Міністерство освіти і науки України

Національний університет “Києво-Могилянська академія”

Факультет інформатики

Кафедра інформатики

Кваліфікаційна робота

Освітній ступінь: магістр

на тему: “**Дослідження галюцинацій великих мовних моделей**”

Виконала: студентка 2 року навчання

Спеціальності 122 Комп’ютерні науки

Чайка Ольга Василівна

Керівник Ігнатенко О. П.

Доцент, доктор фізико-математичних наук

Рецензент

Магістерська робота захищена з оцінкою _____

Секретар ЕК

_____

«___» _____2025

Київ-2025

Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Кафедра інформатики факультету інформатики

ЗАТВЕРДЖУЮ

Зав.кафедри інформатики,

к.ф-м.н., доц. Гороховський С.С.

_____

„____”_____2025 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

ДЛЯ ДИПЛОМНОЇ РОБОТИ СТУДЕНТЦІ

Чайці Ользі Василівні

Тема роботи: “Дослідження галюцинацій великих мовних моделей”

Керівник роботи: Ігнатенко О.П., доцент, доктор фізико-математичних наук

Завдання:

Зміст ТЧ до магістерської роботи:

Зміст

Анотація

Дата видачі „___” _____ 2025 р. Керівник _____

(підпис)

Завдання отримав _____

(підпис)

ГРАФІК ПІДГОТОВКИ ДИПЛОМНОЇ РОБОТИ ДО ЗАХИСТУ

№ п/п	Назва етапу дипломного проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання теми магістерської роботи.	05.11.2024	
2.	Огляд тематичної літератури.	28.11.2024	
3.	Ознайомлення з тематичними матеріалами та побудова структури практичної та теоретичної частин роботи.	15.01.2025	
3.	Пошук та дослідження сучасних алгоритмів визначення галюцинацій у мовних моделях машинного перекладу	10.02.2025	
4.	Підготовка україномовного датасету	05.03.2025	
5.	Перевірка та аналіз ефективності обраних метрик на датасеті	17.03.2025	
6.	Побудова структури додаткового дослідження	10.04.2025	
7.	Генерація галюцинованих перекладів, проведення повторної перевірки метрик	20.04.2025	
8.	Аналіз отриманих результатів	02.05.2025	
9.	Написання пояснювальної роботи.	09.05.2025	
10.	Створення слайдів для доповіді та написання доповіді.	25.05.2025	

11.	Захист магістерської роботи (проєкту)	12.06.2025	

Студентка: Чайка Ольга Василівна

Керівник: Ігнатенко Олексій Петрович, доцент, доктор фізико-математичних наук

“_____”_____

ЗМІСТ

Анотація

Дипломний проєкт присвячено дослідженню наявних рішень визначення галюцинацій у моделях машинного перекладу у контексті перекладів з української мови.

Об'єкт дослідження: галюцинації у великих мовних моделях, зокрема у машинному перекладі.

Предмет дослідження: оцінка ефективності метрик визначення галюцинацій у машинному перекладі для української мови.

Мета дослідження: визначити, чи справляються сучасні метрики із визначенням галюцинацій у перекладі з української мови.

Ключові слова:

Велика мовна модель, машинний переклад, галюцинація мовної моделі, метрика оцінювання.

ВСТУП

У сучасну епоху глобалізації потреба у доступності перекладу є беззаперечною, оскільки давно вийшла за рамки суто ділових відносин. Не дивно, що машинний переклад (переклад текстів із використанням технологій машинного навчання) набув такої популярності: до нього вдаються застосунки, спрямовані на переклад, технологією послуговуються і застосунки більш широкого спрямування. Очевидно, що з розвитком моделей обробки природної мови їх почали залучати і до цього завдання також.

Між тим, моделі обробки природної мови, зокрема й ті, що використовуються у нейронному машинному перекладі наразі нерідко є далекими від досконалості: однією з проблем є їхня вразливість до галюцинацій. Зрозуміло, що складно зовсім уникнути галюцинацій, однак наразі навіть для низки поширених великих мовних моделей частота галюцинацій становить більше 5% [1]. Враховуючи поширеність використання нейронного машинного перекладу, зрозуміло, що ненадійність перекладу, до якої призводять галюцинації, може призвести до значних збитків для компанії, що використовувала такий переклад: чи то незадоволені субтитрами клієнти стрімінгових платформ на кшталт Netflix, чи збої на виробництві через некоректно перекладені вказівки керівництва.

Однак своєчасне виявлення галюцинованого перекладу дозволяє не лише локалізувати і виправити помилку у перекладі. Деякі патерни галюцинацій характерні специфічним помилкам у моделі чи датасеті [2], тобто ідентифікація галюцинацій дозволяє визначити помилки, яких було припущено при навчанні моделі чи розмітці датасету.

При оцінці якості згенерованих перекладів використовують ряд метрик, як-от BLEU, METEOR, Wass-to-Unif, Wass-to-Data, Wass-Combo. Утім, не

кожна з них пристосована визначати не просто загальну зв’язність перекладу саме галюцинації – як відомо, вони нерідко формують цілком зв’язний, хоч і хибний текст. Ще однією проблемою є те, що частина з наведених метрик не була перевірена для української мови. У той же час, українська мова є малоресурсною, тобто недостатньо представленою за лінгвістичними ресурсами, особливо у контексті дослідження галюцинацій. Це може призвести до того, що через нестачу якісних датасетів українською імовірність галюцинацій може зрости, але і у роботі метрик відбудуться викривлення, що зашкодять виявленню цих галюцинацій.

Отже, **об’єктом дослідження** у цій роботі є галюцинації у великих мовних моделях, зокрема у машинному перекладі. **Предметом дослідження** є оцінка ефективності метрик визначення галюцинацій у машинному перекладі для української мови. **За мету дослідження** поставлено визначити, чи справляються сучасні метрики із визначенням галюцинацій у перекладі з української мови. Для досягнення цієї мети потрібно виконати наступні **задачі**:

1. Сформувати і розмітити датасет зі зразками коректних і галюцинованих перекладів з українською на англійську;
2. Проаналізувати використовувані метрики перекладу, визначити такі, що спрямовані саме на визначення галюцинацій у ньому;
3. Реалізувати обчислення обраних метрик на прикладах зі створеного датасету;
4. Проаналізувати результати виконаної класифікації: чи можна вважати їх ефективними для використання для української мови, за потреби виконати додаткові перевірки;

5. Надати рекомендації стосовного подальших досліджень визначення галюцинацій у задачах машинного перекладу для української мови.

Оцінка метрик, що визначали галюцинації у перекладі виконувалась за допомогою AUROC, FPR@90TPR, Precision, Recall, F1-score.

1. ОСНОВНІ ПОНЯТТЯ

1.1. Машинний переклад

Задача машинного перекладу, як можна зрозуміти з назви, полягає у автоматизації перекладу з однієї природної мови на іншу. Різні джерела вказують за рік походження машинного перекладу 1947 або 1949 рік.

Орієнтовно до 2014 року для задач машинного перекладу використовувався підхід статистичного машинного перекладу. За цього підходу модель перекладу навчається на паралельному корпусі (тобто такому, що містить оригінальні тексти та їхні переклади цільовою мовою). Ціллю навчання є визначення імовірностей варіантів перекладу цільовою мовою до набору вхідного тексту. При цьому використовується також і мовна модель, що визначає з цих варіантів найбільш підходящий – такий, що найкраще узгоджується у цільовій мові, тобто статистично проявляє кращу узгодженість [3, 4]. Такий підхід, зокрема, мало підходить для семантично далеких (найчастіше таких, що належать до різних мовних груп) мов, оскільки потребує суттєвих змін у порядок слів і принципи побудови граматичних конструкцій.

