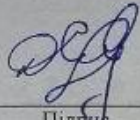
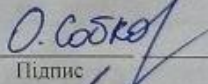
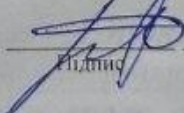


КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студент групи КН-22-2  Богдан ДЕНИСЕНКО
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ
Керівник: Ph.D., ст. викл., каф. КН  Олена СОБКО
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ
Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор  Олександр БАРМАК
Підпис Ім'я, ПРІЗВИЩЕ
17 червня 2026 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

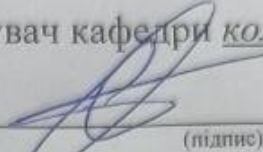
Освітній ступінь бакалавр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук



(підпис)

д.т.н., професор Олександр БАРМАК

« 22 » січня 2026 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА**

1. Тема кваліфікаційної роботи бакалавра: «Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій»

2. Завдання видано студенту Богдан Денисенко
(Ім'я, прізвище)

3. Керівник роботи старший викладач кафедри КН Олена Собко
(посада, ім'я, прізвище)

4. Затверджено наказом університету від « 20 » січня 2026 р. № 7

5. Дата видачі завдання студенту: « 22 » січня 2026 р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Метою кваліфікаційної роботи бакалавра є підвищення ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних. Для досягнення мети необхідно: виконати аналіз предметної області аналізу часової семантичної динаміки висловлювань спікерів у відеоконтенті, моделей глибокого навчання, сучасних наукових публікацій та існуючих програмних засобів; виконати формалізацію задачі аналізу часової семантичної динаміки висловлювань; розробити метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій; виконати програмну реалізацію інтелектуальної системи у вигляді експериментального вебзастосування для дослідження запропонованого методу; провести експериментальне дослідження розробленого методу на реальних даних YouTube-каналів та оцінити його ефективність за обраними метриками.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження теми кваліфікаційної роботи з керівником, складання календарного графіка виконання	січень 2026	Виконано
2	Ознайомлення з предметною областю, формулювання мети і задач дослідження, визначення об'єкта та предмета дослідження	лютий 2026	Виконано
3	Проектування методу розв'язання задачі, опис архітектурних рішень, розроблення математичних моделей та алгоритмів.	березень 2026	Виконано
4	Обґрунтування інструментарію розробки, програмна реалізація розробленого методу, проведення експериментального тестування та оцінювання ефективності.	квітень 2026	Виконано
5	Написання тексту кваліфікаційної роботи, урахування зауважень керівника, оформлення згідно з вимогами	травень 2026	Виконано
6	Розроблення презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2026	Виконано
7	Отримання відгуку керівника, рецензії, перевірка тексту кваліфікаційної роботи на плагіат, нормоконтроль	червень 2026	Виконано
8	Підготовка до захисту та захист кваліфікаційної роботи	червень 2026	Виконано

Виконавець:

студент групи КН-22-2
Група виконавця

Бого
Підпис

Богдан ДЕНИСЕНКО
Ім'я, ПРІЗВИЩЕ

Керівник:

Ph.D., ст. викл., каф. КН
Науковий ступінь, посада

О. Собко
Підпис

Олена СОБКО
Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: «Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій»

Виконавець кваліфікаційної роботи бакалавра: студент групи КН-22-2 Богдан Денисенко

Керівник кваліфікаційної роботи бакалавра: Ph.D., старший викладач кафедри КН Олена Собко

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
68	16	8	43	3

Метою кваліфікаційної роботи бакалавра є підвищення ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних.

Практичне застосування: моніторинг публічного дискурсу аналітичними центрами та редакціями, журналістські розслідування, академічні дослідження часової динаміки риторики у відеоконтенті.

Ключові слова: часова семантична динаміка, обробка природної мови, велика мовна модель, діаризація спікерів, нейромережевий метод.

Виконавець: студент групи КН-22-2
Група виконавця


Підпис

Богдан ДЕНИСЕНКО
Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень.....	4
Вступ.....	6
Розділ 1 Характеристика предметної області аналізу часової семантичної динаміки висловлювань спікерів.....	8
1.1 Аналіз інформаційних моделей семантичної динаміки висловлювань спікерів	8
1.2 Огляд теоретичних підходів до розв’язку подібних задач	10
1.3 Аналіз наукових публікацій.....	13
1.4 Аналіз існуючих програмних засобів.....	17
1.5 Мета та завдання дослідження	18
Розділ 2 Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій.....	19
2.1 Формалізація задачі аналізу часової семантичної динаміки висловлювань	19
2.1.1 Формалізація вхідних та вихідних даних.....	20
2.1.2 Формалізація задачі діаризації спікерів з ідентифікацією за голосовим відбитком.....	22
2.1.3 Формалізація задачі автоматичного розпізнавання мовлення.....	22
2.1.4 Формалізація задачі вилучення ключових ознак висловлювань	23
2.1.5 Формалізація задачі проєкції ключових ознак на тематичні осі та оцінки позиції.....	24
2.1.6 Формалізація задачі часової агрегації оцінок позиції	25
2.2 Кроки методу нейромережевого аналізу часової семантичної динаміки висловлювань	26
2.3 Нейромережеві моделі	29
2.3.1 Модель діаризації спікерів з ідентифікацією за голосовим відбитком ..	29
2.3.2 Модель автоматичного розпізнавання мовлення	30
2.3.3 Велика мовна модель для вилучення ознак та оцінки позиції.....	31
2.4 Опис датасету для експериментального дослідження	33

2.5 Метрики оцінювання ефективності.....	36
2.5.1 Метрики діаризації спікерів	36
2.5.2 Метрики автоматичного розпізнавання мовлення	37
2.5.3 Метрики класифікації позиції спікера	37
2.5.4 Інтегральна оцінка якості методу.....	39
2.6 Сценарії проведення експериментів	39
2.7 Висновки до розділу 2.....	42
Розділ 3 Експериментальне дослідження методу та застосування інтелектуальної системи	43
3.1 Опис експериментальної інтелектуальної системи.....	43
3.1.1 Структура інтелектуальної системи.....	44
3.1.2 Сценарії застосування.....	46
3.2 Результати експериментів.....	51
3.2.1 Оцінка якості діаризації та ідентифікації за голосовим відбитком.....	51
3.2.2 Оцінка якості автоматичного розпізнавання мовлення	52
3.2.3 Оцінка якості класифікації позиції спікера	53
3.2.4 Інтегральна оцінка та виявлення зміни позиції	56
3.2.5 Обговорення обмежень та напрями вдосконалення.....	58
3.3 Висновки до розділу 3.....	61
Загальні висновки.....	62
Перелік посилань	64
Додатки	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
API	Application Programming Interface – прикладний програмний інтерфейс
ASR	Automatic Speech Recognition – автоматичне розпізнавання мовлення
AUC	Area Under Curve – площа під ROC-кривою
CER	Character Error Rate – помилка розпізнавання на рівні символів
DER	Diarization Error Rate – помилка діаризації спікерів
F1	Гармонічне середнє Precision і Recall
GPT	Generative Pre-trained Transformer
JSON	JavaScript Object Notation – формат обміну даними
LLM	Large Language Model – велика мовна модель
MAE	Mean Absolute Error – середня абсолютна похибка
NLP	Natural Language Processing – обробка природної мови
ORM	Object-Relational Mapping – об’єктно-реляційне відображення
REST	Representational State Transfer – архітектурний стиль API
ROC	Receiver Operating Characteristic
RPM	Requests Per Minute – запитів за хвилину
TDT	Token Duration Transducer – трансдьюсер довжин токенів
TPM	Tokens Per Minute – токенів за хвилину
TTL	Time To Live – час життя кешованого об’єкта
VAD	Voice Activity Detection – виявлення мовленнєвої

	активності
WER	Word Error Rate – помилка розпізнавання на рівні слів
ДСТУ	Державний стандарт України
ГН	Глибоке навчання
КРБ	Кваліфікаційна робота бакалавра
ПЗ	Програмне забезпечення
СКБД	Система керування базами даних

Вступ

Кваліфікаційна робота бакалавра присвячена підвищенню ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних.

Актуальність. Споживання інформації в інтернеті змістилося у бік відеоформатів: подкастів, YouTube-каналів, прямих ефірів. Поряд із традиційними редакціями позицію аудиторії щодо суспільно-значущих тем формують окремі автори, які виступають у ролі лідерів думок. Кумулятивний ефект систематичного повторення наративів, тональних та інтонаційних маркерів робить таких спікерів важливими акторами публічного дискурсу.

Під впливом політичної кон'юнктури, фінансових інтересів або появи нових даних спікер змінює позицію щодо тем та подій. Зміна позиції відбувається відкрито, з обґрунтуванням, або приховано, шляхом поступового зміщення акцентів, тональних оцінок та частоти згадок ключових сутностей. Споживач, що звик довіряти автору, не помічає поступового зміщення позиції і продовжує сприймати спікера як носія сталої позиції. Прихована зміна становить ризик для медіагігієни, оскільки залишається непомітною без структурованого аналізу контенту за тривалий період.

Ручне відстеження часової семантичної динаміки висловлювань потребує перегляду десятків годин відео та послідовного аналізу окремих фрагментів, що робить такий процес дорогим та немасштабованим. Нейромережеві засоби обробки природної мови, моделі діаризації спікерів та великі мовні моделі знімають це обмеження: вони дозволяють перенести виявлення семантичних зрушень на машинне опрацювання та отримати кількісну оцінку зміни позиції за обраними тематичними осями.

Тематика роботи перебуває на перетині напрямів обробки природної мови, інтелектуального аналізу даних та аналізу часових рядів і безпосередньо стосується задач stance detection та діахронічного семантичного аналізу.

Об'єкт дослідження – процес нейромережевого аналізу часової семантичної динаміки висловлювань спікера за відеоконтентом.

Предмет дослідження – нейромережеві методи виявлення позиції спікера, моделі діаризації та автоматичного розпізнавання мовлення, великі мовні моделі для семантичного аналізу та засоби часової агрегації результатів.

Мета кваліфікаційної роботи бакалавра полягає у підвищенні ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних. Завдання кваліфікаційної роботи бакалавра:

- виконати аналіз предметної області аналізу часової семантичної динаміки висловлювань спікерів у відеоконтенті, моделей глибокого навчання, сучасних наукових публікацій та існуючих програмних засобів;

- виконати формалізацію задачі аналізу часової семантичної динаміки висловлювань;

- розробити метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій;

- виконати програмну реалізацію інтелектуальної системи у вигляді експериментального вебзастосунку для дослідження запропонованого методу;

- провести експериментальне дослідження розробленого методу на реальних даних YouTube-каналів та оцінити його ефективність за обраними метриками.

Практичне значення роботи полягає у можливості автоматизованого відстеження зміни позиції публічних спікерів за тривалий період, що раніше потребувало ручного перегляду десятків годин відео.

Розділ 1 Характеристика предметної області аналізу часової семантичної динаміки висловлювань спікерів

1.1 Аналіз інформаційних моделей семантичної динаміки висловлювань спікерів

Структура медіаспоживання за останнє десятиліття змінилася [1, 2]: разом із друкованими виданнями та телебаченням джерелом політичної, соціальної та економічної інформації стали відеоблоги, подкасти, авторські YouTube-канали та коментарі публічних експертів у соціальних мережах. У дослідженні [1] на трьох послідовних експериментах із аудиторією майже 1300 респондентів показано, що приналежність теми до «профільного» автора активує когнітивні упередження читачів і впливає на поширення контенту. У дослідженні [2] зафіксовано, що так звані лайфстайл-інфлюенсери дедалі частіше інтегрують політичні питання у звичні теми та формують політичні переконання саме у тих сегментів аудиторії, які уникають традиційних новин.

Цих авторів узагальнено називають лідерами думок, а саму їхню роль описують у термінах двоступеневого потоку комунікації: інформація з первинних джерел потрапляє до аудиторії опосередковано, через коментар спікера, що додає до фактів власну тональну та оцінювальну рамку. У дослідженні [3] на лонгitudному датасеті повідомлень показано, що лідери думок не лише поширюють позицію, а й формують ехо-камери та посилюють поляризацію навколо тематичних кластерів. Автори висловлювань, які залишаються активними протягом багатьох місяців і років, накопичують довіру у своєї аудиторії: систематичне повторення певних формулювань, інтонацій та оціночних маркерів закріплює у глядачів стійкі переконання.

Стійкість позиції спікера є умовною. Під дією зовнішніх та внутрішніх чинників, таких як зміна політичної кон'юнктури, фінансові інтереси, тиск спонсорської або партійної спільноти, поява нових фактів чи особистий перегляд цінностей, оцінки автора щодо однієї і тієї ж теми зміщуються. Перехід відбувається у двох режимах: відкритому, коли автор окремо проговорює зміну

позиції та обґрунтовує її, і прихованому, коли позиція зміщується без явного коментаря, а зміна впізнається лише за зіставленням висловлювань у різні моменти часу.

Прихований режим становить найбільш складний випадок для медіагігієни [4]. Глядач, що звик до спікера, продовжує атрибутувати йому стару позицію, тоді як висловлювання вже зміщені. У дослідженні [5] на даних повідомлень про вакцинацію проти COVID-19 продемонстровано, що при хронологічному поділі вибірки точність моделей виявлення позиції знижується у середньому на 8–15 % порівняно з випадковим поділом, а отже стандартні класифікатори без урахування часової динаміки не помічають саме повільної зміни позиції. Така зміна виявляється лише за умови систематичного агрегованого аналізу великого обсягу контенту того ж спікера.

Семантичні маркери, за якими розпізнають зміну позиції, мають кілька рівнів. Перший рівень – тональний: підтримка певної події з часом замінюється скепсисом, скепсис переростає у відкриту критику. Другий рівень – лексичний: змінюється набір ключових сутностей, які спікер активно згадує, з'являються нові акценти, а попередні теми поступово зникають з порядку денного. Третій рівень – стратегія аргументації: спікер відмовляється від фактологічного обґрунтування на користь емоційних звернень, іронії та персональних випадів. Четвертий рівень – частота згадок: різке зменшення згадок про подію або, навпаки, аномальний сплеск інтересу до неї сигналізує про коригування позиції.

Часовий горизонт відстеження залежить від мети дослідження. Для оперативного моніторингу актуальних подій достатньо тижневого або місячного інтервалу [6], оскільки зрушення в риторичі на цій шкалі добре фіксуються за частотою згадок. Аналіз часової семантичної динаміки висловлювань потребує піврічних і річних інтервалів, на яких семантичні зрушення стають статистично значущими та не зводяться до короткочасної реакції на одну подію [7].

Сутностями предметної області є: спікер як автор висловлювань, тема або подія як об'єкт оцінки, окреме висловлювання як одиниця аналізу, позиція спікера щодо теми як атрибут із часовою прив'язкою. Атрибутами

висловлювання є дата виходу, тривалість, ключові слова, тональність, стратегія аргументації та оцінка позиції автора. Подальшу автоматизацію цих сутностей доцільно виконати засобами обробки природної мови та нейромережевих моделей [8], аналіз яких наведено у наступному пунктах.

