

У сучасному світі стрімко набирає популярності азійська культура. Одним з найбільших пластів цього контенту можна сміливо назвати японські комікси – мангу. Однак зі зростанням числа читачів за межами Японії почали відкриватися деякі проблеми. Якісний переклад, найголовніший критерій для дистрибуції манги, займає багато часу та сил, та вимагає від локалізатора наявності людини з високим рівнем японської мови. Аналіз існуючих готових рішень перекладу манги виявив певні обмеження функціональності. Майже жоден з них базується на технології нейронного машинного перекладу (Neural Machine Translation, NMT) та оптичного розпізнавання символів (Optical Character Recognition, OCR). Це призводить до неточностей перекладу та втраті закладеного оригіналом контексту. У свою чергу технології мультимодальних великих мовних моделей (Multimodal Large Language Model, MLLM), аналізуючи візуальний контекст, допомагають усунути ці проблеми.

У попередній роботі було розглянуто існуючі сервіси для перекладу манги, зазначено їх сильна та слабкі сторони та було зазначено, що імплементація технологій штучного інтелекту, зокрема MLLMs, які здатні аналізувати візуальні контексти сторінок, можуть значно покращити якість та знизити трудомісткість перекладу [1]. Порівняльний приклад перекладу методами на основі різних підходів наведено на рисунку 1 [2]. Серед представлених варіантів найбільш близьким до оригінального перекладу є метод на основі MLLM, заснований на мультимодальному аналізі зображення і тексту, оскільки він точніше передає зміст реплік і зберігає правильну комунікативну форму висловлювання. Таким чином, можемо бачити на ринку відсутність сервісів, які б дозволили здійснювати якісний контекстний переклад без необхідності використовувати людський труд та часи. Такий

застосунок міг би значно покращити швидкість перекладу та масштабувати рівень дистрибуції манги по всьому світу, у тому числі на український ринок, так як MLLM не потребує окремого налаштування для кожної пари мов.



● – NMT; 📄 – OCR; 👁 – MLLM;

Рисунок 1 – Порівняння результатів перекладу [2]

Метою роботи є проведення практичних досліджень для обрання найкращого підходу перекладу манги з точки зору точності та швидкості для подальшого його використання у мобільному застосунку.

Було вирішено дослідити два підходи:

– двох кроковий, де на першому кроці проводиться екстракція тексту на основі ORC (детектування та розпізнавання тексту), а на другому вже відбувається переклад розпізнаного тексту на основі MLLM;

– одно кроковий, де переклад отримується за допомогою MLLM на основі безпосередньо зображення без проміжного етапу розпізнавання.

Для проведення досліджень було проаналізовано існуючі манга-орієнтовані датасети та зроблено висновок, що такі датасети, наприклад, як Manga109, не покривають питання перекладу, а зосереджуються лише на аналізі зображення та використанні розпізнавання тексту. Тож для проведення більш розширеного аналізу, що містить й оцінку кінцевого перекладу, було створено власний датасет, що складається з 10 зображень сторінок оригінальної манги та їх офіційного перекладу на англійську мову.

Для аналізу першого підходу спочатку кроці було проведені експерименти для визначення якості розпізнавання тексту за допомогою комбінації спеціалізованих фреймверків Comic Text Detector (для нормалізації зображень) та MangaOCR (для розпізнавання на базі OCR), та за допомогою підходу на основі MLLM. Оцінка якості розпізнавання проводилася за допомогою відстанні Левенштейна. Експерименти показали середню збіжність

розпізнаних текстів за допомогою OCR з оригіналом 0,7266, у той час як показник MLLM становив 0,9021. Тож подальші дослідження щодо якості перекладу на основі результатів OCR було вирішено не проводити, та зосередитися на пошуку найкращої MLLM для перекладу. Таким чином, дослідження двохкрокового та однокрокового підходів звузилися до аналізу перекладу на основі MLLM, з використанням проміжного результату розпізнавання, або без нього відразу отримання перекладу.

Для експериментів було відібрано 3 популярні моделі, які показали найкращі результати у попередніх тестах: Gemini 3 Pro, Gemini 3 Flash, та GPT 5.2. В результаті було отримано, що найкращий час показала модель Gemini 3 Flash (без проміжного розпізнавання), наступною йде GPT 5.2, суттєво відстає Gemini 3 Pro (табл. 1). Найкращу якість за експертною оцінкою показала модель Gemini 3 Pro, трохи їй потупається Gemini 3 Flash.

Таблиця 1 – Результати оцінювання часових витрат

Варіанти однокрокового підходу	Gemini 3 Pro, сек	GPT 5.2, сек	Gemini 3 Flash, сек
Два запуски (розпізнавання, переклад)	30,7	21,7	4,6
Один запуск (переклад без проміжного розпізнавання)	32,1	3,8	3,3

Таким чином, можна зробити висновок, що однокроковий підхід на основі моделі Gemini 3 Flash без проміжного розпізнавання є компромісом між якістю та швидкістю, та його доцільно використати у майбутньому застосуванні для перекладу манги.

Список використаних джерел:

1. Yakovleva O., Ostapchuk V. Comparative analysis of the quality and problem areas of current manga translation tools. Proceedings of the VI Correspondence International Scientific and Practical Conference “Scientific vector of various sphere’ development: reality and future trends” in International scientific journal «Grail of Science». (February, 2026). Vinnytsia, Ukraine – Vienna, Austria. 2026. Vol. 62. P. 1096-1098. <https://doi.org/10.36074/grail-of-science.20.02.2026.118>

2. Philip L., Konrad S., Joshua T., Shonosuke I., Jie Y. Context-informed machine translation of manga using multimodal large language models. *arXiv.org e-Print archive*. (November, 2024). Vol. 1. P. 2 <https://arxiv.org/html/2411.02589v1>