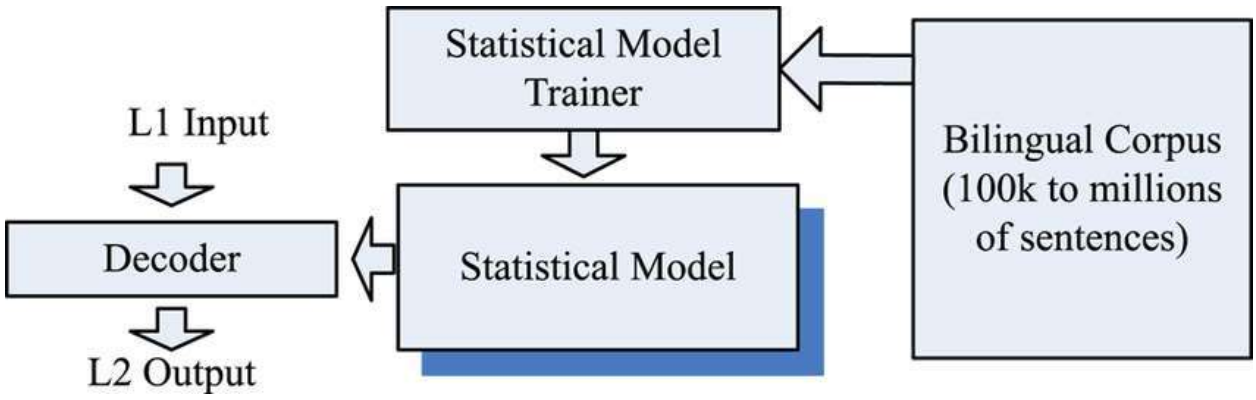

Рис. 1.1. Візуалізація статистичного машинного перекладу, взято з [5]

У 2014 році, втім, було запропоновано новий підхід до задачі – нейронний машинний переклад. Тут використовуються дві складові: енкодер, що перетворює текстове джерело однією мовою на вектор, і декодер, що перетворює уже вектор на текст цільовою мовою. [6]. Хоча конвенційні моделі машинного перекладу базуються на архітектурі енкодер-декодер, деякі роботи пропонують використовувати моделі GPT, що є суто декодерними: для обробки контексту і джерела та генерації вихідного тексту використовуються одні і ті ж параметри [7]. На рис.1.2 наведено схематичне відображення роботи енкодер-декодерної моделі машинного перекладу.

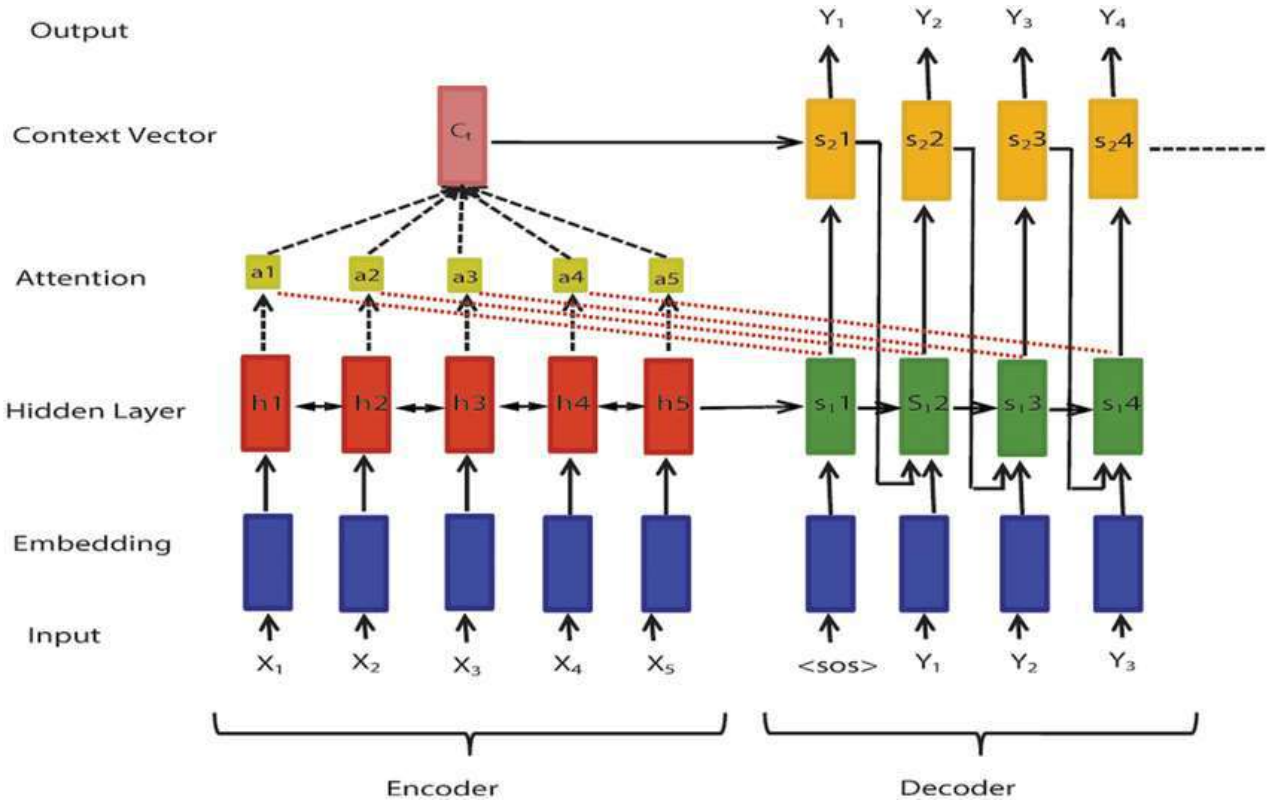

Рис. 1.2. Візуалізація роботи моделі нейронного машинного перекладу, взято з [8]

І за статистичного, і за нейронного машинного перекладу важливою складовою є обчислення механізму уваги: він визначає та оцінює зв’язок між цільовим токеном та словами з джерела.

1.2. Галюцинації у мовних моделях

Точне формулювання поняття галюцинації у різних джерелах розходиться, однак суть залишається одна: галюцинованим перекладом є такий, що містить у собі неточності, що призводять до спотворення суті оригінального речення. Таким чином, далеко не кожна помилка перекладу вважатиметься галюцинацією – до них не можна зарахувати граматичні помилки, що не мають впливу на суть речення (наприклад, некоректне вживання модальних дієслів).

У [9] визначають ряд можливих причин виникнення галюцинацій у мовних моделях:

- Галюцинації, викликані розбіжністю у даних: поширеною проблемою є евристичний збір даних для датасету і недостатньо ретельну фільтрацію значень у ньому, унаслідок чого у цільовому перекладі можуть опинитися речення, які неможливо підтвердити з джерела.
- Галюцинації, викликані похибками при тренуванні: тут навіть незначна похибка у датасеті здатна призвести до виклику галюцинації, оскільки обрана модель є схильною до галюцинацій. До цього можуть призвести хибні кореляції між різними частинами тренувальних даних, некоректний вибір стратегії декодування, параметричне знання (при переднавчанні модель запам’ятовує різні дані через параметри, а надалі вже при генерації

перекладу надає перевагу завченим параметричним знанням, а не тексту джерела).

Окрім розподілу за джерелами, що викликали галюцинації, їх найчастіше поділяють на intrinsic та extrinsic (надалі внутрішні та зовнішні відповідно) залежно від того, наскільки викривлена інформація пов’язана з оригіналом.

1.2.1. Intrinsic галюцинації

Внутрішніми галюцинаціями вважаються такі, що прямо протирічать вмісту джерела [9]: наприклад, маємо прогноз погоди, де українською сказано, що температура становитиме 25 градусів за Цельсієм, а в перекладі на англійську змінюється шкала вимірювання: замість градусів за Цельсієм отримуємо градуси за Фаренгейтом.

1.2.2. Extrinsic галюцинації

Зовнішньою ж галюцинацією буде така, що може містити і коректну інформацію, однак перевірити її у рамках джерела ми не в змозі. [9] У прикладі з прогнозом погоди такою галюцинацією було б, якби у англомовному перекладі, окрім інформації про температуру, отримали б інформацію і про швидкість вітру – в оригіналі про неї ані слова.

1.3. Оцінка якості машинного перекладу

1.3.1. Метрики оцінки якості машинного перекладу

При оцінці машинного перекладу найчастіше використовують BLEU, METEOR та BLEURT.