1.2 Огляд теоретичних підходів до розв'язку подібних задач

Розв'язок задачі аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій передбачає кілька послідовних підзадач: відокремлення мовлення цільового спікера у багатоголосому записі, перетворення мовлення в текст, вилучення семантичних ознак висловлювання, оцінка позиції спікера щодо тем та агрегація результатів у часі. Для кожної з підзадач у літературі запропоновано окремий клас нейромережевих моделей глибокого навчання.

Задача діаризації спікерів формулюється як «хто говорив коли» і розв'язується через ланцюжок із виявлення мовленнєвої активності, видобутку голосових ембедингів та кластеризації сегментів за спікером [9]. Сучасні нейромережеві підходи замінили класичний модульний пайплайн end-to-end архітектурами [10], які спільно навчаються на сегментацію та кластеризацію. Серед відкритих моделей виокремлюють трансформерні сегментаційні мережі з функцією втрат powerset multi-class cross entropy, які досягають рівня помилки діаризації Diarization Error Rate (DER) близько 11–19 % на стандартних бенчмарках [9, 11]. Окрему лінію складають end-to-end трансформерні архітектури з упорядкуванням за часом надходження мовлення, оптимізовані під обмежену кількість спікерів. Для тривалих записів та автоматичної ідентифікації спікера за наперед заданим голосовим відбитком застосовуються хмарні сервіси діаризації, що базуються на тих самих принципах, але донавчені на більших закритих даних. Для актуальної моделі precision-2 у документації [12] повідомлено усереднене значення DER 11,2 % на мультязичному бенчмарку та зниження DER на 14 % порівняно з попередньою версією моделі precision-1.

Задача автоматичного розпізнавання мовлення (ASR) у багатомовних умовах розв'язується трансформерними моделями типу encoder-decoder. Слабокеровані великомасштабні моделі [13], навчені на сотнях тисяч годин аудіо, охоплюють понад дев'ять десятків мов і досягають Word Error Rate близько 8–10 % на чистих даних [14]. Окремим класом є архітектури з FastConformer-енкодером та трансдьюсер-декодером, які підтримують 25 європейських мов, включно з українською, та досягають середнього WER 6,3 % у чистих умовах і 11,7 % при співвідношенні сигнал–шум 0 дБ [15]. Для україномовного відеоконтенту, який є пріоритетним у роботі, обрано саме цей клас моделей через підтримку мови, високу швидкість інференсу (приблизно у 10–12 разів за реальний час) та сумісність із розглянутим сервісом діаризації за форматом виходу.

Точність діаризації та розпізнавання мовлення безпосередньо впливає на якість семантичного аналізу. Помилка ASR на ім'я особи або топонім зміщує статистику ключових сутностей, а помилка діаризації приписує висловлювання іншому спікеру і викривляє оцінку його позиції. Зведене порівняння нейромережових моделей розпізнавання мовлення, актуальне на 2025–2026 рр., наведено у таблиці 1.1 за даними технічних звітів та карток відповідних моделей [13, 14, 15].

Таблиця 1.1 – Порівняння нейромережових моделей автоматичного розпізнавання мовлення

Модель	Архітектура	WER (чисті дані), %	Кількість підтримуваних мов	Реліз
Whisper large-v3 [13]	Encoder-decoder, слабкокеровано	8–10	>95	2023
WhisperX [14]	Whisper + forced alignment	8–10	>95	2023
Parakeet TDT 0.6b v3 [15]	FastConformer + TDT	6,3	25	2025
NeMo Canary 1B	Encoder-decoder	7,1	4	2024
Whisper turbo	Encoder-decoder	9,2	>95	2024

Значення WER (чисті дані) наведено за офіційними технічними звітами та картками моделей: Whisper [13], WhisperX [14], Parakeet TDT 0.6b v3 [15]; для NeMo Canary 1B – за офіційною картою моделі NVIDIA NeMo (huggingface.co/nvidia/canary-1b).

Задача виявлення позиції спікера (stance detection) формулюється як класифікація тексту відносно цільового об'єкта (теми, події, особи) з мітками типу «підтримка», «протиція», «нейтральне». В оглядовій роботі [16] зведено еволюцію методів, починаючи з ранніх бенчмарків stance detection [17] від класичних моделей машинного навчання до моделей на основі двонаправленого трансформера BERT [18], що розширює базову архітектуру трансформера [19] та сучасних великих мовних моделей сімейств GPT, Claude та Gemini; додатково еволюцію підходів узагальнено у [20] і [21]. У дослідженні [22] на десяти моделях розміром від десятків мільйонів до сотень мільярдів параметрів показано, що найбільші моделі без донавчання вже працюють на рівні (за принципом малосемплового навчання, описаним у [23]) донавчених менших трансформерів, а у багатьох соціально-наукових задачах перевершують і людських розмітників. У роботі [24] експериментально підтверджено, що велика мовна модель класу GPT-5 з нульовою температурою забезпечує відтворюваність метрик у межах $\pm 5\%$ при оцінці позиції у повідомленнях соціальних мереж.

Підвищення точності досягається і гібридними архітектурами. У роботі [25] логіко-тензорна мережа об'єднує велику мовну модель з логічним виведенням для пояснення передбачень. У роботі [26] графова нейромережа над гетерогенною стансовою мережею використовує великі мовні моделі для збагачення вузлів та ребер семантичними ознаками. Такі підходи дозволяють працювати з імпліцитною позицією, коли об'єкт оцінки не названо явно у тексті.

Окреме місце посідає часова складова задачі. Архітектура TempoFormer, описана у [27], інтегрує часові ознаки безпосередньо у механізм уваги трансформера через темпоральну варіацію rotary-позиційних кодувань і фіксує зміни громадської думки точніше за послідовне застосування статичних

класифікаторів. Великі мовні моделі додатково використовуються для пост-обробки результатів діаризації: у роботі [28] описано систему DiarizationLM, яка коригує помилки атрибуції спікерів за допомогою генеративних моделей і знижує DER на 25–55 % залежно від домену.

Семантичні зміни на рівні окремих слів і виразів є предметом окремої галузі – діахронічного аналізу семантики. В оглядовій публікації [29] систематизовано підходи до виявлення таких зрушень за допомогою контекстуалізованих ембедингів та великих мовних моделей. У роботі [30] показано, що використання визначень, згенерованих великою мовною моделлю, як семантичних представлень слова дає точність, конкурентну зі спеціально натренованими моделями, без потреби у вручну створених лексичних базах.

Підсумовуючи огляд, з огляду переваг для розв’язання задачі обрано таку конфігурацію типів засобів: хмарний сервіс діаризації precision-2 з підтримкою ідентифікації спікера за голосовим відбитком, мультимовну модель автоматичного розпізнавання мовлення Parakeet TDT 0.6b v3, велика мовна модель класу GPT-5 з підтримкою інструкційного режиму та регульованого reasoning effort для вилучення ключових ознак висловлювання та оцінки позиції за конфігурованими промптами, а часова агрегація реалізована на рівні бази даних з подальшою візуалізацією тренду оцінок.

1.3 Аналіз наукових публікацій

Задача аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій є актуальною науково-прикладною задачею у сферах обробки природної мови (NLP), інтелектуального аналізу аудіосигналів та діахронічного семантичного аналізу. Протягом останніх двох-трьох років у світовій практиці було опубліковано низку експериментальних досліджень, які частково або повністю охоплюють ключові компоненти цієї задачі.

У роботі [16] подано систематичний огляд методів виявлення позиції із застосуванням великих мовних моделей. Опрацьовано 144 публікації за період 2019 – квітень 2025 рр. і запропоновано трирівневу таксономію: режим навчання (з вчителем, нульовий, малосемпловий, інструкційний), модальність даних (тексти, мультимодальні дані) та архітектурна категорія моделі (енкодерна, декодерна, гібридна). Експериментально продемонстровано перевагу великих моделей сімейств GPT-4 та Claude на стандартних бенчмарках SemEval-2016 і EZ-STANCE, F_1 -метрика моделей цього класу досягає 76–82 % проти 61–68 % для дотренованих BERT-моделей. Робота слугує орієнтиром при виборі архітектури аналізатора висловлювань.

У роботі [5] вперше досліджено вплив часового концептуального дрейфу на задачу виявлення позиції щодо вакцинації проти COVID-19. На багатомовному датасеті відгуків у соціальних мережах (раніше досліджуваному у [31]) за 2020–2022 рр. порівняно випадковий і хронологічний поділ навчальної та тестової вибірок. На останньому точність F_1 спадає у середньому на 8–15 %, а в окремих експериментах сягає падіння до 22 %. Висновок робить недоцільним використання класифікаторів без часової адаптації для довгострокового моніторингу та доводить необхідність агрегованого аналізу за датою публікації висловлювання.

У публікації [27] описано архітектуру TempoFormer для часово-обізнаних представлень при виявленні зміни позиції. Часові інтервали між висловлюваннями інтегруються безпосередньо у позиційне кодування механізму уваги трансформера через темпоральну варіацію rotary-кодувань, що дозволяє моделі враховувати не лише послідовність, а й щільність подій у часі. На трьох бенчмарках змін настрою та позиції модель показала state-of-the-art результати за метрикою macro- F_1 порівняно з базовими лонгітюдними трансформерами. Підхід безпосередньо застосовний при формалізації часової динаміки у запропонованому методі.

У роботі [32] представлено фреймворк аналізу президентських дебатів за допомогою великих мовних моделей. Транскрипти дебатів обробляються

моделями класу GPT-4o і Claude 3 за фіксованим набором критеріїв «3P-3I» (policy, persona, perspective – interests, ideologies, identity). На рівні агрегованих оцінок зафіксовано систематичну різницю між кандидатами та між роками; суміжний підхід до аналізу персональних випадів у дебатах подано у [33], що демонструє здатність великих мовних моделей виявляти багатокomпонентні риторичні стратегії у довгих транскриптах. Робота підтверджує застосовність подібної конфігурованої схеми до україномовного відеоконтенту.

У роботі [14] представлено архітектуру для точного за часом транскрибування довгого аудіо: базова модель розпізнавання мовлення поєднується із примусовим вирівнюванням за фонемами та зовнішнім модулем діаризації. На бенчмарках TED-LIUM 3 та AMI автори фіксують Word Error Rate 8–10 % та зниження помилок атрибуції спікерів у довгих записах за рахунок вирівнювання на рівні слів. Підхід являє базовий блок транскрибування з поsegmentною атрибуцією, який може використовуватися як вхід для подальшого аналізу позиції спікера.

У роботі [22] на десяти моделях розміром від десятків мільйонів до сотень мільярдів параметрів порівняно режими zero-shot інференсу великих мовних моделей з інструкційним донавчанням трансформерних енкодерів. Для класифікації позиції великі моделі без донавчання досягають F_1 80–84 %, що збігається з донавченими базовими класифікаторами, а при малому обсязі розмічених даних перевищують їх у середньому на 5–12 %. Окремо виявлено, що великі моделі менш чутливі до зміни теми аналізу, що робить їх придатними для задач з відкритою множиною тематичних осей.

У роботі [25] запропоновано логіко-тензорну мережу для виявлення позиції, яка об'єднує велику мовну модель із логічним виведенням на рівні предикатів. Модель повертає не лише мітку позиції, але й ланцюжок логічних правил, що привели до рішення, що формує інтерпретований аналіз. На бенчмарках SemEval-2016 та P-Stance підхід досягає приросту macro- F_1 на 4–7 % порівняно з базовими моделями, що демонструє бажаний для формалізованого обґрунтування баланс точності та пояснювальності.

У роботі [24] експериментально досліджено відтворюваність оцінювання позиції та сентименту великими мовними моделями при різних налаштуваннях температури та версії моделі. Показано, що з нульовою температурою розбіжність F_1 між повторними запусками однієї моделі залишається в межах $\pm 5\%$, що робить великі моделі придатними для відтвореного аналізу. Результат важливий для довгострокового моніторингу, де оцінки позиції різних відео порівнюються в часі.

У роботі [29] подано систематичний огляд методів виявлення семантичних зрушень на рівні слів і виразів за допомогою контекстуалізованих ембедингів та великих мовних моделей. Узагальнено три типи зрушень: розширення значення, звуження значення та зміна тональності. Робота слугує концептуальною основою для побудови часової шкали оцінок позиції за тематичними осями.

У роботі [34] запропоновано двофазний підхід до виявлення позиції у діахронічному режимі: перша фаза формує початкову класифікацію за поточним вікном, друга фаза коригує оцінки з урахуванням зовнішніх факторів впливу. Експерименти на лонгітюдних даних демонструють приріст $\text{macro-}F_1$ на 3–6% порівняно з моделями без часової адаптації. Підхід прямо застосовний у задачі агрегації оцінок за тривалий період.

У роботі [30] показано, що використання визначень, згенерованих великою мовною моделлю, у ролі семантичних представлень слова дає точність, конкурентну зі спеціально натренованими моделями, без потреби у вручну створених лексичних базах. Підхід зменшує залежність від ресурсів для нових мов та доменів, що особливо суттєво для україномовного контенту.

Прослідковується закономірність: більшість досліджень фокусується або на статичному виявленні позиції в окремому тексті, або на короткостроковій динаміці у соціальних мережах за коментарями та постами. Окремо описані компоненти – транскрибування з таймкодами, класифікація позиції великими мовними моделями, часова обізнаність представлень, логіко-тензорні розширення – показують високу зрілість підходів. Цілісний пайплайн, який

поєднував би перелічені підходи у послідовному аналізі відеоконтенту з часовою агрегацією оцінок позиції за конкретним автором за роки публікацій, у переглянутих джерелах не зафіксовано.

1.4 Аналіз існуючих програмних засобів

Для перевірки повноти існуючих готових рішень виконано порівняння найближчих за функціональністю програмних засобів. Огляд охоплює інструменти, що працюють із довгим відеоконтентом або довгими транскриптами та підтримують одну або кілька стадій розглядуваного пайплайну: завантаження відео, діаризацію, ASR, аналіз семантичних ознак та часову візуалізацію. Зведене порівняння наведено у таблиці 1.2.

Таблиця 1.2 – Порівняння найближчих за функціональністю програмних засобів

Засіб	Діаризація з вр	Укр. ASR	Аналіз позиції з настройкою тем	Часова агрегація	Відкрите ДЗ
AssemblyAI Studio	+	+ (Universal)	–	–	–
pyannoteAI Hub	+	+ (Parakeet)	–	–	+ (Hub)
Whisper + WebUI	-/+	± (Whisper)	–	–	+
NeMo Speech Toolkit	+	+	–	–	+
Otter / Notta / Tactiq	+	–	–	–	–
Hugging Face Pipelines	+	+	± (через NLI)	–	+
Запропонований у роботі	+	+	+	+	+

Як видно з таблиці 1.2, готових засобів, які покривають усі чотири ключові складові методу (діаризація з ідентифікацією за голосовим відбитком, україномовний ASR, конфігурований аналіз позиції за тематичними осями, часова агрегація оцінок), у відкритому доступі не зафіксовано. Близькі за функціональністю напрацювання у напрямі застосування великих мовних моделей до політичної аналітики описано у [35]. Більшість комерційних продуктів зосереджена на транскрибуванні та діаризації як на кінцевому результаті; україномовні рішення для NLP-аналізу тональності та політичної

пропаганди розвиваються переважно у академічному середовищі [36], [37] і [38] та не передбачає семантичного аналізу довільних тем. Відкриті бібліотеки ruannotate, NeMo та Whisper надають окремі ланки пайплайну, але не містять інтеграційного шару з великою мовною моделлю та часовою агрегацією на рівні автора.