BLEU [10] спирається на модифіковану влучність (modified precision) перекладу. Вона обчислюється наступним чином: спершу обчислюється

максимальна кількість появ певного слова у референсних реченнях. Після цього вона так само обчислюється для кожного варіанту перекладу, а далі обрізається так, аби не перевищувати максимальну кількість появ у джерелі. Зрештою, визначається співвідношення кількості спільних n-грам до загальної кількості n-грам у реченні кандидаті. Зрозуміло, що узгодженість n-грам не гарантує змістову узгодженість перекладу. Через це за такого підходу галюцинований переклад все ще може мати хороші бали: скажімо, текст доволі довгий, але при цьому лексично переважно близький до референсу – тоді навіть суттєве відхилення від змісту за 1-2 словами мало вплине на загальну оцінку.

METEOR [11] також визначає подібність за уніграмами, однак враховує не лише влучність, але і повноту (recall) перекладу. На першому кроці обчислення метрики створюються набори усіх можливих мапінгів уніграм перекладу, що оцінюється, та референсного перекладу. Якщо кожна уніграма одного з перекладів мапиться не більше, ніж однієї уніграми в іншому, вважається, що було досягнуто вирівнювання. На кожному етапі мапуються лише ті уніграми, для яких раніше не було досягнуто підходящого відповідника. Крім того, на кожному з етапів використовуються різні модулі: порядок їхнього застосування не є строго визначеним, однак за замовчуванням на першому етапі використовується точна відповідність, на другому стемінг, на третьому синонімія. По завершенню вирівнювання обчислюються Precision та Recall та визначається середнє гармонійне для Precision та 9* Recall. Це середнє, а також штраф за надто довгі n-грами, і формують METEOR score. Варто відзначити, що METEOR при цьому також не підходить для визначення галюцинацій – при оцінці не враховується оригінальне речення, тобто відповідність вмісту перевірити не вдасться.

BLEURT [12] у обчисленні використовує натренований трансформер, наприклад, BERT (Bidirectional Encoder Representations from Transformers). Модель попередньо навчають на розмічених даних із пар перекладів та референсів, а також оцінками перекладів, наданих експертами. Ціллю навчання є максимальне наближення оцінки до людської. Хоча підхід суттєво відрізняється від двох попередніх, він так само не бере до уваги оригінальний текст (лише еталонні переклади). Таким чином, навіть якщо наданий переклад міститиме вигадану інформацію, але стилістично і граматично коректну, є високий ризик того, що такий переклад отримає хороші бали від BLEURT.

Отож, як бачимо, при оцінці якості машинного перекладу переважно покладаються на оцінку за метриками, що оцінюють стилістичну якість загалом, а не саме наявність галюцинацій. Водночас, як уже було сказано раніше, наявність галюцинацій може свідчити про помилки у датасеті чи архітектурі моделі машинного перекладу. Визначення галюцинованого контенту дозволить проаналізувати його патерн і, можливо, ідентифікувати джерело, однак ані BLEU, ані METEOR, ані BLEURT на це неспроможні.

1.3.2. Метрики визначення галюцинацій для машинного перекладу

Утім, існує ряд метрик, що специфікуються саме на визначенні галюцинацій, зокрема у машинному перекладі. Так, у 2023 році було запропоновано підхід, що базується на визначенні відстані Вассерштейна між розподілами уваг для перекладу, що оцінюється, та деяким еталонним розподілом.

За увагу у контексті глибинного навчання прийнято вважати вектор, який визначає розподіл ваг між токенами вхідного речення відповідно до їхньої значущості для певного токена у вихідному реченні [13]. Нижче наведено ілюстрацію використання шару уваги.

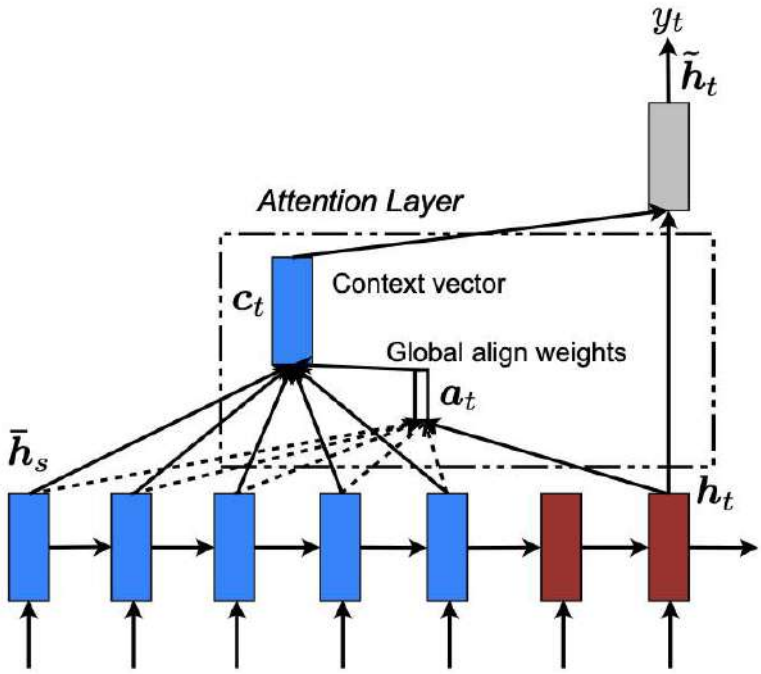

Рис. 1.1. Приклад використання механізму уваги, взято з [14]

У рамках цього підходу описано три наступних метрики [15]:

- Wass-to-Unif – тут до визначається розподіл уваг π_μ для тексту, що оцінюється, після чого між обчисленим розподілом та універсальним обчислюється відстань Вассерштейна. Чим більшою є ця відстань, тим вищою є імовірність того, що переклад, що оцінюється, є галюцинацією.
- Wass-to-Data за еталонний розподіл бере уже набір референсних розподілів уваг. Перш за все, обирається набір зразкових перекладів. З цього набору конструюється множина розподілів уваг. Зрештою, так само визначається розподіл уваг для перекладу, що оцінюється, після чого між цими розподілами визначається відстань Вассерштейна.
- Wass-Combo об’єднує попередні дві метрики, обчислюючи їх покроково. На першому кроці за визначеним пороговим скалярним значенням оцінюється, чи можна вважати оцінюваний переклад

галюцинацією згідно з Wass-to-Unif. Якщо значення Wass-to-Unif перевищує визначений поріг, спираємося на нормалізоване значення Wass-to-Unif. Якщо ні, додатково обчислюється Wass-to-Data. Більш формально:

$$s_{wc}(x) = 1\,[s_{wtu}(x) > \tau_{wtu}] \times \tilde{s}_{wtu}$$
$$+1\,[s_{wtu}(x) \leq \tau_{wtu}] \times s_{wtd},$$

де $s_{wc}(x)$ – Wass-Combo score для оцінюваного перекладу x, s_{wtu} – Wass-to-Unif score, τ_{wtu} – скалярне порогове значення, за перевищення якого переклад вважаємо галюцинацією, а s_{wtd} – Wass-to-Data score.

- STARE (Simple deTectors AggREgation) було запропоновано у 2024 році. Цей підхід пропонує використовувати агреговану оцінку від декількох метрик: метрики при цьому повинні бути спрямовані на визначення галюцинацій. Оскільки кожна з цих метрик може мати різну шкалу оцінювання, спершу виконується мінімаксна нормалізація для визначення ваг для балів метрик. За цими вагами далі складається зважене середнє оцінок, що і вважатиметься оцінкою STARE [16].

З урахуванням того, що описані метрики були розроблені безпосередньо для визначення галюцинацій, а не загальної оцінки стилістичної якості перекладу, було прийнято рішення у роботі розглянути і оцінити ефективність саме цих метрик. Важливо зауважити, що, хоч метрика STARE також міститься у переліку, наведеному вище, за результатами перевірок ефективності метрик, що спираються на відстань Вассерштейна було прийнято рішення цю метрику не оцінювати, оскільки успішність її використання напряму залежить від успішності метрик, що враховуються при обчисленні

STARE (тут до обчислення планувалося доєднати щонайменше дві метрики, Wass-to-Data і Wass-Combo).