Це закріплює необхідність розробки відповідного методу, запропонованого у наступному розділі кваліфікаційної роботи, а також програмної реалізації, яка поєднує вище наведені складові у єдиний пайплайн опрацювання відеоконтенту.

1.5 Мета та завдання дослідження

Метою кваліфікаційної роботи бакалавра є підвищення ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних.

Для досягнення поставленої мети необхідно виконати наступні задачі:

- виконати аналіз предметної області аналізу часової семантичної динаміки висловлювань спікерів у відеоконтенті, моделей глибокого навчання, сучасних наукових публікацій та існуючих програмних засобів;
- виконати формалізацію задачі аналізу часової семантичної динаміки висловлювань;
- розробити метод неймережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій;
- виконати програмну реалізацію інтелектуальної системи у вигляді експериментального вебзастосунку для дослідження запропонованого методу;
- провести експериментальне дослідження розробленого методу на реальних даних YouTube-каналів та оцінити його ефективність за обраними метриками.

Розділ 2 Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

2.1 Формалізація задачі аналізу часової семантичної динаміки висловлювань

Для реалізації методу нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій необхідно формалізувати процес перетворення множини відеоматеріалів спікера у часові ряди оцінок його позиції щодо визначених тематичних осей. З огляду на складність такого перетворення, загальна задача декомпонується на послідовність взаємопов'язаних підзадач, кожна з яких виконує окремий етап обробки даних та формує проміжне представлення, що використовується на наступному кроці. Схему декомпозиції задачі наведено на рисунку 2.1. Перетворення розкладається на шість послідовних підзадач (підпункти 2.1.1–2.1.6), кожній з яких відповідає окремий крок методу у складі загального пайплайну.

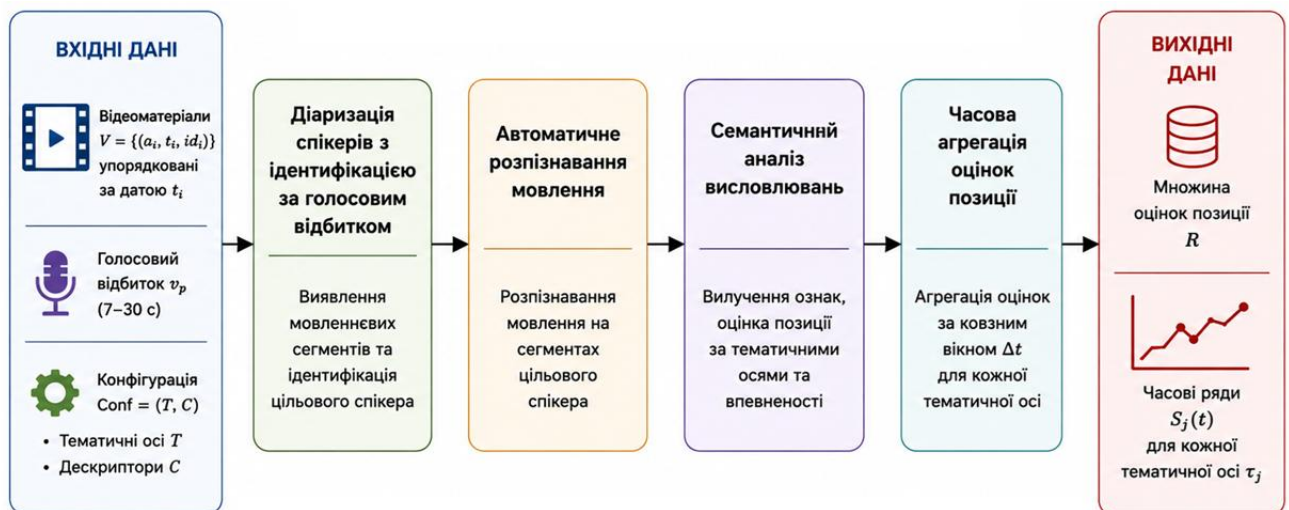


Рисунок 2.1 – Схеми декомпозиції задачі

Загальну логіку подано у вигляді алгоритму 2.1; детальну формалізацію кожного кроку наведено у підпунктах 2.1.1–2.1.6, а детальну схему методу наведено у пункті 2.2.

Алгоритм 2.1 – Метод нейромережевого аналізу часової семантичної

динаміки висловлювань

Вхід: $V = \{(idi, ti)\}$ – множина посилань на відеоматеріали;
 vp – голосовий відбиток спікера; $Conf = (T, C)$ – конфігурація аналізу;
 Δt – ширина вікна часової агрегації; $confmin$ – поріг впевненості.

Вихід: R – множина оцінок позиції; $\{S_j(t)\}$ – часові ряди за тематичними осями.

1. $R \leftarrow \emptyset$
2. для кожного $(idi, ti) \in V$ виконати:
3. $a_i \leftarrow$ завантажити_аудіо(idi) // крок 1
4. $Seg_i \leftarrow$ діаризувати(a_i, vp) // крок 2, ф.(2.6)
5. $Seg_iTARGET \leftarrow \{(s, e) \in Seg_i \mid speaker = TARGET\}$
6. $Utterances_i \leftarrow \emptyset$
7. для кожного $(startk, endk) \in Seg_iTARGET$ виконати:
8. $textk \leftarrow$ розпізнати_мовлення($a_i[startk : endk]$) // крок 3
9. $Utterances_i \leftarrow Utterances_i \cup \{(textk, startk, endk)\}$
10. для кожного $(textk, startk, endk) \in Utterances_i$ виконати:
11. $KFk \leftarrow$ вилучити_ознаки($textk, prompt_extract$) // крок 4
12. для кожного $kfm \in KFk$ виконати:
13. $(s_{ij}, conf_{ij}, \langle c \rangle_{ij}) \leftarrow$
агрегувати_полярність($kf_m, Conf$) // крок 5, ф. (2.9)
14. якщо $conf_{ij} \geq confmin$ і τ_j визначено, то:
15. $R \leftarrow R \cup \{(\tau_j, idi, ti, s_{ij}, conf_{ij}, \langle c \rangle_{ij})\}$
16. для кожного $\tau_j \in T$ виконати:
17. $S_j \leftarrow$ агрегувати_за_часом($R, \tau_j, \Delta t$) // крок 6, ф.(2.10)
18. повернути $(R, \{S_j(t)\})$

Задача формалізується як перетворення впорядкованої за датою публікації множини відеоматеріалів спікера у часовий ряд оцінок позиції за конфігурованим набором тематичних осей.

2.1.1 Формалізація вхідних та вихідних даних

Вхідними даними методу є впорядкована за датою публікації множина відеоматеріалів одного спікера. Кожен відеоматеріал подається трійкою «аудіодоріжка – дата публікації – ідентифікатор у джерелі». Запис множини вхідних даних подано у формулі (2.1).

$$V = \{(a_i, t_i, id_i) \mid i = 1, \dots, N\}, t_1 \leq t_2 \leq \dots \leq t_N, \quad (2.1)$$

де V – множина відеоматеріалів спікера; a_i – аудіодоріжка i -го відео; t_i – дата публікації i -го відео; id_i – ідентифікатор i -го відео у джерелі; N – кількість відеоматеріалів у множині.

Окремо подається голосовий відбиток цільового спікера vp , отриманий з контрольного фрагмента аудіо тривалістю від 7 до 30 секунд. Конфігурація

аналізу позиції задається кортежем $Conf = (T, C)$, де T – упорядкована множина тематичних осей, C – упорядкована множина дескрипторів позиції (формула 2.2).

$$Conf = (T, C), T = \{\tau_1, \tau_2, \dots, \tau_M\}, C = \{c_1, c_2, \dots, c_K\}. \quad (2.2)$$

Вихідними результатами методу є множина оцінок позиції спікера R , елементи якої прив'язані до конкретної тематичної осі та конкретного відеоматеріалу (формула 2.3), а також похідний від неї часовий ряд $S_j(t)$ агрегованих оцінок позиції за кожною темою (формула 2.4):

$$R = \{(\tau_j, id_i, t_i, s_{ij}, conf_{ij}, \langle c \rangle_{ij}) | \tau_j \in T, id_i \in V\}, \quad (2.3)$$

$$S_j(t) = \frac{1}{|W_t|} \sum_{i \in W_t} conf_{ij} \cdot s_{ij}, W_t = \{i | t - \Delta t \leq t_i < t\}, \quad (2.4)$$

де $s_{ij} \in [-1; +1]$ – числова оцінка позиції спікера щодо теми τ_j у відеоматеріалі i ; $conf_{ij} \in [0; 1]$ – впевненість моделі; $\langle c \rangle_{ij} \subseteq C$ – підмножина дескрипторів позиції; $|W_t|$ – потужність вікна; Δt – ширина ковзного вікна. Завдання методу – побудувати відображення F за формулою (2.5).

$$F: (V, vp, Conf, \Delta t) \rightarrow (R, \{S_j(t) | j = 1, \dots, M\}) \quad (2.5)$$

Псевдокод побудови вхідно-вихідних структур наведено нижче.

Алгоритм 2.1.1 – Підготовка вхідних і вихідних структур

Вхід: посилання $\{idi\}$, vp , $Conf = (T, C)$, Δt .

Вихід: $V = \{(ai, ti, idi)\}$, $R \leftarrow \emptyset$, $\{S_j(t)\}$.

1. для кожного idi :
 2. $ai \leftarrow$ завантажити_аудіо(idi)
 3. $ti \leftarrow$ дата_публікації(idi)
 4. сформувати $V = \{(ai, ti, idi)\}$, відсортовану за ti
 5. ініціалізувати $R \leftarrow \emptyset$
 6. ініціалізувати $S_j(t) \leftarrow$ порожній часовий ряд для кожного $\tau_j \in T$
-

Структури, побудовані на цьому кроці – впорядкована множина відеоматеріалів V , порожня множина оцінок позиції R та заготовки часових рядів $S_j(t)$ за кожною тематичною віссю – утворюють спільний контекст для всіх наступних підзадач 2.1.2–2.1.6 і визначають формат їхніх вхідних та вихідних даних.

2.1.2 Формалізація задачі діаризації спікерів з ідентифікацією за голосовим відбитком

Задача діаризації формалізується як виявлення у аудіодоріжці множини часових інтервалів, що відповідають мовленнєвій активності, та подальша атрибуція кожного інтервалу до окремого спікера. У контексті розглядуваного методу цільовий спікер є асоційованим з наперед зареєстрованим голосовим відбитком vp ; для решти учасників запису ідентифікація не виконується, а вони об'єднуються у сукупність OTHER. Відображення діаризації D подано у формулі (2.6).

$$D: (a_i, vp) \rightarrow \{(start_k, end_k, speaker_k) | k = 1, \dots, K_i, speaker_k \in \{TARGET, OTHER\}\}, \quad (2.6)$$

де K_i – кількість мовленнєвих сегментів у відеоматеріалі i ; $start_k, end_k$ – часові межі k -го сегмента; $speaker_k$ – мітка приналежності до цільового спікера, що визначається порівнянням ембедингу сегмента з голосовим відбитком vp за метрикою косинусної відстані з порогом `cosine_threshold`.

Алгоритм 2.1.2 – Діаризація з ідентифікацією за голосовим відбитком

Вхід: a_i – аудіодоріжка, vp – голосовий відбиток.

Вихід: $Seg_i = \{ (start_k, end_k, speaker_k) \}$.

1. виявлення_активності(a_i) → набір мовленнєвих фреймів
 2. для кожного фрейму: видобути ембединг через x -vectors
 3. кластеризація фреймів → попередні групи спікерів
 4. для кожної групи: усереднити ембединг → `group_emb`
 5. для кожної групи:
 6. якщо `cosine(group_emb, vp) ≥ cosine_threshold`:
 7. `speaker_k ← TARGET`
 8. інакше:
 9. `speaker_k ← OTHER`
 10. сформувати Seg_i зі знайдених меж сегментів та міток
-

Сформульована задача забезпечує атрибуцію мовленнєвих сегментів до цільового спікера, що є необхідною передумовою для коректного семантичного аналізу його висловлювань на наступних кроках методу.

2.1.3 Формалізація задачі автоматичного розпізнавання мовлення

Задача автоматичного розпізнавання мовлення формалізується як відображення з простору мовленнєвих сегментів у простір вирівняних за часом

текстових транскрипцій (формула 2.7). У розглядуваному методі застосування ASR обмежується лише сегментами цільового спікера, отриманими на попередньому кроці діаризації.

$$\text{ASR: } a_i[\text{start}_k:\text{end}_k] \rightarrow \{(\text{word}_l, t_l^{\text{start}}, t_l^{\text{end}}, \text{conf}_l) | l = 1, \dots, L_k\}, \quad (2.7)$$

де word_l – текст l -го слова; $t_l^{\text{start}}, t_l^{\text{end}}$ – часові межі слова; conf_l – впевненість моделі у розпізнаванні; L_k – кількість слів у k -му сегменті. Послідовні слова одного спікера об'єднуються у смислові висловлювання text_k за обмеженням довжини у токенах та паузами між словами понад 1,5 секунди.

Алгоритм 2.1.3 – Автоматичне розпізнавання мовлення і формування

ВИСЛОВЛЮВАНЬ

Вхід: $\text{SegiTARGET} = \{(\text{start}_k, \text{end}_k)\}, a_i$.

Вихід: $\text{Utterances}_i = \{(\text{text}_k, \text{start}_k, \text{end}_k)\}$.

1. $\text{Utterances}_i \leftarrow \emptyset$
 2. для кожного $(\text{start}_k, \text{end}_k) \in \text{SegiTARGET}$:
 3. $\text{words} \leftarrow \text{ASR}(a_i[\text{start}_k : \text{end}_k])$
 4. $\text{text}_k \leftarrow \text{об'єднати_слова}(\text{words})$ із розривами при паузі $> 1,5$ с
 5. якщо $\text{токенів}(\text{text}_k) > \text{max_tokens}$:
 6. розбити text_k на під-висловлювання
 7. $\text{Utterances}_i \leftarrow \text{Utterances}_i \cup \{(\text{text}_k, \text{start}_k, \text{end}_k)\}$
-

Отриманий у такому вигляді результат розпізнавання забезпечує точну часову прив'язку слів до моменту їх виголошення та слугує безпосереднім входом для семантичного аналізу висловлювань.

2.1.4 Формалізація задачі вилучення ключових ознак висловлювань

Задача вилучення ключових ознак висловлювань формалізується як відображення з простору текстів у простір розмічених сутностей (формула 2.8). Кожна ознака характеризується типом сутності (особа, група, організація, подія, поняття, статистика тощо) та підмножиною тегів модальності з конфігурації S (підтримка, критика, іронія, сарказм, скептицизм, нейтральність, адвокація, виправдання, тривога, спекуляція, насмішка, обурення).

$$\text{Extract: } \text{text}_k \rightarrow KF_k = \{(\text{keyword}_m, \text{type}_m, \text{tags}_m, \text{conf}_m, \text{context}_m) | m = 1, \dots, M_k\} \quad (2.8)$$

де keyword_m – згадана у висловлюванні сутність у початковій формі; $\text{type}_m \in \{\text{person, group, organization, place, event, date, object, concept, phrase, statistic}\}$ –

тип сутності; $tags_m \subseteq C$ – підмножина тегів модальності; $conf_m \in [0; 1]$ – впевненість моделі; $context_m$ – цитата з оригінального тексту.