2. ОЦІНКА ЕФЕКТИВНОСТІ МЕТРИК ОЦІНЮВАННЯ ГАЛЮЦИНАЦІЙ, ЗАСНОВАНИХ НА ВІДСТАНІ ВАССЕРШТЕЙНА

2.1. Бенчмарки для оцінки ефективності метрики

Оскільки уже визначено, що у рамках роботи як метрики визначення галюцинацій оцінюватися будуть Wass-to-Unif, Wass-to-Data та Wass-Combo, необхідно також визначити метрики, за якими буде оцінюватися їхня точність.

Перш за все, було обрано AUROC (Area Under Curve — Receiver Operating Characteristics) та FPR@90TPR (False positive rate at 90% of true positives), оскільки ними послуговувалися автори оригінальної статті, тож, визначивши їхню поточну оцінку, можна порівняти результати.

AUROC використовується для задач бінарної класифікації (такою і є задача визначення галюцинацій у згенерованому перекладі) та описує площу, що потрапляє під криву ROC. ROC-крива, у свою чергу описується через ще дві метрики: True Positive Rate (TPR, частка зразків коректно визначених як істинні) та False Positive Rate (FPR, частка зразків, хибно визначених як істинні), які обчислюються наступним чином:

$$TPR = \frac{TP}{TP+FN},$$

$$FPR = \frac{FP}{FP+T},$$

де TP – кількість істинно позитивних, FN – кількість хибно негативних, FP – хибно позитивних, а TN – істинно негативних значень. [17]

Графічно AUROC можна інтерпретувати наступним чином:

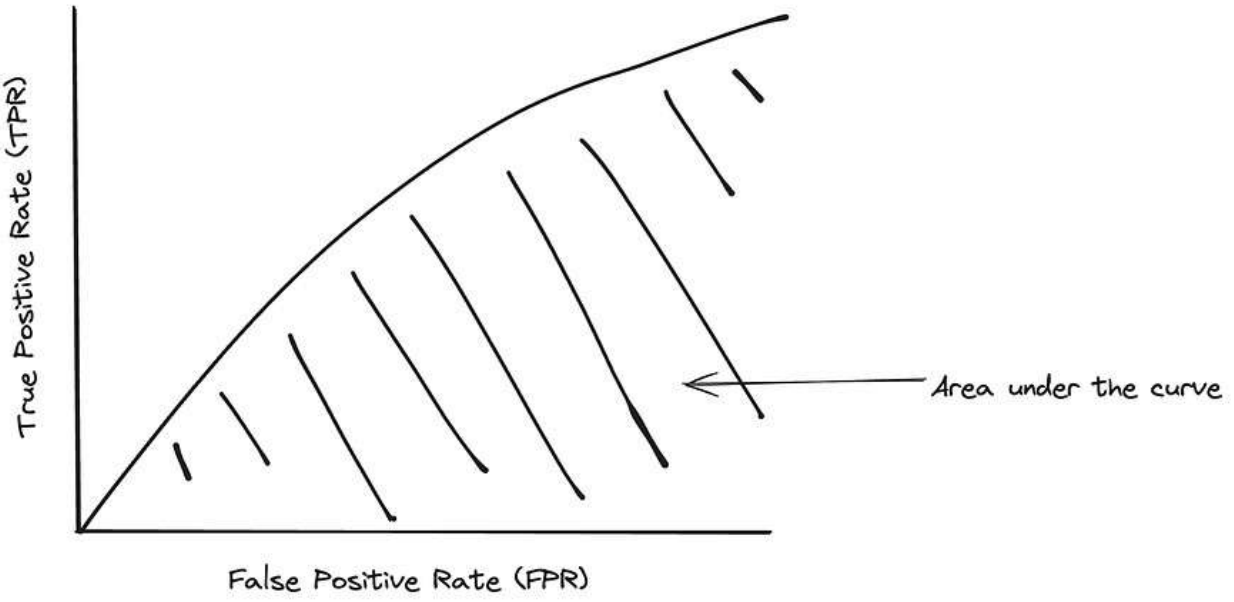

Рис. 2.1: Графічна характеристика AUROC та відображення кривої ROC, взято з [17]

FPR@90TPR визначає очікуваний відсоток хибно позитивних значень за умови, що коректно ідентифіковано 90% істинних значень.

Окрім метрик, що розглядалися в оригінальній статті, для ретельнішого аналізу було також обрано Precision, Recall, F1-score:

2.2. Підготовка україномовного датасету

Як уже було вказано, за мету було поставлено перевірити ефективність метрик на галюцинаціях, отриманих в українсько-англійському перекладі. Оскільки українська мова є не надто поширеною у дослідженні галюцинацій у моделях обробки природньої мови (зокрема і машинного перекладу), наразі немає розмічених україномовних датасетів із галюцинаціями у перекладі, тож розмітка виконувалась вручну.

За основу було взято датасет Ukrainian-English parallel corpus MaCoCu-uk-en 1.0. Він містить вибірку перекладів на англійську мову текстів з різних

веб-ресурсів, наприклад, онлайн-магазинів, сайтів туристичних агентств, громадських і релігійних організацій, витримок із прогнозів погоди тощо. [18] Оскільки датасет не містить міток, що стосуються галюцинаторності контенту, його було розмічено вручну. Для цього було відібрано 1000 прикладів (у оригінальній статті для перевірки використовували датасет на 3000 оцінених експертами прикладів, із них > 2.5 тисяч були коректними, тобто не містили галюцинацій). Було використано мітки “extrinsic”, “intrinsic”, “correct” та “other” (для помилок, що не є галюцинаціями). Мітки “extrinsic” та “intrinsic” замість єдиної мітки “hallucination” було використано, аби уможливити подальше використання для розширеного аналізу. Наявність у перекладі галюцинації, а також її тип визначалися згідно з визначеннями, наведеними у розділі 1. Приклади текстів для кожної з міток наведено у таблиці 2.1.

Таблиця 2.1: Приклад розмітки датасету

Оригінал	**Переклад**	**Мітка**
(82 відгуки) Emona — це бюджетний готель, розташований приблизно за 25 км від аеропорту "Любляна" та за декілька хвилин ходьби від Vidmarjeva vila. Гості, які подорожують автомобілем, можуть скористатися безкоштовною парковкою на території.	(367 reviews) Vault Hotel Ljubljana is a 4-star property situated 15 minutes' walk of distance from Tivoli Castle. Parking spaces is available on site as well.	extrinsic
В даному розділі «Погода у місті Мельдола» кожен відвідувач зможе знайти просте рішення, представлене у вигляді таблиць та графіків.	In this section "Weather today in the city Borgo" each visitor can find a simple solution presented in the form of tables and graphs.	intrinsic

Продовження таблиці 2.1.

Анотація: У статті розглядаються репрезентації Першої світової війни на сторінках щоденного російськомовного, правого видання «Киевлянин» за матеріалами 1914-1917 років.	Abstract: The article deals with the representation of the First World War on the pages of the daily Russian-language, right-wing newspaper «Kievlianin» in the materials of 1914-1917.	correct
Ясна річ, що при такій ситуації струмочок інвестицій в будівництво повільно, але вірно вичерпується, - Адже люди обережні вважають доцільними інвестиції лише в готові об'єкти або, принаймні, в ті КГ, де реально спостерігається будівництво будинків.	Clearly, under such situation trickle of investment in building slowly but surely running dry - Because people are cautious consider appropriate investments in a ready-made objects, or at least those in the CG, which is actually observed building homes.	other

Зрештою, мітки було розподілено наступним чином:

- 161 extrinsic;
- 129 intrinsic;
- 555 correct;
- 155 other.

Для визначення референсного розподілу уваг із датасету було виокремлено 200 записів із міткою correct. При обчисленні Wass-to-Data ці приклади не оцінювалися.

2.3. Опис використаних моделей

Перш ніж перейти до застосування власне метрик, необхідно визначитися з моделлю для обчислення розподілу уваг. До моделі було висунуто наступні вимоги:

- Підтримка української мови;
- Відносна легковісність.

З урахуванням цих двох обмежень, було обрано MarianMT та mBART.