Сформульована задача формалізує лексичний рівень аналізу висловлювань. Отримані ключові ознаки з тегами думки слугують входом для подальшої проєкції на тематичні осі.

2.1.5 Формалізація задачі проєкції ключових ознак на тематичні осі та оцінки позиції

Множина ключових ознак проєктується на конфігуровану множину тематичних осей T . Для кожної осі τ_j обчислюється усереднена оцінка позиції за відповідними ключовими ознаками з урахуванням впевненості моделі (формула 2.9).

$$s_{ij} = \frac{\sum_{kf_m \in KF_k^{(j)}} conf_m \cdot polarity(tags_m)}{\max(1, |KF_k^{(j)}|)} \quad (2.9)$$

де $KF_k^{(j)} \subseteq KF_k$ – підмножина ключових ознак, що проєктуються на тематичну вісь τ_j ; $polarity(tags_m) \in [-1; +1]$ – числова полярність набору тегів модальності, обчислена за фіксованою таблицею відповідності («підтримка» $\rightarrow +1$, «критика» $\rightarrow -1$, «нейтральність» $\rightarrow 0$ тощо). На відміну від часової агрегації (2.10), зваженої за впевненістю оцінок, проєкція ознак одного висловлювання на вісь використовує рівнозначне усереднення полярностей, оскільки ознаки в межах одного висловлювання вважаються рівноправними.

Алгоритм 2.1.5 – Семантичний аналіз висловлювань

Вхід: $Utterances_i$, $Conf = (T, C)$, $prompt_extract$.

Вихід: R_i – множина оцінок позиції за відеоматеріалом i .

1. $R_i \leftarrow \emptyset$
 2. для кожного $(text_k, start_k, end_k) \in Utterances_i$:
 3. $KF_k \leftarrow LLM_extract(text_k, prompt_extract)$
 4. для кожного $\tau_j \in T$:
 5. $KF_k(j) \leftarrow \text{проєкція_на_вісь}(KF_k, \tau_j)$ // ф. (2.9)
 6. якщо $KF_k(j) \neq \emptyset$:
 7. $s_{ij} \leftarrow \text{обчислити_оцінку}(KF_k(j))$ // формула (2.9)
 8. $conf_{ij} \leftarrow \text{середнє}(conf_m), m \in KF_k(j)$
 9. якщо $conf_{ij} \geq conf_{min}$:
 10. $R_i \leftarrow R_i \cup \{(\tau_j, idi, ti, s_{ij}, conf_{ij}, \langle c \rangle_{ij})\}$
-

Сформульована задача замикає семантичний рівень формалізації: ключові ознаки агрегуються у скалярні оцінки позиції за кожною тематичною віссю, що утворюють вхід для часової агрегації на наступному кроці.

2.1.6 Формалізація задачі часової агрегації оцінок позиції

Заключна підзадача методу полягає у побудові часового ряду агрегованих оцінок позиції спікера за кожною тематичною віссю. Для кожної осі $\tau_j \in T$ агрегація виконується за ковзним вікном тривалістю Δt ; усередненою величиною є середньозважена за впевненістю модель оцінок з відеоматеріалів, дати публікації яких потрапляють у вікно. Формальне визначення часового ряду подано у формулі (2.4), а розгорнуту процедуру обчислення – у формулі (2.10).

$$\text{Aggregate: } (R, T, \Delta t) \rightarrow \{S_j(t) : S_j(t) = \frac{\sum_{i \in W_t} \text{conf}_{ij} \cdot s_{ij}}{\sum_{i \in W_t} \text{conf}_{ij}}\}, \quad (2.10)$$

де $W_t = \{ i : t - \Delta t \leq t_i < t \}$ – вікно агрегації; знаменник у правій частині – сумарна впевненість оцінок у вікні, що нормалізує середнє за вагою. Часовий ряд $S_j(t)$ фіксує тренд позиції спікера щодо теми τ_j і слугує вхідним сигналом для виявлення точок зміни позиції – моментів, у яких ковзне середнє перетинає нейтральний рівень або змінює напрямок.

Алгоритм 2.1.6 – Часова агрегація оцінок позиції

Вхід: R – множина всіх оцінок позиції; T – множина тематичних осей; Δt .

Вихід: $\{ S_j(t) \mid j = 1 \dots M \}$ – набір часових рядів.

1. для кожного $\tau_j \in T$:
2. обрати точки часу $\{t_n\}$, що покривають період роботи спікера
3. для кожного t_n :
4. $W_t \leftarrow \{ i : t_n - \Delta t \leq t_i < t_n \}$
5. $\text{numerator} \leftarrow \sum_{i \in W_t} \text{conf}_{ij} \cdot s_{ij}$
6. $\text{denominator} \leftarrow \sum_{i \in W_t} \text{conf}_{ij}$
7. якщо $\text{denominator} > 0$:
8. $S_j(t_n) \leftarrow \text{numerator} / \text{denominator}$
9. інакше:
10. $S_j(t_n) \leftarrow \text{undefined}$

Розв'язання задач 2.1.1–2.1.6 у їхній послідовній композиції утворює метод нейромережевого аналізу часової семантичної динаміки висловлювань, детальний опис якого наведено у підпункті 2.2.

Якість побудованого відображення F оцінюється композицією трьох функцій втрат: помилки атрибуції мовленнєвих сегментів до спікера, помилки

розпізнавання мовлення та помилки оцінки позиції спікера. Інтегральна функція втрат подається у формулі (2.11).

$$\mathcal{L}(F) = \alpha \cdot \text{DER}(F) + \beta \cdot \text{WER}(F) + \gamma \cdot (1 - F1_{\text{stance}}(F)), \quad (2.11)$$

де $\text{DER}(F)$ – Diarization Error Rate, помилка атрибуції мовленнєвих сегментів до спікера; $\text{WER}(F)$ – Word Error Rate, помилка розпізнавання мовлення; $F1_{\text{stance}}(F)$ – макроусереднена F_1 -міра класифікації позиції за тематичними осями; α, β, γ – вагові коефіцієнти, що задають відносну важливість кожної складової (у роботі обрано $\alpha = 0,20, \beta = 0,20, \gamma = 0,60$).

Інтегральна оцінка $L(F)$ дозволяє звести оцінювання якості методу до єдиного безрозмірного показника, придатного як для порівняння конфігурацій, так і для контролю якості при модифікаціях окремих компонентів пайплайну.

2.2 Кроки методу нейромережевого аналізу часової семантичної динаміки висловлювань

Запропонований метод реалізує відображення $F: (V, v_p, \text{Conf}, \Delta t) \rightarrow (R, \{S_j(t)\})$ (формула 2.5) у вигляді послідовності з шести покрокових перетворень. Загальна схема методу, що поєднує входні дані, проміжні представлення та вихідні результати, наведена на рисунку 2.2.

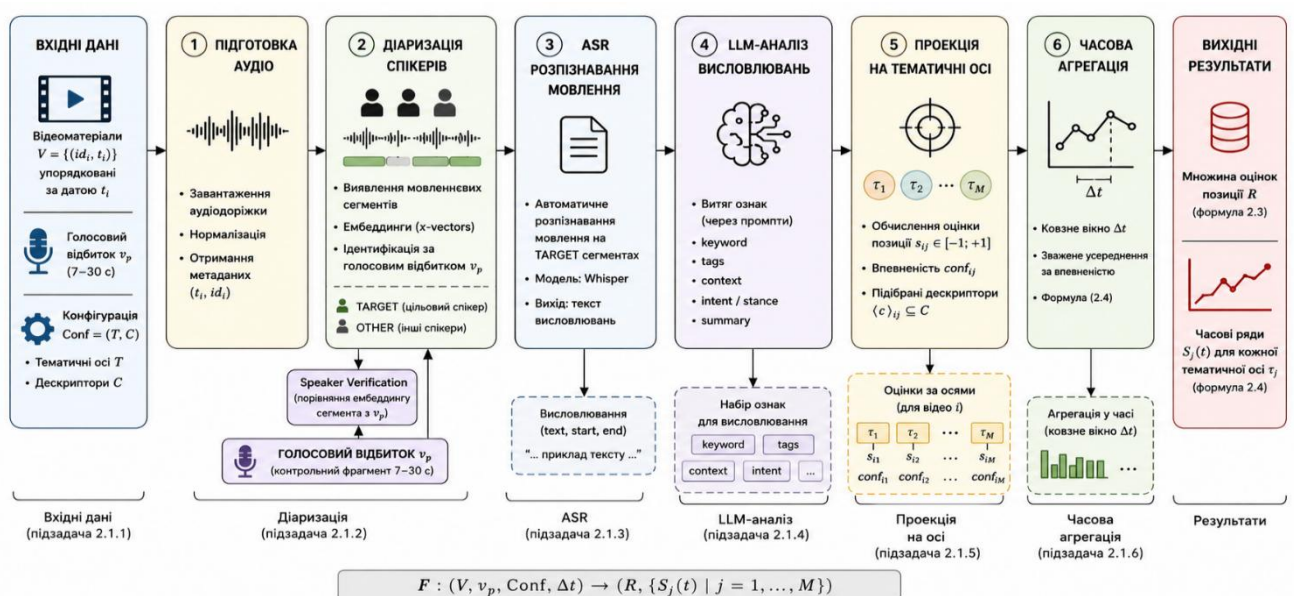


Рисунок 2.2 – Схема методу нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

Кожен з шести кроків методу формалізує одне з перетворень, запроваджених у підпунктах 2.1.1–2.1.6, та конкретизує його у контексті єдиного пайплайну.

Крок 1 – підготовка аудіоданих. На вхід кроку подається множина посилань $V = \{(id_i, t_i)\}$ та голосовий відбиток vp . Результатом виконання є множина пар (a_i, t_i) , у якій a_i – нормалізована до моно-сигналу 16 кГц 16-бітного РСМ аудіодоріжка, а t_i – дата публікації, зчитана з метаданих джерела. Для відеоматеріалів тривалістю понад 90 хвилин аудіодоріжка розбивається на блоки по 60 хвилин з перекриттям 5 секунд, що дозволяє зберегти контекст на межах блоків при подальшому семантичному аналізі. Виходом є множина пар (a_i, t_i) – аудіодоріжок та дат публікації. Реалізація: для кожного посилання id_i завантажується відеоматеріал з джерела, з нього виокремлюється аудіодоріжка та нормалізується до моно-сигналу 16 кГц 16-бітного РСМ. Тривалі записи (понад 90 хвилин) розбиваються на блоки по 60 хвилин з перекриттям 5 секунд. Дата публікації t_i зчитується з метаданих джерела.

Крок 2 – діаризація з ідентифікацією за голосовим відбитком, формалізована у формулі (2.6). Перетворення приймає аудіодоріжку a_i та голосовий відбиток vp і повертає множину мовленнєвих сегментів $\{(start_k, end_k, speaker_k)\}$ з мітками TARGET для цільового спікера та OTHER для інших учасників запису.

Крок 3 – автоматичне розпізнавання мовлення. Вхідною множиною кроку є мовленнєві сегменти цільового спікера, виокремлені на попередньому кроці. Результатом є послідовність текстових транскрипцій $\{(text_k, start_k, end_k)\}$. Послідовні сегменти об'єднуються у смислові висловлювання за обмеженням довжини у токенах (до 4 тис. токенів) та паузою у мовленні понад 1,5 секунди. Вихід: послідовність $\{(text_k, start_k, end_k)\}$ текстових транскрипцій сегментів. Реалізація: послідовні сегменти одного спікера об'єднуються у смислові висловлювання за обмеженням довжини у токенах (до 4 тис. токенів) та паузою у мовленні понад 1,5 секунди.

Крок 4 – вилучення ключових ознак висловлювання. Вхідним об'єктом є текст висловлювання $text_k$, отриманий на попередньому кроці. Результатом – множина ключових ознак $\{kf_m\} = \{(keyword_m, type_m, tags_m, conf_m, context_m)\}$, де $keyword_m$ – згадана у висловлюванні сутність, $type_m$ – її тип (тема, подія, особа, поняття тощо), $tags_m \subseteq C$ – теги модальності, $conf_m$ – впевненість моделі, $context_m$ – цитата з оригінального тексту. Вихід: множина ключових ознак $\{kf_m\} = \{(keyword_m, type_m, tags_m, conf_m, context_m)\}$, де $keyword_m$ – згадана у висловлюванні сутність, $type_m$ – її тип (тема, подія, особа, поняття), $tags_m \subseteq C$ – теги модальності, $conf_m$ – впевненість моделі, $context_m$ – цитата з оригінального тексту.

Крок 5 – проєкція ключових ознак на конфігуровані тематичні осі, формалізована у формулі (2.9). Для кожної тематичної осі $\tau_j \in T$ ключові ознаки, що стосуються теми осі, відбираються за збігом сутностей, а оцінка позиції s_{ij} обчислюється детерміновано як середня полярність тегів думки відібраних ознак з урахуванням їхньої впевненості. Цей крок не потребує окремого звернення до великої мовної моделі – він агрегує теги, вилучені на кроці 4.

Крок 6 – часова агрегація, що відображає формулу (2.4) та (2.10). Для кожної тематичної осі $\tau_j \in T$ будується послідовність вікон W_t з кроком $\Delta t/2$; у кожному вікні обчислюється середнє значення s_{ij} з вагами $conf_{ij}$; додатково обчислюється кількість висловлювань у вікні та розкид оцінок. Результат зберігається у вигляді часового ряду для подальшої візуалізації.

Базові операції завантаження аудіо, діаризації, розпізнавання мовлення, вилучення ключових ознак та оцінки позиції реалізуються нейромережевими моделями, розглянутими у підпункті 2.3. Системні інструкції великій мовній моделі (промпти вилучення ознак та оцінки позиції) наведено у додатку Б.

Оригінальність методу полягає у поєднанні діаризації з ідентифікацією за голосовим відбитком, двоетапного аналізу великою мовною моделлю (вилучення ознак \rightarrow проєкція на конфігуровані тематичні осі) та часової агрегації оцінок позиції на рівні автора. У відомих публікаціях, розглянутих у підпункті 1.3, ці компоненти описані окремо: робота [14] поєднує діаризацію з

ASR без аналізу позиції; робота [32] аналізує транскрипти за фіксованим набором критеріїв без атрибуції до спікера у багатоголосому записі; робота [27] вводить часову обізнаність у трансформер без етапу вилучення ключових ознак. Запропонований метод вперше об'єднує перелічені компоненти у послідовний пайплайн з конфігурованим набором тематичних осей.

2.3 Нейромережеві моделі

Для реалізації запропонованого методу необхідно використати три класи нейромережевих моделей: моделі діаризації з ідентифікацією спікера за голосовим відбитком, моделі автоматичного розпізнавання мовлення та великій мовній моделі для двоетапного семантичного аналізу – вилучення ключових ознак висловлювання та оцінки позиції спікера. Усі моделі застосовуються в режимі інференсу без донавчання на матеріалах роботи; поведінка моделей налаштовується через системні промпти та параметри декодування, що забезпечує відтворюваність результатів при повторних запусках.