MarianMT має модель “opus-mt-uk-en”, що призначена для генерації перекладів з українською на англійську, за архітектурою є sequence-to-sequence трансформером, отже, може надати перелік уваг для визначення розподілу, і важить 298 МБ, тобто не є надмірним навантаженням. [19] Цю ж модель також було використано для обчислення референсного розподілу уваг.

Як альтернатива при обчисленні Wass-to-Unif та Wass-to-Data для обчислення розподілу уваг у перекладі, що оцінюється, було використано mBART. Це також sequence-to-sequence трансформер, призначений у першу чергу для вирішення задачі машинного перекладу. Модель “mbart-large-50-many-to-many-mmt” підтримує 50 мов, у тому числі двосторонній переклад з української на англійську. Через це модель є важкою, важить близько 1.5 ГБ. [20] У зв’язку з цим розподіл уваг в оцінюваному тексті за допомогою mBART визначався лише для Wass-to-Unif та Wass-to-Data.

2.4. Перевірка ефективності метрик на україномовному датасеті

2.4.1. Wass-to-Unif

Перш за все було обчислено Wass-to-Unif. Як описано у описі метрики у розділі 1, для цього було визначено розподіл уваг між токенами цільового та вхідного речень: спершу було визначено матрицю уваги, далі її було переведено у вектор та нормалізовано. Між утвореним розподілом та універсальним розподілом тієї ж довжини було визначено відстань Вассерштейна. У коді ця частина виражається наступним чином:

```
attn = output.cross_attentions[-1]

            attn = attn.mean(dim=1)[0]          # [tgt_len, src_len]

            attention_mass = attn.sum(dim=0)

            attention_mass = attention_mass / attention_mass.sum()

            uniform    =    torch.full_like(attention_mass,    1.0    /
attention_mass.numel())

            wass  =  0.5  *  torch.nn.functional.l1_loss(attention_mass,
uniform, reduction="sum")
```

Зрештою, за використання MarianMT та mBART для визначення матриці уваг отримали показники Wass-to-Unif, наведені у таблиці нижче. Тут Precision, Recall та F1-score обчислювалися окремо для extrinsic та intrinsic галюцинацій.

Таблиця 2.2. Результати обчислення Wass-to-Unif для україномовного датасету

	MarianMT	**mBART**
AUROC	0.4941	0.5089

Продовження таблиці 2.2

FPR@90TPR	0.938	0.8986
Precision	Intrinsic: 0.0749 Extrinsic: 0.2406	Intrinsic: 0.196 Extrinsic: 0.1307
Recall	Intrinsic: 0.1085 Extrinsic: 0.2795	Intrinsic: 0.3023 Extrinsic: 0.1615
F1-score	Intrinsic: 0.0886 Extrinsic: 0.2586	Intrinsic: 0.2378 Extrinsic: 0.1444

Варто взяти до уваги, що авторами статті для Wass-to-Unif вказані значення AUROC = 0.8037, а FPR@90TPR = 0.7222. Неважко помітити, що отримані у ході експерименту результати до них навіть не наближаються. Бачимо, що за умови використання mBART отримали хоч і несуттєво, але переважно кращі результати, отож можна припустити, що щонайменше однією з причин поганих результатів за використання MarianMT є недостатньо коректне обчислення розподілу уваг. Варто також зауважити, що існує помітна різниця у обчисленні Precision, Recall та F1-score залежно від типу галюцинації: MarianMT проявляє суттєво гірші результати для intrinsic галюцинацій, а mBART, хоч і не з такою різницею, але краще проявляє себе на extrinsic. Останнє є дещо контрінтуїтивним, оскільки очікується, що для extrinsic галюцинацій їхнє виявлення, особливо засноване на розподілі уваг, повинно бути легшим, адже вони часто містять інформацію, яка не має нічого спільного із джерелом. Однак, зважаючи на те, що результати в обох випадках близькі до випадкових, не є доцільним спиратися на них для висування остаточних гіпотез стосовно їхніх причин – розбіжності можуть бути наслідком статистичної похибки.

2.4.2. Wass-to-Data

Як уже було вказано у попередніх розділах, обчислення Wass-to-Data у плані реалізації не надто відрізняється від Wass-to-Unif: основною різницею є те, що замість універсального розподілу для порівняння береться такий, що відповідає розподілу уваг коректних перекладів. У даному випадку було виокремлено приклади, позначені як коректні. Результати обчислення наведено у таблиці нижче.

Таблиця 2.3. Результати обчислення Wass-to-Data для україномовного датасету

	MarianMT	**mBART**
AUROC	0.4301	0.5090
FPR@90TPR	0.9042	0.8944
Precision	Intrinsic: 0.1168 Extrinsic: 0.1316	Intrinsic: 0.1373 Extrinsic: 0.1602
Recall	Intrinsic: 0.4264 Extrinsic: 0.3851	Intrinsic: 0.7442 Extrinsic: 0.6957
F1-score	Intrinsic: 0.1833 Extrinsic: 0.1962	Intrinsic: 0.2319 Extrinsic: 0.2605

Для порівняння, автори статті для Wass-to-Data отримали AUROC = 0.842, а FPR@90TPR = 0.4815, тобто слід було очікувати на суттєве покращення результативності, адже за референсний розподіл береться такий, що вже засновано на коректних переклад. Як бачимо, суттєвого покращення не відбулося, результати все ще залишаються на рівні випадкових. Обчислення за розподілом уваг, отриманим за використання mBART знову проявив себе дещо краще за те, що було обчислене через MarianMT.

2.4.3. Wass-Combo

Оскільки Wass-Combo спирається на результати, що отримуються за рахунок обчислення Wass-to-Unif і Wass-to-Data, зрозуміло, що на суттєве покращення очікувати немає сенсу, отож обчислення виконувалося лише для одного розподілу уваг, а саме отриманого за допомогою MarianMT. За порогове значення було взято середнє між максимальним та середнім значенням Wass-to-Unif для прикладів із міткою “correct”.

У розділі 1 надано текстовий опис розрахунку Wass-Combo, у коді це виражається наступним чином:

```
high_mask = wtu > threshold

# нормалізація wtu ТІЛЬКИ ДЛЯ ЗНАЧЕНЬ > threshold

max_above = wtu[high_mask].max() if np.any(high_mask) else 1.0

norm_wtu = np.zeros_like(wtu)

norm_wtu[high_mask] = (wtu[high_mask] - threshold) / (max_above - threshold + 1e-8)

# комбінування згідно з формулою

combo = np.where(high_mask, norm_wtu, wtd)
```

Результати обчислення Wass-Combo наведено у таблиці нижче.

Таблиця 2.4. Результати обчислення Wass-Combo для україномовного датасету

	MarianMT
AUROC	0.4670
FPR@90TPR	0.8634

Продовження таблиці 2.4

Precision	Intrinsic: 0.1379 Extrinsic: 0.1609
Recall	Intrinsic: 0.9302 Extrinsic: 0.8696
F1-score	Intrinsic: 0.2402 Extrinsic: 0.2716

Як уже і очікувалося, помітного покращення у результативності не відбулося, хоча автори метрики продемонстрували AUROC = 0.8717, а FPR@90TPR = 47.56.

На цьому етапі вже можна навести загальні висновки за усіма трьома метриками і висунути кілька гіпотез стосовно таких результатів. По-перше, очевидно, що метрики Wass-to-Unif, Wass-to-Data та Wass-Combo проявили дуже погані результати; результативність класифікації галюцинацій з їхнім використанням близька до випадковості. Тим не менш, стабільно кращі, хоч і все ще не переконливі, результати демонстрували метрики, які у обчисленнях використовували розподіл уваг, отриманий з mBART. Це свідчить, що MarianMT, ймовірно, недостатньо коректно визначає уваги для такої малоресурсної мови як українська. Маємо наступні припущення:

1. Wass-to-Unif, Wass-to-Data та Wass-Combo, незважаючи на ствердження авторів, не є результативними у визначенні галюцинацій у цілому.
2. Для української мови визначення розподілу уваг не працює достатньо коректно, щоб отриманий розподіл можна було використовувати для оцінки галюцинаторності контенту.

Для перевірки першої гіпотези було прийнято рішення знайти датасет із перекладами багаторесурсними мовами (наприклад, іспанською) та повторно оцінити метрики.

2.5. Підготовка іспаномовного датасету

З цією метою було обрано датасет Europarl, що містить переклади засідань Європарламенту для 21 мови. Зокрема було обрано датасет, що містить цитати з засідань іспанською та англійською. Цього разу потреби у розмітці не було, оскільки всі переклади у датасеті є коректними – вони надані перекладачами Європарламенту. [21]

Це, однак, означає, що потрібно отримати також і галюцинований переклад, адже у наведеному датасеті такі приклади не представлені. Для цього було взято 10000 речень і до них застосовано метод, описаний у [22]:

1. Із цільового речення (перекладу) видаляється/ маскується частина слів – це формує спотворене речення.
2. Зашумлене речення передається до BART-моделі, яка генерує нове речення шляхом beam search: оскільки частину вхідного тексту приховано, BART вставляє додатковий зміст, який не був у початковому реченні, таким чином формуючи галюцинований переклад.

У коді цей метод було реалізовано наступним чином:

```
# Маскування частини слів у реченні
def mask_tokens(text, mask_prob=0.3):
    words = text.split()
    masked = [w if random.random() > mask_prob else "<mask>" for w in words]
    if "<mask>" not in masked:  # гарантуємо наявність <mask>
```

```
        masked[random.randint(0, len(words) - 1)] = "<mask>"
    return " ".join(masked)
# Галюцинування перекладу
noised = mask_tokens(ref_translation)
inputs = bart_tokenizer(noised, return_tensors="pt")
with torch.no_grad():
    output_ids = bart_model.generate(**inputs, num_beams=5)
hallucinated = bart_tokenizer.decode(output_ids[0], skip_special_tokens=True)
```

У таблиці нижче наведено кілька прикладів згенерованих таким чином галюцинацій.

Таблиця 2.5. Приклади галюцинованих перекладів

Оригінал	**Коректний переклад**	**Галюцинований переклад**
Le confieso que, de momento, las cosas me parecen un tanto confusas.	I admit that, at present, the matter seems to be somewhat confused.	I admit that, at present, the matter seems to be settled.
Las preguntas contestadas con anterioridad se referían a la intervención, en determinado expediente, de la Sra. de Palacio, no a esas declaraciones aparecidas en el diario ABC el 18 de noviembre pasado.	The questions answered previously referred to Mrs de Palacio' s intervention, on another occasion, and not to these comments which appeared in the ABC newspaper on 18 November.	We have written to Mrs de Palacio' s office on another occasion, and to respond to her comments which appeared in the ABC newspaper on 18 May. We have questions answered below.

Продовження таблиці 2.5

El Grupo del Partido de los Socialistas Europeos solicita que se incluya una declaración de la Comisión sobre sus objetivos estratégicos para los próximos cinco años, así como sobre la reforma administrativa de la Comisión.	The Group of the Party of European Socialists requests that a Commission statement be included on its strategic objectives for the next five years and on the administrative reform of the Commission.	The Group of Ministers of the Party of German Socialists requests that a joint statement be included on its strategic objectives for the next five years and on the administrative reform of the European Commission.

Для обчислення розподілу коректних перекладів було використано ще 5000 зразків із оригінального датасету. У якості моделі для визначення розподілів уваг було знову використано MarianMT, цього разу “opus-mt-ROMANCE-en”, що призначена для перекладу з іспанської, французької, італійської, португальської та румунської на англійську [23].

2.6. Перевірка ефективності метрик на іспаномовному датасеті

Оскільки для даного датасету не виконувалась розмітка на intrinsic та extrinsic галюцинації, Precision, Recall та F1-Score оцінювалися незалежно від типу галюцинації.

Таблиця 2.6. Результати перевірки Wass-to-Unif та Wass-to-Data на іспаномовному датасеті

	Wass-to-Unif	**Wass-to-Data**
AUROC	0.7581	0.7987
FPR@90TPR	0.7334	0.6224
Precision	0.7594	0.6740

Продовження таблиці 2.6

Recall	0.5679	0.8406
F1-score	0.6498	0.7481

Цього разу бачимо, що результати вже кращі, хоча не остаточно досягають результатів, наведених авторами статті. Тут додатковою перепоною могло виявитися те, що датасет складається з речень із прямої мови, тобто розподіл уваг у будь-якому разі виглядатиме дещо неприроднім у порівнянні із письмовим текстом. Крім цього, не виключено, що частина прикладів закоротка.

Для додаткового аналізу результатів було відібрано 20 прикладів, 15 із них є найдовшими і ще 5 коротші, але при першій оцінці отримали високі значення Wass-to-Data. У коді відбір даних було реалізовано наступним чином:

```
df["len_perturb"] = df["en_perturb"].fillna("").apply(lambda x: len(x.split()))

# топ-15 найдовших галюцинованих

long_top = df.sort_values(by=["len_perturb", "wass_to_data"], ascending=False).head(15)

# 5 коротших з високим wass_to_data

short_top = df[df["len_perturb"] < 10].sort_values(by="wass_to_data", ascending=False).head(5)
```

Для кожного із цих 20 речень було згенеровано ще по 5 галюцинованих перекладів та додатково оцінено уже їх. Тут оцінка розподілу показників

проводилась без додаткових метрик, вручну, оскільки їхня кількість замала. У таблиці нижче наведено декілька показових прикладів.

Таблиця 2.6. Коректні переклади та їхні оцінки

Переклад	**Wass-to-Unif**	**Wass-to-Data**
We should remember that the most disadvantaged include immigrants and refugees.	0.7329	0.6588
Berger Report (A5-0007/2000) and Berger Report (A5-0012/2000)	0.6609	0.5054

Таблиця 2.7. Галюциновані переклади з оцінками

Галюцинацинований переклад	**Wass-to-Unif**	**Wass-to-Data**
We should remember that the disadvantaged are disadvantaged.	0.6974	0.6525
Berger Report (A5-0007/2000) and (A6-0008/2001) Report (B2-0009/2001).	0.4416	0.5413

Бачимо, що на коротших прикладах не надто добре себе проявляє Wass-to-Unif – імовірно, на коротких реченнях розраховується недостатньо репрезентативний розподіл уваг. Окрім цього, спостерігаємо, що галюцинації, котрі включають у себе відхилення за числовими значеннями, далеко не завжди негативно впливають на оцінку метрики. Подібну ж поведінку можна було спостерігати і при оцінці україномовного датасету – так, було відмічено приклад галюцинації при перекладі прогнозу погоди, де Wass-to-Unif та Wass-to-Data не перевищили навіть порогових значень.

Оригінал: днем повітря прогріється до +24...+26°C, точка роси: +18,08°C; співвідношення температури і вологості: неприємно сприймається більшістю людей; очікуються опади і гроза, рекомендується захопити з собою парасольку, ветер слабкий, що дме з північного сходу зі швидкістю 2-4 м/сек, небо затягнуте хмарами; одяг по сезону: відкриті сандалі, шльопанці, шорти, спідниця, легке плаття, футболка; астрономічний сезон: літо;

Переклад: in the afternoon the air temperature warms up to +23...+26°C, dew point: +19,82°C; ratio of temperature and humidity: Somewhat uncomfortable for most people; rain is expected, it is recommended to take an umbrella, gentle breeze wind blowing from the west at a speed of 11-14 km/h, overcast sky; what to wear: open sandals, flip flops, shorts, skirt, easy dress, t-shirt; astronomical season: summer;

Незважаючи на деякі винятки, з більшості прикладів бачимо, що метрики відпрацьовують адекватно, тобто можна зробити висновок, що на багаторесурсних мовах вони дійсно проявляють себе краще.

3. ОБМЕЖЕННЯ ТА РЕКОМЕНДАЦІЇ ДЛЯ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

3.1. Обмеження

Хоча за отриманими результатами можна зробити певні висновки, які до того ж відповідають розрізненним зауваженням деяких інших дослідників, при інтерпретації результатів варто взяти до уваги наступні обмеження цієї роботи:

- Розмір україномовного датасету: хоча автори статті демонстрували успішні результати обчислень метрик на датасеті, співставному за розмірами, все ж лишається імовірність, що на більш об’ємному датасеті метрики можуть проявити себе краще.
- Довжина речень: в україномовному датасеті були присутні зразки речень довжиною до 10 слів; на таких коротких реченнях, як показала подальша перевірка на іспаномовному датасеті, attention-based метрики, якими є Wass-to-Unif, Wass-to-Data та Wass-Combo, працюють гірше.
- Моделі визначення розподілу уваг: як показало дослідження, моделі MarianMT гірше визначають коректні розподіли уваг. Можливо, mBART також недостатньо пристосований до якісного визначення розподілу уваг для української мови.