2.3.1 Модель діаризації спікерів з ідентифікацією за голосовим відбитком

Модель діаризації формалізується як відображення з простору вхідних сигналів у простір розмічених сегментів за формулою (2.6). Вхідними величинами є аудіодоріжка a_i довільної тривалості, попередньо зареєстрований голосовий відбиток спікера vp та опціональне обмеження на максимальну кількість спікерів у записі. Структурно модель складається з трьох послідовних блоків: блоку виявлення мовленнєвої активності (Voice Activity Detection), блоку видобутку голосових ембедингів за кожен фрейм мовлення та блоку кластеризації фреймів за спікерами з прив'язкою до зареєстрованого відбитку. Структура моделі наведено на рисунку 2.3.

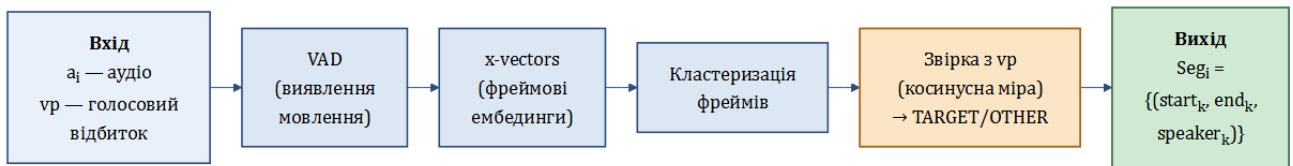


Рисунок 2.3 – Структура моделі діаризації з ідентифікацією за голосовим відбитком

Базовий блок виявлення мовленнєвої активності реалізується трансформерною сегментаційною мережею з функцією втрат powerset multi-class cross entropy, що дозволяє моделі одночасно прогнозувати перекриття мовлення кількох спікерів [11]. Видобуток ембедингів виконується нейромережею на основі x-vectors, яка проєктує фрейм мовлення тривалістю 1,5 с у вектор фіксованої розмірності 256. Кластеризація використовує комбінацію агломеративного групування з порогом косинусної відстані та зіставлення з голосовим відбитком vp за порогом $\text{cosine_threshold} = 0,55$. Результатом роботи моделі є послідовність кортежів $\{(start_k, end_k, speaker_k) \mid k = 1 .. K\}$, де $speaker_k$ набуває значення цільового спікера або «інший».

Описана архітектура поєднує статистичну кластеризацію x-векторів із цілеспрямованою зв'язкою з голосовим відбитком, що дозволяє коректно ідентифікувати цільового спікера навіть у багатоспікерних діалогах та фоновій студійній мові.

2.3.2 Модель автоматичного розпізнавання мовлення

Модель автоматичного розпізнавання мовлення формалізується як відображення з простору мовленнєвих сегментів у простір розмічених за часом текстових транскрипцій за формулою (2.7). Вхідним об'єктом є виокремлений мовленнєвий сегмент тривалістю до 30 секунд; результатом – текстова транскрипція з посимвольним вирівнюванням за часом. Структуру модель належить до сімейства трансформерних мереж із FastConformer-енкодером та трансдьюсер-декодером (Token Duration Transducer, TDT), навчених на наборі

мультимовних аудіоданих, який охоплює 25 європейських мов, у тому числі українську [15]. Структуру моделі наведено на рисунку 2.4.

Енкодер складається з блоків FastConformer з функцією уваги до контексту до трьох годин у режимі local attention; декодер генерує токени тексту авторегресійно з використанням транздюсер-стека. Вихідним результатом є послідовність $\{(word_l, start_l, end_l, conf_l) \mid l = 1 .. L\}$, де $word_l$ – текст слова, $start_l$, end_l – часові межі слова, $conf_l$ – впевненість моделі у розпізнаванні. Сегментація на смислові висловлювання виконується об'єднанням послідовних слів одного спікера за обмеженням довжини у токенах та часовою паузою між словами.

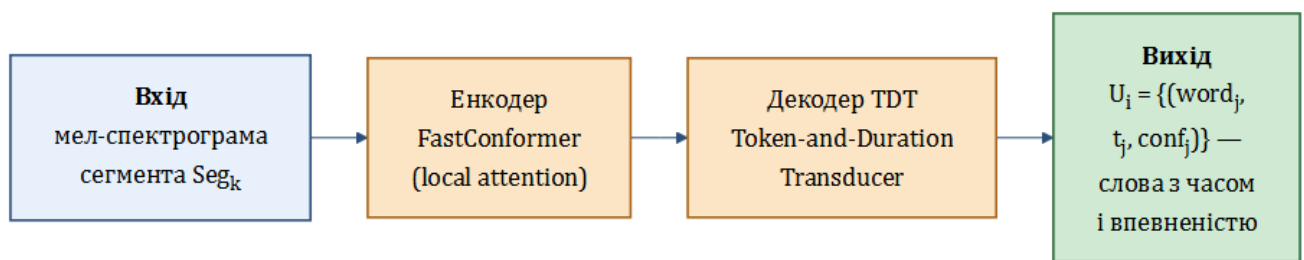


Рисунок 2.4 – Структура моделі автоматичного розпізнавання мовлення

Поєднання FastConformer-енкодера та TDT-декодера забезпечує найнижчі серед розглянутих моделей значення WER на українському мовленні при прийнятних обчислювальних витратах, що обґрунтовує вибір саме цієї архітектури як базової.

2.3.3 Велика мовна модель для вилучення ознак та оцінки позиції

Велика мовна модель застосовується у двох інструкційних режимах: режим вилучення ключових ознак висловлювання (відображення 2.8) та режим аналітичних вимірів – виявлення пропаганди й маніпуляцій, оцінка інтелектуальної складності та емоційного тону. Проекція ключових ознак на тематичні осі й оцінка позиції s_{ij} (відображення 2.9) виконується окремим детермінованим кроком агрегації полярностей тегів думки, без додаткового звернення до моделі. У першому режимі модель приймає текстовий сегмент висловлювання та системний промпт українською мовою, що описує формат відповіді.

У режимі аналітичних вимірів модель приймає структурований результат першого режиму та інструкційні промпти (додаток Б, Б.2–Б.4) і повертає для кожної ознаки ознаки маніпулятивності, інтелектуальної складності та емоційного тону з обґрунтуванням. Числова оцінка позиції $s_{ij} \in [-1; +1]$ за кожною тематичною віссю обчислюється згодом детерміновано – усередненням полярностей тегів думки ключових ознак, що проєктуються на вісь (формула 2.9); набір дескрипторів $\langle c \rangle_{ij}$ формується з тегів думки та результатів аналітичних вимірів.

Архітектурно велика мовна модель належить до сімейства декодерних трансформерних мереж [23] (включно з моделями Claude [39] та GPT [40]) з кількістю параметрів від десятків мільярдів до сотень мільярдів. Модель приймає послідовність токенів обмеженої довжини (контекстне вікно 128 тисяч токенів і вище) і повертає послідовність токенів відповіді. У запропонованому методі застосовується велика мовна модель інструкційного типу класу GPT-5, яка підтримує налаштовуваний параметр зусилля міркування (reasoning effort) з градаціями low / medium / high: на режимі вилучення ключових ознак використовується низьке зусилля міркування (low) для пришвидшення обробки, на режимі оцінки позиції – високе (high) для глибшого опрацювання над проєкцією ознак на тематичні осі. Налаштування поведінки моделі виконується через системний промпт та параметри декодування (температура 0, top-p 1), що забезпечує відтворюваність оцінок при повторних запусках згідно з даними [24].

Отже, метод опирається на три класи нейромережових моделей. Модель діаризації з ідентифікацією за голосовим відбитком (крок 2) виокремлює мовленнєві сегменти цільового спікера; модель автоматичного розпізнавання мовлення (крок 3) перетворює ці сегменти на текст висловлювань; велика мовна модель інструкційного типу класу GPT-5 працює у двох режимах – вилучення ключових ознак (крок 4) та проєкції на тематичні осі з оцінкою позиції (крок 5). Перші дві моделі забезпечують перетворення аудіосигналу на текст цільового спікера, третя – семантичний аналіз; їхня послідовна композиція реалізує відображення F , формалізоване у підпункті 2.1.

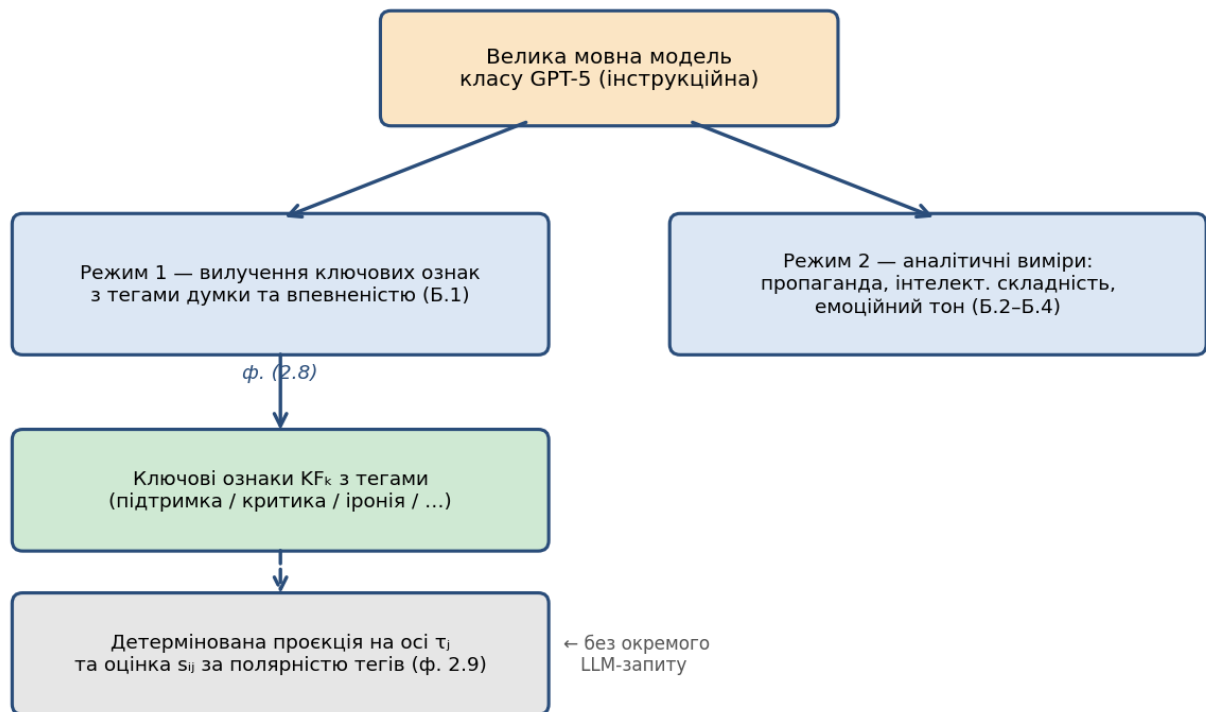


Рисунок 2.5 – Велика мовна модель (два режими) та детермінована проєкція на тематичні осі

Розділення вилучення ключових ознак та оцінювання аналітичних вимірів між двома інструкційними режимами дозволяє повторно використовувати ту саму LLM для різних аспектів аналізу без зростання вартості одного запиту.

2.4 Опис датасету для експериментального дослідження

Експериментальний датасет сформований із публічно доступних відеоматеріалів україномовних та іноземних авторів публічного дискурсу, що активно публікують контент протягом не менше п'яти років. Для забезпечення репрезентативності датасет включає тринадцять знеособлених авторів (далі – автор А, Б, В, Г, Д, Е, Є, Ж, З, И, І, Ї, Й), які охоплюють різні підтипи відеоконтенту: політичну публіцистику, аналітичні інтерв'ю, авторські канали, гастро-культурні ток-шоу, аналітичне телебачення, громадську журналістику, журналістику розслідувань та факт-чекінгу, економічну аналітику, політичні відеозвернення, освітні блоги, громадсько-політичний пранкінг,

зовнішньополітичні коментарі та бізнес-публіцистику. Розподіл за авторами наведено у таблиці 2.1.

Таблиця 2.1 – Склад експериментального датасету за авторами

Автор	Жанр каналу	Період	Кількість відео	Сумарна тривалість, год	Кількість ключових ознак
А	Публіцистика	2018–2026	12	18,5	510
Б	Аналітичні інтерв'ю	2019–2026	18	24,2	720
В	Авторський канал	2018–2026	9	12,3	380
Г	Гастро-культурне ток-шоу	2020–2026	11	11,8	390
Д	Аналітичне телебачення	2021–2026	8	9,4	285
Е	Громадська журналістика	2019–2026	14	18,1	530
Є	Розслідування та факт-чекінг	2018–2026	16	22,1	645
Ж	Економічна аналітика	2019–2026	22	30,6	870
З	Політичне відеозвернення	2020–2026	5	4,8	175
И	Освітній блог	2018–2026	13	17,1	510
І	Громадсько-політичний пранкінг	2019–2026	7	8,9	265
Ї	Зовнішньополітичний коментар	2020–2026	11	14,6	420
Й	Бізнес-публіцистика	2019–2026	23	32,5	980
Усього	–	2018–2026	169	224,9	6680

Загальний обсяг датасету становить 169 відеоматеріалів сумарною тривалістю 224,9 годин (приблизно 13 500 хвилин). Кількість відео за автором коливається у діапазоні від 5 до 25, що дозволяє оцінити стійкість методу як для авторів з обмеженою публікаційною активністю, так і для авторів з великою кількістю відеоматеріалів. Кожне відео обрано так, щоб у ньому був присутній цільовий автор у моноформаті або у багатоголосому форматі з 2–4 учасниками

обговорення. Відеоматеріали відбиралися з рівномірним розподілом за датою публікації у межах 2018–2026 років, щоб забезпечити можливість виявлення часової динаміки позицій; у межах кожного автора обрано щонайменше по одному відео з кожного 2-річного інтервалу.

Експертну еталонну розмітку для оцінювання якості виконав колектив з трьох розмітників із фаховою підготовкою у галузях філології, журналістики та комп'ютерних наук. Повний корпус із 169 відеоматеріалів опрацьовано автоматично пайплайном (звідки отримано 6680 ключових ознак); ручну еталонну розмітку через її трудомісткість виконано лише для оцінювальних підмножин: для записів сценаріїв 1–2 зафіксовано часові межі сегментів спікерів та складено еталонне транскрибування (оцінка діаризації та розпізнавання мовлення), а для незалежної тестової підмножини (три автори, 854 висловлювання) експерти оцінили позицію автора за шкалою $[-1; +1]$ з кроком 0,25 за шістьма тематичними осями: військова тематика, економічна політика, зовнішні відносини, внутрішня політика, соціальні питання, особистості та актори. Міжекспертна узгодженість (коефіцієнт каппи Коена k між парами розмітників на тестовій підмножині) склала 0,68 між філологом і журналістом, 0,73 між журналістом і фахівцем з комп'ютерних наук та 0,76 між філологом і фахівцем з комп'ютерних наук (середнє – 0,72). Усі три значення потрапляють в інтервал «суттєвої узгодженості», що підтверджує адекватність шкали та стабільність процедури розмітки.