3.2. Рекомендації для подальшого дослідження

З урахуванням описаних обмежень, рекомендується розширити датасет щонайменше до 10000 зразків. Крім цього, слід виконати додаткову розмітку за патернами походження галюцинацій, аби проаналізувати які саме галюцинації краще ідентифікуються. Також рекомендується провести додатковий аналіз мовних моделей та підібрати для визначення розподілу уваг

таку, що демонструє кращі результати українською, та за потреби донавчити її. Для цього, імовірно, слід підібрати або створити більш якісний датасет із перекладами з українською на англійську, оскільки Ukrainian-English parallel corpus MaCoCu-uk-en 1.0 містить велику кількість помилкових перекладів, у тому числі і таких, що не є і галюцинаціями.

Зрештою, варто дослідити наявні не attention-based метрики та перевірити доцільність їхнього використання для україномовних текстів.

ВИСНОВКИ

У рамках магістерської роботи вдалося перевірити ефективність attention-based метрик для визначення галюцинацій у перекладі на вручну розміченому датасеті з перекладами текстів з української мови. Перевірка показала, що метрики при класифікації текстів проявляють точність, близьку до випадкової, причому незалежно від того, із яким розподілом порівнюється розподіл уваг у текстах, що перекладаються. Для пояснення таких результатів було висунуто гіпотезу щодо того, що подібні метрики не дозволяють якісно визначати галюцинований переклад у малоресурсних мовах, якою є українська.

Було проведено додаткову перевірку на даних, що містили переклади з більш поширеної мови – іспанської. Перевірка продемонструвала, що Wass-to-Unif, Wass-to-Data та Wass-Combo адекватно проявили себе на датасеті з галюцинованими перекладами з іспанської, що може слугувати за підтвердження висунутої гіпотези.

Таким чином, результати роботи свідчать про те, що наразі використання таких метрик як Wass-to-Unif, Wass-to-Data та Wass-Combo для визначення галюцинацій у перекладах з/на українську мову не є результативним, тож пошук ефективніших метрик лишається актуальним. Рекомендовано провести додаткові дослідження із метриками, що у своїй оцінці не спираються на розподіл уваг.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Statistical machine translation. *Machine Translate*. URL: https://machinetranslate.org/statistical-machine-translation
2. Raunak V., Menezes A., Junczys-Dowmunt M. The Curious Case of Hallucinations in Neural Machine Translation. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online. Stroudsburg, PA, USA, 2021. URL: https://doi.org/10.18653/v1/2021.naacl-main.92
3. Lopez A. Statistical machine translation. *ACM Computing Surveys*. 2008. Vol. 40, no. 3. P. 1–49. URL: https://doi.org/10.1145/1380584.1380586
4. Statistical machine translation. *Machine Translate*. URL: https://machinetranslate.org/statistical-machine-translation
5. Saini S., Sahula V. A novel model based on Sequential Adaptive Memory for English–Hindi Translation. *Cognitive Computation and Systems*. 2021. Vol. 3, no. 2. P. 142–153. URL: https://doi.org/10.1049/ccs2.12011
6. Progress in Machine Translation / H. Wang et al. *Engineering*. 2021. URL: https://doi.org/10.1016/j.eng.2021.03.023
7. How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation / A. Hendi et al. 2023. URL: https://doi.org/10.48550/arXiv.2302.09210.
8. Neural Machine Translation Models with Attention-Based Dropout Layer / H. Israr et al. *Computers, Materials & Continua*. 2023. Vol. 75, no. 2. P. 2981–3009. URL: https://doi.org/10.32604/cmc.2023.035814
9. Survey of Hallucination in Natural Language Generation / Z. Ji et al. *ACM Computing Surveys*. 2022. URL: https://doi.org/10.1145/3571730

10. BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, 6 June 2002 – 12 June 2025. 2002.
11. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, Ann Arbor, 8 June 2005. 2005.
12. Sellam T., Das D., Parikh A. BLEURT: Learning Robust Metrics for Text Generation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online. Stroudsburg, PA, USA, 2020. URL: https://doi.org/10.18653/v1/2020.acl-main.704
13. GeeksforGeeks. ML - Attention mechanism - GeeksforGeeks. *GeeksforGeeks*. URL: https://www.geeksforgeeks.org/ml-attention-mechanism/
14. Papers with Code - Multiplicative Attention Explained. *The latest in Machine Learning | Papers With Code*. URL: https://paperswithcode.com/method/multiplicative-attention
15. Optimal Transport for Unsupervised Hallucination Detection in Neural Machine Translation / N. M. Guerreiro et al. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Toronto, Canada. Stroudsburg, PA, USA, 2023. URL: https://doi.org/10.18653/v1/2023.acl-long.770
16. Enhanced Hallucination Detection in Neural Machine Translation through Simple Detector Aggregation / A. Himmi et al. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, USA. Stroudsburg, PA, USA, 2024. P. 18573–18583. URL: https://doi.org/10.18653/v1/2024.emnlp-main.1033

17. Regmi S. Mathematical Interpretation of AUROC. *Medium*. URL: https://surajregmi.medium.com/mathematical-interpretation-of-auroc-9fc8d4f43ce3
18. Ukrainian-English parallel corpus MaCoCu-uk-en 1.0 – ELRC-SHARE. *ELRC-SHARE*. URL: https://elrc-share.eu/repository/browse/ukrainian-english-parallel-corpus-macocu-uk-en-10/56421b82261811eeab3000155d026706dae4e20d10584ce6857768de82baff9a/
19. MarianMT – transformers 3.5.0 documentation. *Hugging Face – The AI community building the future.* URL: https://huggingface.co/transformers/v3.5.1/model_doc/marian.html
20. facebook/mbart-large-50-many-to-many-mmt · Hugging Face. *Hugging Face – The AI community building the future.* URL: https://huggingface.co/facebook/mbart-large-50-many-to-many-mmt
21. Papers with Code - Europarl Dataset. *The latest in Machine Learning | Papers With Code*. URL: https://paperswithcode.com/dataset/europarl
22. Detecting Hallucinated Content in Conditional Neural Sequence Generation / C. Zhou et al. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Online. Stroudsburg, PA, USA, 2021. URL: https://doi.org/10.18653/v1/2021.findings-acl.120
23. Helsinki-NLP/opus-mt-ROMANCE-en · Hugging Face. *Hugging Face – The AI community building the future.* URL: https://huggingface.co/Helsinki-NLP/opus-mt-ROMANCE-en

ДОДАТОК А

5 із 20 згенерованих перекладів з іспанської та їхніх галюцинацій:

Речення 1:

Коректний переклад: We are now seeing the beginning of a strong recovery, which it appears may even be sustainable for a good while yet if we are prudent with our policies. This recovery is the logical consequence of the efforts we have made, but it comes at a price: it follows the restructuring of public accounts in European countries, which have kept inflation under control with a policy wisely intended to control costs and increase productivity in a Europe that has begun a vigorous restructuring of its own industries, banks and public services, even though this process is not yet complete and still has a long way to go.

Галюцинація 1: We are now seeing the beginning of a strong recovery, which it appears will be sustainable for a while yet if we are prudent in our policies. This is the consequence of the fiscal consolidation we have undertaken, but it comes at a price: it requires the restructuring of public services. It is Europe that has begun a vigorous restructuring of its industries, banks and public services, even though this process is not yet complete and still has a long way to go. It should be the European Central Bank which has kept inflation under control with monetary policy wisely intended to control costs and increase productivity in the economy.