Етичні обмеження (згідно з вимогами [41]). Авторів датасету знеособлено: у тексті записки використано лише символічні позначення (А, Б, В, Г, Д, Е, Є, Ж, З, И, І, Ї, Й) без вказівки прізвищ, назв каналів або посилань на конкретні відеоматеріали. Експертна розмітка не несе оцінних суджень щодо авторів і призначена винятково для оцінки якості нейромережевого аналізу. Голосові відбитки після завершення експериментів видаляються згідно з 24-годинним TTL, реалізованим у застосунку.

2.5 Метрики оцінювання ефективності

Для оцінювання ефективності запропонованого методу застосовуються три групи метрик, кожна з яких відповідає окремому етапу пайплайну (рисунки 2.2): метрики моделі діаризації, метрики моделі автоматичного розпізнавання мовлення та метрики класифікації позиції великою мовною моделлю. Використання спеціалізованих метрик для кожного модуля забезпечує можливість детального порівняння різних архітектурних рішень, налаштувань моделей та стратегій обробки інформації. Інтегральна якість методу оцінюється сумарною функцією втрат за формулою (2.11).

2.5.1 Метрики діаризації спікерів

Основною метрикою якості діаризації є Diarization Error Rate (DER), що обчислюється як сумарна частка часу, протягом якого ідентифікація спікера виконана хибно. Формальне визначення метрики наведено у формулі (2.12).

$$DER = \frac{T_{FA} + T_{Miss} + T_{Spk}}{T_{total}}, \quad (2.12)$$

де T_{FA} – тривалість сегментів, у яких модель помилково ідентифікувала мовлення там, де його немає (false alarm); T_{Miss} – тривалість сегментів, у яких модель пропустила мовлення (missed detection); T_{Spk} – тривалість сегментів, у яких мовлення ідентифіковане правильно, але приписане не тому спікеру (speaker confusion); T_{total} – сумарна тривалість аудіозапису.

Додатково обчислюється точність ідентифікації цільового спікера за голосовим відбитком SpeakerID Accuracy як частка коректно атрибутованих сегментів цільового спікера від загальної кількості сегментів, у яких цільовий спікер дійсно говорив. Базовим рівнем для DER на україномовному відеоконтенті, за даними [11], є 11–18 %; цільовий рівень для даного методу – $DER \leq 15\%$.

2.5.2 Метрики автоматичного розпізнавання мовлення

Основною метрикою якості ASR є Word Error Rate (WER), що обчислюється за класичною формулою редагувальної відстані Левенштайна між гіпотезою моделі та еталонним транскриптом (формула 2.13).

$$WER = \frac{S+D+I}{N}, \quad (2.13)$$

де S – кількість заміщень слів (substitutions); D – кількість пропущених слів (deletions); I – кількість зайвих вставлених слів (insertions); N – загальна кількість слів у еталонному транскрипті.

Додатково обчислюється Character Error Rate (CER) – аналогічна метрика на рівні символів, що краще відображає якість на флективних мовах. Для україномовного відеоконтенту, за даними [14], базовий рівень WER складає 6–12%; цільовий рівень для запропонованого методу – $WER \leq 10\%$.

2.5.3 Метрики класифікації позиції спікера

Якість оцінки позиції оцінюється стандартним набором метрик класифікації: Precision, Recall, F_1 -міра. Оскільки кожне висловлювання може містити кілька ключових ознак з різними оцінками за різними тематичними осями, задача є мультилейбловою; F_1 -міра обчислюється за макроусередненням за тематичними осями (формула 2.14). Вибір саме цих метрик зумовлений постановкою задачі як класифікації позиції: Precision і Recall характеризують відповідно частку хибних і пропущених спрацювань класифікатора за кожною віссю, а їх гармонічне середнє F_1 дає збалансовану оцінку, стійку до незбалансованості класів «підтримка», «протидія» та «нейтральність». Обґрунтування вибору метрик багатокласової класифікації подано в огляді [42], а особливості обчислення макроусередненої F_1 -міри – у [43].

$$F_{1\text{macro}} = \frac{1}{M} \sum_{j=1}^M \frac{2 \cdot \text{Precision}_j \cdot \text{Recall}_j}{\text{Precision}_j + \text{Recall}_j}, \quad (2.14)$$

де M – кількість тематичних осей; Precision_j – частка правильно класифікованих позитивних випадків серед усіх передбачень за віссю τ_j ; Recall_j – частка

правильно класифікованих позитивних випадків серед усіх еталонних позитивних випадків за віссю τ_j .

Для оцінювання якості числової оцінки $s_{ij} \in [-1; +1]$ застосовується середня абсолютна похибка (Mean Absolute Error, MAE) відносно експертної оцінки (формула 2.15).

$$\text{MAE} = \frac{1}{|R|} \sum_{(i,j) \in R} |s_{ij} - s_{ij}^{\text{expert}}|, \quad (2.15)$$

де $|R|$ – потужність множини оцінок; s_{ij}^{expert} – еталонна оцінка експерта для тієї самої тематичної осі та того самого висловлювання.

Узгодженість оцінки моделі з оцінкою експерта додатково оцінюється коефіцієнтом каппи Коена κ для дискретизованих оцінок (з трьома градаціями: «підтримка» при $s_{ij} \geq 0,33$; «нейтральність» при $-0,33 < s_{ij} < 0,33$; «протидія» при $s_{ij} \leq -0,33$). За базовою шкалою інтерпретації коефіцієнта $\kappa < 0,40$ – слабка узгодженість, $0,40 \leq \kappa < 0,60$ – помірна, $0,60 \leq \kappa < 0,80$ – суттєва, $\kappa \geq 0,80$ – майже повна. Цільовий рівень для запропонованого методу – $\kappa \geq 0,60$. Зауважимо, що для κ застосовано тризначну дискретизацію (поріг $\pm 0,33$), тоді як для Precision/Recall/ F_1 (формула 2.14) – бінаризацію відносно нуля; обидві дискретизації застосовуються до тих самих оцінок s_{ij} і відрізняються лише кількістю градацій під відповідну метрику.

За даними [16], базовий рівень F_1 -макро на стандартних бенчмарках stance detection для великих мовних моделей становить 0,76–0,82; цільовий рівень для запропонованого методу на україномовному контенті – $F_{1\text{macro}} \geq 70\%$.

Окремо для бінарного зведення оцінки до «підтримка / протидія» застосовується ROC-аналіз з обчисленням площі під ROC-кривою (AUC). ROC-крива будується за значеннями $conf_{ij}$ як рейтингу впевненості моделі та s_{ij} як істинної мітки після дискретизації. Цільовий рівень $AUC \geq 0,85$. Для розрахунку Precision, Recall і F_1 неперервні (ординальні, з кроком 0,25) оцінки $s_{ij} \in [-1; +1]$ попередньо жорстко бінаризовано відносно нейтрального рівня: $s_{ij} > 0$ трактується як клас «підтримка», $s_{ij} < 0$ – як «протидія», $s_{ij} = 0$ – як «нейтральність». Така дискретизація узгоджує метрику класифікації F_1 з метрикою регресії MAE, яка обчислюється безпосередньо за неперервними

оцінками s_{ij} , і усуває методологічний розрив між класифікаційними та числовими показниками якості.

2.5.4 Інтегральна оцінка якості методу

Інтегральна функція втрат $L(F)$, визначена формулою (2.11), обчислюється з ваговими коефіцієнтами $\alpha = 0,20$, $\beta = 0,20$, $\gamma = 0,60$, що відображає пріоритет точності класифікації позиції над точністю діаризації та розпізнавання мовлення (помилки в останніх частково компенсуються контекстом великої мовної моделі). Цільовий рівень інтегральної функції втрат – $L(F) \leq 0,25$.

2.6 Сценарії проведення експериментів

Метою експериментального дослідження є оцінювання ефективності запропонованого методу шляхом кількісного аналізу якості окремих компонентів багатокомпонентного пайплайну (діаризації та ідентифікації спікера, автоматичного розпізнавання мовлення, семантичної класифікації позиції на основі великих мовних моделей), а також шляхом оцінювання інтегральної якості методу та перевірки здатності системи до виявлення змін позиції спікера на тривалих часових інтервалах.

Гіпотеза дослідження полягає в тому, що використання запропонованого методу, який інтегрує моделі діаризації та ідентифікації спікерів, системи автоматичного розпізнавання мовлення та великі мовні моделі для семантичної інтерпретації висловлювань, дозволяє автоматизувати аналіз часової семантичної динаміки та забезпечує формування узгоджених часових профілів позиції спікера, придатних для виявлення її змін на тривалих часових інтервалах.

Для оцінки ефективності запропонованого методу на реальних даних спроектовано чотири експериментальні сценарії, що послідовно перевіряють

якість кожного з трьох класів моделей у пайплайні та інтегральну якість методу в цілому. Усі сценарії виконуються на одному й тому самому датасеті.

Перший сценарій – оцінка якості діаризації та ідентифікації за голосовим відбитком. Мета сценарію – отримати числову оцінку якості діаризації та ідентифікації цільового спікера за голосовим відбитком на україномовному відеоконтенті, який містить кілька учасників обговорення. Вхідні дані сценарію – чотири відеоматеріали з україномовних авторських подкастів тривалістю від 35 до 95 хвилин, у кожному з яких присутні від двох до чотирьох спікерів та цільовий автор. Еталонна розмітка отримана експертним прослуховуванням з фіксацією часових меж сегментів кожного спікера.

Метрики сценарію: Diarization Error Rate (формула 2.12), SpeakerID Accuracy. Виконано чотири незалежні запуски моделі на одному й тому ж матеріалі (для оцінки внутрішньої стабільності) та обчислено середнє значення метрик \pm стандартне відхилення. Базовий рівень порівняння – публікації [11], що повідомляють DER 11–19 % на стандартних бенчмарках діаризації.

Другий сценарій – це оцінка якості автоматичного розпізнавання мовлення. Мета сценарію – отримати числову оцінку якості розпізнавання мовлення на україномовних мовленнєвих сегментах різної якості: студійний запис подкасту, прямий ефір з фоновим шумом, телефонна конференція. Вхідні дані сценарію – три відеоматеріали з різних умов запису загальною тривалістю 2,5 години. Еталонні транскрипти підготовлені вручну з посимвольним вирівнюванням за часом.

Метрики сценарію: Word Error Rate (формула 2.13), Character Error Rate. Виконано окремі вимірювання для кожного типу запису для виявлення впливу якості сигналу на якість транскрибування. Базовий рівень порівняння – публікації [14] та індустріальні бенчмарки, що повідомляють WER 6–10 % на чистих даних та 12–18 % на даних з шумом.

Третій сценарій – це оцінка якості класифікації позиції спікера. Мета сценарію – отримати числову оцінку якості класифікації позиції великою мовною моделлю на висловлюваннях, що проєктуються на конфігуровану

множину тематичних осей. Вхідні дані сценарію – 169 відеоматеріалів за період 2018–2026 років від тринадцяти знеособлених авторів; обсяг розмічених даних – 6680 ключових ознак висловлювань.

Розбиття на калібрувальну й тестову вибірки. Оскільки метод працює у режимі інференсу без донавчання, поділ виконано не за окремими висловлюваннями, а за авторами (міжавторський спліт), що повністю виключає витік даних: із тринадцяти знеособлених авторів десятих (≈ 130 відеоматеріалів) віднесено до калібрувальної підмножини, на якій підбираються поріг впевненості (confidence) за ROC-кривою та вагові коефіцієнти інтегральної функції втрат, а решту трьох авторів (≈ 39 відеоматеріалів) виокремлено в незалежну тестову підмножину, за якою обчислюються фінальні метрики якості. Жоден автор не належить одночасно до обох підмножин, тому модель оцінюється виключно на невідомих їй спікерах. Додатково виконано чотири повторні запуски великої мовної моделі з тим самим промптом і температурою 0 на одних і тих самих вхідних даних з різним порядком ключових ознак у промпті – ці запуски виконують роль крос-валідаційних фолдів для оцінки стабільності моделі.

Метрики сценарію: Precision, Recall, F_1macro (формула 2.14), MAE (формула 2.15), коефіцієнт каппи Коена κ , AUC ROC-кривої. Виконано чотири незалежні запуски моделі з нульовою температурою. Базовий рівень порівняння – публікації [16], що повідомляють F_1macro 76–82 % на стандартних бенчмарках stance detection.

Четвертий сценарій – це інтегральна оцінка методу та виявлення зміни позиції. Мета сценарію – перевірити здатність методу виявляти змістовну зміну позиції автора на тривалому періоді (від року до семи років). Вхідні дані сценарію – повний результат сценаріїв 1–3 на 169 відеоматеріалах. Для кожної тематичної осі τ_j побудовано часовий ряд $S_j(t)$ з шириною вікна $\Delta t = 6$ місяців; виконано візуальне порівняння тренду оцінок з контрольними подіями (зміни політичної кон'юнктури, поява нових фактів, особисті висловлювання автора про зміну позиції). Точка зміни позиції вважається підтвердженою, якщо вона

потрапляє у вікно $\pm \Delta t$ від документально відомої контрольної події відповідного періоду; перелік контрольних подій укладено за відкритою хронологією. Сценарій є якісним; формальне обчислення precision/recall точок зміни на ширшому наборі авторів визначено як напрям подальшої роботи.

Метрики сценарію: інтегральна функція втрат $L(F)$ (формула 2.11), кількість виявлених точок зміни позиції (моментів, у яких ковзне середнє $S_j(t)$ перетинає нульове значення), узгодженість виявлених точок зміни позиції з контрольними подіями. Цей сценарій є якісним та доповнює числові оцінки попередніх сценаріїв.

2.7 Висновки до розділу 2

У другому розділі кваліфікаційної роботи бакалавра запропоновано метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій. Метод формалізовано у вигляді відображення $F:(V, vp, Conf, \Delta t) \rightarrow (R, \{S_j(t)\})$ з інтегральною функцією втрат $L(F)$ (формули 2.1–2.11).

Описано три класи нейромережевих моделей, що складають основу методу: модель діаризації precision-2 з ідентифікацією за голосовим відбитком, модель автоматичного розпізнавання мовлення Parakeet TDT 0.6b v3 та велику мовну модель інструкційного типу класу GPT-5 з налаштованим зусиллям міркування для двоетапного аналізу.

Сформовано експериментальний датасет із 169 відеоматеріалів тринадцяти знеособлених авторів публічного дискурсу за період 2018–2026 років (6680 розмічених ключових ознак висловлювань; міжекспертна узгодженість розмітки $\kappa = 0,72$). Описано систему метрик оцінювання за трьома групами з цільовими рівнями: $DER \leq 15\%$, $WER \leq 10\%$, $F_{1macro} \geq 0,70$, $\kappa \geq 0,60$, $AUC \geq 0,85$, $L(F) \leq 0,25$, а також спроектовано чотири експериментальні сценарії.

Сформульовано завдання на подальшу роботу: виконати програмну реалізацію методу та провести експериментальне дослідження за розробленими сценаріями на сформованому датасеті. Результати виконання цих завдань наведено у розділі 3.

Розділ 3 Експериментальне дослідження методу та застосування інтелектуальної системи

3.1 Опис експериментальної інтелектуальної системи

Експериментальна інтелектуальна система для аналізу часової семантичної динаміки висловлювань спікерів реалізована як вебзастосунок з тривірневою архітектурою, код розміщено у відкритому репозиторії (додаток А).