Галюцинація 2: We are now in the midst of a strong recovery, which it appears may even be sustainable for a good long time if we are prudent with our policies. This recovery is a logical result of the efforts we have made, but it comes at a price: it follows the restructuring of public accounts in the 1990s, which kept inflation under control, with a policy of austerity to control costs and increase productivity in a difficult economic environment. China has begun a vigorous restructuring of its own industries, banks and public services, even though the process is not yet complete and still has a long way to go.

Галюцинація 3: We are now seeing the beginning of a strong recovery, which it appears may even be sustainable for a while yet if we are prudent with our policies. This recovery is a credit to the efforts we have made, but it comes at a price: it follows the collapse of bank accounts in many countries, which have been brought under control with a policy of austerity intended to control costs and increase

productivity in a Europe that has begun a vigorous restructuring of its banks and public services, even though this process is not yet complete and there is still a long way to go.

Галюцинація 4: We are now seeing the beginning of a strong recovery, which it appears will be even more sustainable for a good while longer if we are careful with our policies. This recovery is the logical consequence of the reforms we have made, but it comes at a price: we have seen the collapse of public finances in European countries, which have kept inflation under control, a policy wisely intended to control costs and increase productivity in a healthy economy. The United States has begun a similar restructuring of its own industries, infrastructure and public services, even though this process is not yet complete and still has a long way to go.

Галюцинація 5: We are now seeing the beginning of a strong economic recovery, which it appears may be sustained for a good while yet if we are able to continue with our policies. This recovery is the logical consequence of the reforms we have made, but it comes at a price: it requires the restructuring of public finances in European countries, which means keeping inflation under control with a policy wisely designed to control growth and increase employment in a sustainable way. Germany has already begun a vigorous restructuring of its banks and public services, even though this process is not yet complete and there is still a long way to go.

Речення 2:

Коректний переклад:

My party opposed this measure and the proposed amendments.

Галюцинація 1:

My organization opposed this measure and the proposed amendment.

Галюцинація 2:

My organization opposed the measure and its amendments.

Галюцинація 3:

The Democratic party opposes this measure and the proposed amendments. https://www.youtube.com/watch?v=0xVnLkFjYg0

Галюцинація 4:

My party opposed the measure and proposed amendments.

Галюцинація 5:

The Democratic party opposed this measure and the proposed amendments. â€“ Rappler.com

Речення 3:

Коректний переклад:

This means that we move away from shredding.

Галюцинація 1:

This means that it's from shredding.

Галюцинація 2:

This is that far away from shredding.

Галюцинація 3:

It means that we move away from shredding. We move towards shredding.

Галюцинація 4:

This is how that we move away from shredding.

Галюцинація 5:

This is that kind of shredding.

Речення 4:

Коректний переклад:

As for the other points, Mr Prodi, I will take economic and social policy as an example: the European social model is certainly experiencing a major crisis and it is certainly not coincidental that unemployment tops the list of problems that none of our policies has managed to resolve, but it is not certain that it can be resolved by going through a list of problems that we have discussed in the past, without having a clear idea or proposal which takes note of the fact that Europe contains economies

that are growing at some of the most sustainable rates and that it is not by chance that these are the economies which have made labour market flexibility and enterprise their prime objective.

Галюцинація 1:

As for the other question, Mr Prodi, I will take it and answer it as follows: the European social security system is certainly experiencing a crisis and it is not coincidental that unemployment tops the list of problems that some of our policies have managed to solve, but it certainly is not certain that it can be resolved by going through the kinds of problems we have discussed in the past, without a clear idea or strategy, which I note the fact that Europe contains economies that are growing at some of the most sustainable rates and that it is by chance that these are the economies that have made labour market flexibility and flexibility prime

Галюцинація 2:

As for the other points, Mr Prodi, I will use the economic and social situation as an example: the European economic model is certainly experiencing a crisis, and it is not surprising that unemployment tops the list of problems that none of our policies has been able to solve, but it is also not certain that these problems can be resolved by going through a list of measures that we have discussed in the past, without having a new idea or a new strategy. One takes note of the fact that Europe contains economies which are growing at some of the most sustainable rates in the world, and that it is a good chance that these

Галюцинація 3:

As for the other points, I will take economic and social policy as an example. The fact that Europe contains economies that are growing at some of the most sustainable rates shows that it is not by chance that these are the economies which have made the market economy and enterprise a success. On the other hand, the European social model is certainly experiencing a crisis, and it is certainly not coincidental that unemployment tops the list of problems that none of our policies have managed to resolve, but there is a certain comfort that it can be solved by going through a list of the problems that we have faced in the past, and having a

Галюцинація 4:

As for the other points, Mr President, I will take the labour market and social policy as an example: the European labour market model is currently experiencing a major crisis and it is not coincidental that this tops the list of problems that none of the current policies has managed to resolve, and is not certain that it can be resolved by going through a list of solutions, as we have discussed in the past, without having a clear strategy or proposal which takes note of the fact that there are some advanced economies that are growing at some of the most sustainable rates and that it are not by chance that these economies are those which have

Галюцинація 5:

As for the other points, Mr President, I agree that the economic and social as well as the European model is experiencing a crisis and it is not coincidental that unemployment tops the list of problems that our policies have managed to resolve, but it's not certain that it can be resolved by going back to the kind of problems we have discussed in the past, without having a clear idea or proposal which takes note of the fact that Europe contains economies which are growing at some of the most rapid rates and that there is a chance that these are economies which have given market and enterprise their legitimacy.

Речення 5:

Коректний переклад:

I have therefore asked my services to be willing to consider a reasonable increase in the number of towns eligible under the URBAN programme conditional upon certain guarantees, firstly on concentration - I am thinking particularly of the target areas which must have a minimum of 20 000 or, in exceptional cases, 10 000 inhabitants - and also upon the critical mass of funding: we must keep to EUR 500 per inhabitant so that the European Union action may be both effective and clear, and so that it does not resemble a sort of scattering of resources which would have no real impact either for citizens or for municipalities.

Галюцинація 1:

I have therefore asked my services to be willing to provide a reasonable amount of funding to municipalities under the EU programme conditional upon certain guarantees, particularly on population concentration - I am particularly of the

opinion that areas must have a minimum of 20 000 or, in exceptional cases, 10 000 inhabitants - and also upon the quality of funding: it must be close to â,¬ 500 per inhabitant so that the European Union action may be both effective and clear, and so that it does not resemble a waste of public resources which would have no real benefit either for the EU or for municipalities.

Галюцинація 2:

I have therefore asked the services to be able to consider a reasonable increase in the number of towns eligible under the URBAN programme based upon two criteria: firstly on population - I am thinking particularly of the target areas which must have a minimum, or, in some cases, 10 000 inhabitants - and also upon the mass factor: we want to keep the EUR per inhabitant so that the European Commission may be clear and clear, and so that it does not resemble a waste of resources which would have a real impact on the quality of life or the municipalities. Thank you.

Галюцинація 3:

I have asked my MP to be willing to consider a reasonable increase in the number of people eligible under the URBAN programme contingent upon certain guarantees, firstly on concentration - I am particularly interested in the target population which must have a minimum of 20 000 or, exceptional cases, 10 000 - and also upon mass - we must keep it under 500 per cent, so that Union action may be both effective and efficient and so that it does not resemble a sort of scattering of resources which will have no real impact either on citizens or the economy.

Галюцинація 4:

I have therefore asked my services to be willing to consider a reasonable increase in the number of towns eligible under the programme based upon certain criteria based on concentration of population; I am thinking particularly of the following criteria: towns must have a minimum of 20 000 or, in exceptional cases, a maximum of 30 000 inhabitants - and based upon the critical mass of funding: towns should keep to no more than â,¬ 500 per inhabitant - so that the European action may be both effective and clear, and so that it does not resemble a sort of scattering of resources that would have no real impact either for citizens or for municipalities.

Галюцинація 5:

I have therefore made my proposal to the European Commission to make a reasonable increase in the number of towns eligible for the URBAN programme conditional upon certain conditions - firstly on the inclusion of rural areas which must have a minimum of 20 000 or, in some cases, 30 000 inhabitants - and also upon the critical mass of funding: towns must keep to 500 euros per inhabitant so that the European Union programme may be both effective and clear, and so that it does not resemble a sort of scattering of resources which would have a real impact on the quality of life for citizens and for municipalities. Thank you.