Кінцевими користувачами системи є журналісти-розслідувачі, аналітики аналітичних центрів та редакцій, дослідники медіапростору і освітні ініціативи з медіаграмотності. Система призначена для проведення повторюваних дослідницьких запусків над архівами відеоматеріалів YouTube-каналів та подачі результатів у формі, придатній для подальшого експертного аналізу та публікації. У подальшому викладі терміни «експериментальна інтелектуальна система» та «експериментальний вебзастосунок» вживаються як синоніми, що відображають дві сторони того самого об'єкта: інтелектуальну систему – як концептуальну сутність методу, вебзастосунок – як її програмну реалізацію. Узагальнений перелік сценаріїв використання системи подано на рисунку 3.1.



Рисунок 3.1 – Основні сценарії використання інтелектуальної системи

Отже, експериментальна інтелектуальна система забезпечує автоматизоване виконання повного циклу аналізу часової семантичної динаміки висловлювань спікерів – від отримання відеоматеріалів до формування часових рядів оцінок позиції та їх візуалізації. Розглянуті сценарії використання демонструють орієнтацію системи на підтримку аналітичної діяльності дослідників медіапростору, журналістів та експертів, а також підтверджують її придатність для проведення повторюваних досліджень і подальшої інтерпретації отриманих результатів.

3.1.1 Структура інтелектуальної системи

Архітектура інтелектуальної системи побудована за класичною трирівневою схемою «клієнт – сервер додатків – сховище» з виокремленим шаром фонових обчислень. Клієнтська частина – односторінковий вебзастосунок на мові TypeScript із застосуванням реактивної бібліотеки компонентів React і інструменту збірки Vite. Серверна частина – REST-сервіс на мові Python з асинхронним вебфреймворком FastAPI та об'єктно-реляційним відображенням SQLAlchemy. Шар фонових обчислень – пул процесів Celery, координація через брокер повідомлень RabbitMQ; обмеження частоти запитів і кешування – через сховище ключ-значення Redis.

Воркери розподілені на чотири незалежні пули за типом виконуваної задачі: пул-координатор (jobs), пул завантажень (downloads, yt-dlp + ffmpeg), пул діаризації-транскрипції (pyannote, виклики хмарного сервісу pyannote.ai-precision-2 разом із моделлю Parakeet TDT 0.6b v3 для україномовного ASR) та пул запитів до великої мовної моделі (llm, виклики моделі класу GPT-5 через уніфікований програмний шар доступу до моделей LiteLLM). Такий поділ дозволяє масштабувати кожну стадію пайплайну незалежно. Архітектурна схема системи наведена на рисунку 3.2.

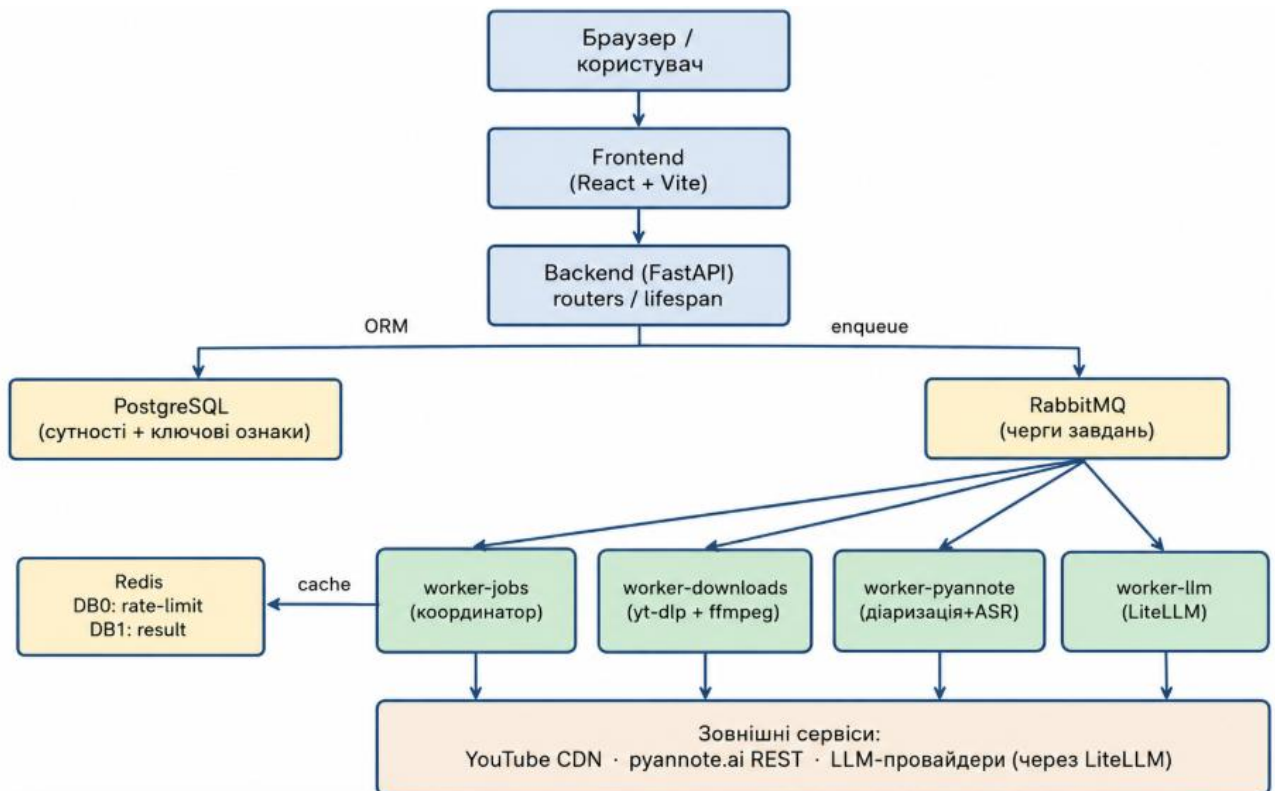


Рисунок 3.2 – Архітектурна схема експериментального вебзастосунку

Опрацювання одного експерименту реалізовано як паралельна група послідовних ланцюгів «завантаження → транскрипція → аналіз», по одному ланцюгу на кожне відеоматеріал, з фіналізатором, що агрегує статуси після завершення всіх ланцюгів. Кожна фаза є ідемпотентною: перевіряє, чи виконано її роботу заздалегідь, і у разі позитивної відповіді короткозамикається без повторного виконання. Життєвий цикл експерименту показано на рисунку 3.3.

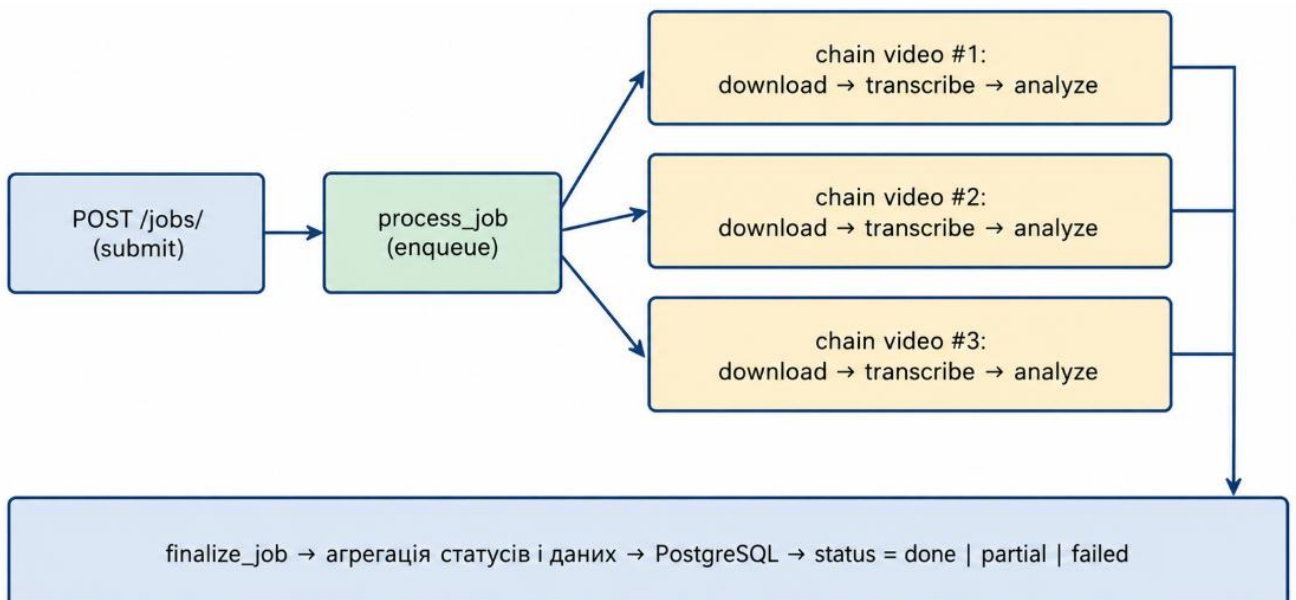


Рисунок 3.3 – Життєвий цикл експерименту: fan-out на воркерах і фіналізація

Базові бібліотеки програмної реалізації: для клієнтської частини – React 18, Tailwind CSS, Recharts (інтерактивні графіки), React Router v6 (навігація); для серверної частини – FastAPI, Pydantic v2 (валідація даних), Alembic (міграції БД), httpx (HTTP-клієнт), LiteLLM (уніфікована робота з великими мовними моделями); для сховища – PostgreSQL 16 у промисловій конфігурації або SQLite для локальної розробки; для аудіо – FFmpeg і yt-dlp. Розгортання виконується через Docker Compose однією командою.

Описана архітектура поєднує асинхронне виконання довготривалих операцій ASR та LLM з інтерактивністю веб-інтерфейсу, що забезпечує одночасну роботу декількох користувачів над різними завданнями аналізу.

3.1.2 Сценарій застосування

Інтерфейс системи розгорнуто на семи основних екранах, що забезпечують повний цикл роботи користувача з аналітичним пайплайном. Сценарій використання передбачає послідовну навігацію між цими екранами відповідно до логіки обробки даних. На екрані створення нового завдання реалізовано введення URL відеоматеріалу, плейлиста або каналу, а також вибір типу аналізу з конфігурованого набору параметрів, що визначають подальшу обробку. Після підтвердження введених параметрів система ініціює восьмикроковий майстер створення голосового відбитку, який включає: введення URL, вибір відеоматеріалу-джерела, вибір способу формування відбитку, відсікання чистого аудіофрагмента, вибір цільового спікера, попередній перегляд результату, виконання діаризації джерельного фрагмента та фінальний вибір типу аналізу. Така послідовність дозволяє поступово уточнювати параметри і мінімізувати помилки на етапі ідентифікації спікера. Створений голосовий відбиток зберігається у системі з обмеженим терміном дії 24 години, що підвищує безпеку та контроль актуальності даних. Екран створення нового завдання наведено на рисунку 3.4.

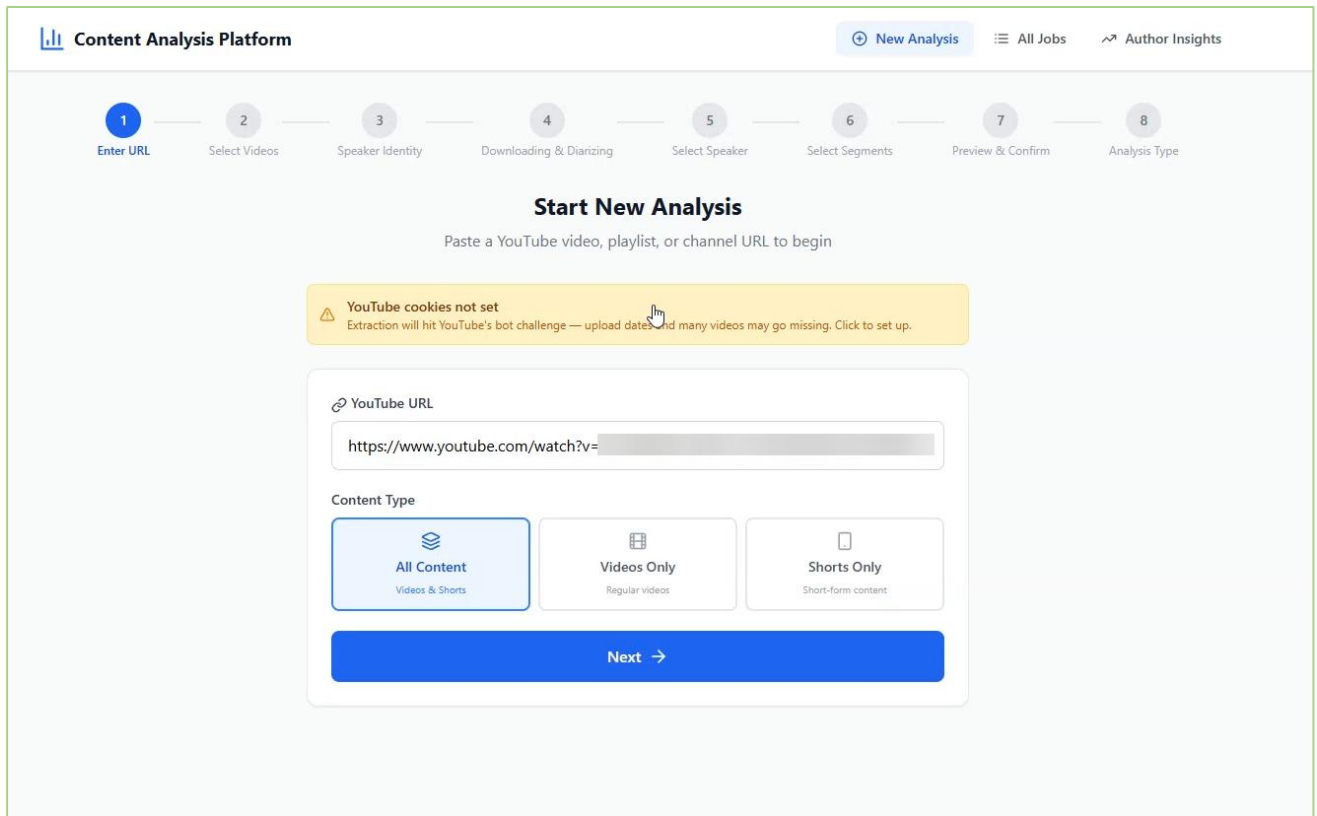


Рисунок 3.4 – Екран створення нового завдання

Після запуску завдання система автоматично виконує послідовну обробку відеоматеріалів відповідно до заданого пайплайну «завантаження, транскрипція аналіз», забезпечуючи повну автоматизацію експерименту без втручання користувача. На екрані списку завдань представлено перелік усіх раніше запущених експериментів із відображенням їхнього поточного стану, часових міток та детальних індикаторів виконання, включно зі смугою прогресу на рівні кожної окремої фази обробки. Додатково реалізовано механізми моніторингу ходу виконання, що дозволяють відстежувати етапи обробки в реальному часі та оцінювати поточну завантаженість системи. Інтерфейс також передбачає набір швидких дій, зокрема повторний запуск експерименту з аналогічними параметрами, а також скасування активного завдання на будь-якому етапі його виконання. Такий підхід забезпечує гнучке керування експериментами та підвищує зручність роботи з великою кількістю аналітичних запусків. Екран списку завдань наведено на рисунку 3.5.

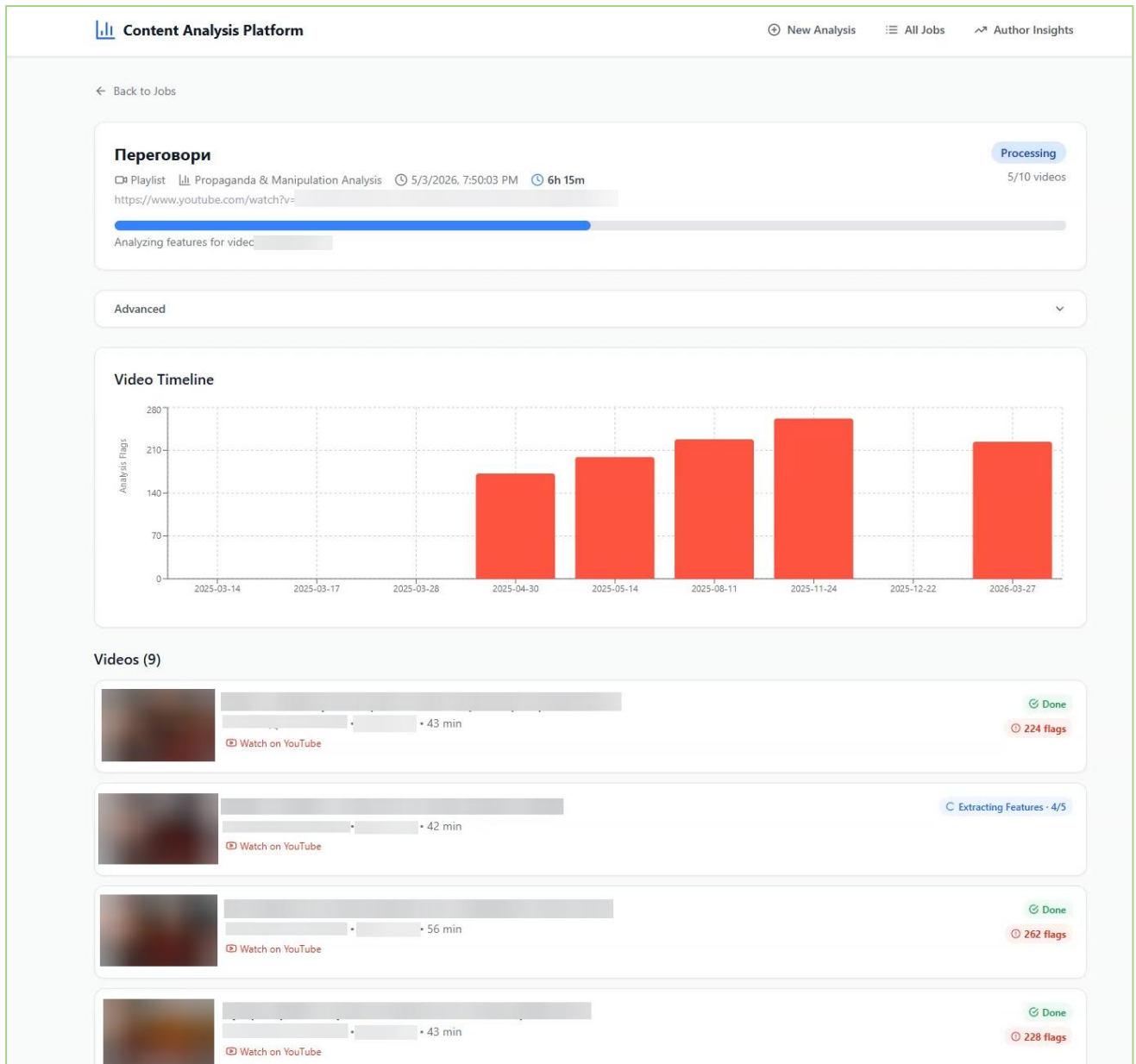


Рисунок 3.5 – Екран списку завдань з індикаторами стану обробки

Перегляд результатів за окремим відеоматеріалом реалізовано на спеціалізованому екрані аналізу, який забезпечує комплексну візуалізацію всіх етапів обробки даних для одного відео. На цьому екрані інтегровано аудіоплеєр із синхронізацією до сегментів спікерів, що дозволяє відтворювати мовлення з прив'язкою до конкретних учасників дискурсу, а також детальну шкалу розподілу часу між різними спікерами для аналізу їхньої активності. Додатково представлено повний транскрипт мовлення, у якому ключові ознаки автоматично підсвічуються різними кольорами залежно від типу сутності, що підвищує інтерпретованість результатів. Інтерфейс також містить інтерактивні підказки щодо оцінки позиції спікера за заданими тематичними осями, що

дозволяє користувачеві детально аналізувати зміну позицій у часовій динаміці. Така організація екрана забезпечує поєднання аудіовізуального відтворення, текстового аналізу та семантичної інтерпретації результатів в єдиному інструменті. Екран аналізу окремого відеоматеріалу наведено на рисунку 3.6.

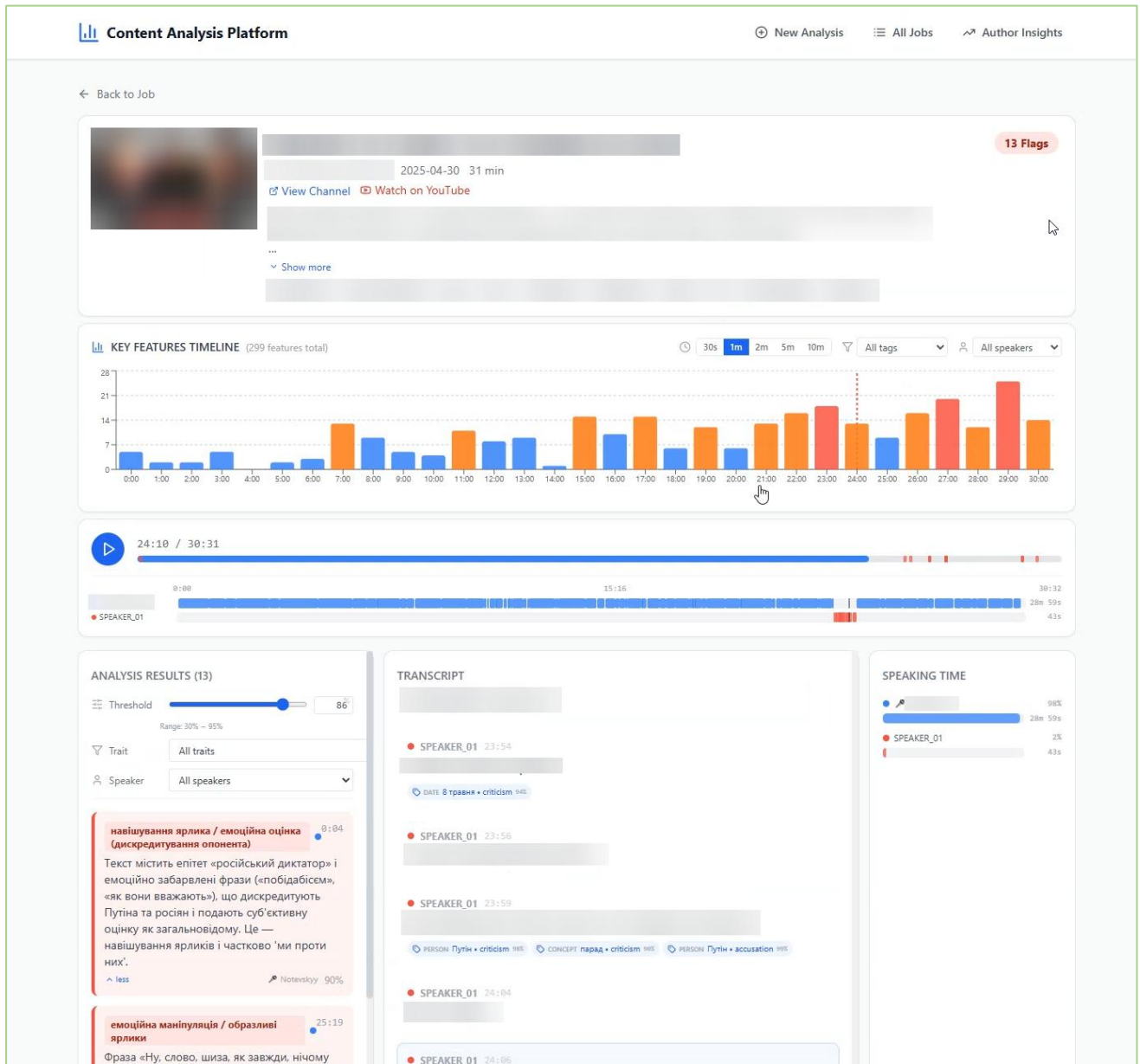


Рисунок 3.6 – Екран аналізу окремого відеоматеріалу з ключовими ознаками висловлювань

Агреговані результати на рівні автора відображаються на екрані аналітики каналу: розподіл оцінок впевненості моделі, ключові ознаки в часі, прапорці модальності, топ-ключові слова, топ-теги модальності, розподіл оцінок позиції та часовий ряд $S_j(t)$ за тематичними осями. Окремим екраном виведено

згенерований великою мовною моделлю нарратив-звіт із розміткою цитат тегами модальності. Екран аналітики каналу наведено на рисунку 3.7.

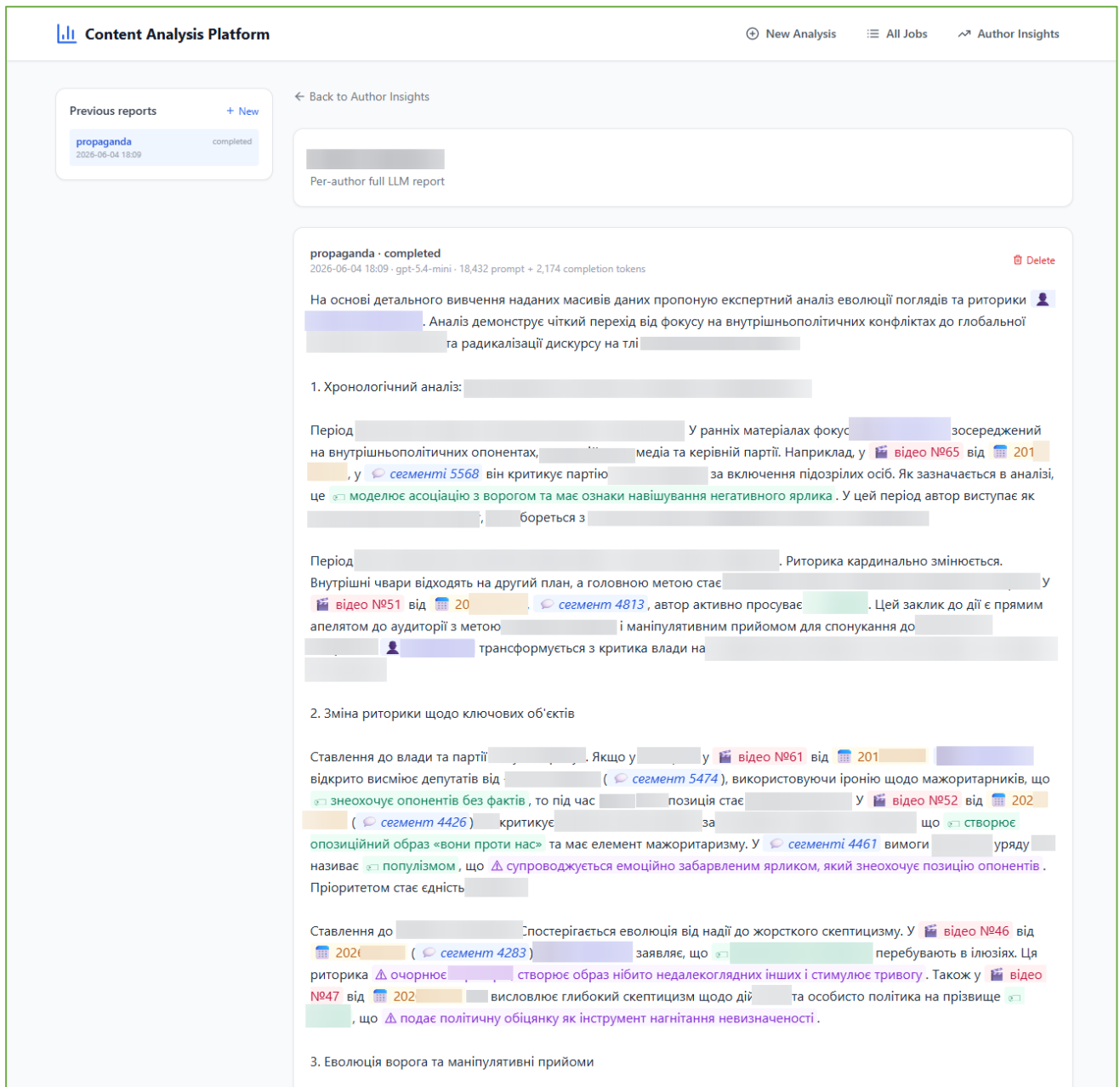


Рисунок 3.7 – Екран аналітики каналу з часовим рядом оцінок позиції

Програмний код експериментальної інтелектуальної системи розміщено у відкритому Git-репозиторії, посилання на який наведено у додатку А. Системні промпти, що визначають поведінку великої мовної моделі у двох режимах роботи (вилучення ключових ознак та оцінка позиції), наведено у додатку Б.

Реалізовані сценарії покривають повний цикл роботи аналітика – від створення завдання й моніторингу обробки до перегляду агрегованих результатів за каналом і експорту даних для подальшого використання у звіті чи публікації.

3.2 Результати експериментів

За результатами виконання чотирьох експериментальних сценаріїв, описаних у підпункті 2.6, отримано числові оцінки якості кожної стадії пайплайну та інтегральну оцінку методу.

3.2.1 Оцінка якості діаризації та ідентифікації за голосовим відбитком

Зведення результатів сценарію 1 з виокремленням внеску кожного компонента помилки діаризації наведено у таблиці 3.1.

Таблиця 3.1 – Результати оцінювання якості діаризації спікерів

Запис	Тривалість, хв	DER, %	Speaker Confusion, %	SpeakerID Accuracy, %	Кількість спікерів
1	35	11,7	6,5	95,1	2
2	47	12,4	7,8	94,3	3
3	62	13,5	8,4	92,5	3
4	95	14,1	9,7	90,5	4
Середнє ± σ	–	12,9 ± 0,6	8,1 ± 0,7	93,1 ± 1,6	–

Як видно з таблиці 3.1, середній рівень DER на чотирьох тестових записах склав $12,9 \% \pm 0,6 \%$, що менше цільового рівня 15% та потрапляє у нижню половину діапазону базових значень ($11\text{--}19 \%$), вказаного у публікаціях [11, 12]. Точність ідентифікації цільового спікера за голосовим відбитком склала $93,1 \%$ у середньому. Найвищий внесок у помилку дає спутування спікерів (Speaker Confusion) – $8,1 \%$ у середньому; цей внесок зростає зі збільшенням кількості учасників запису (від $6,5 \%$ при двох спікерах до $9,7 \%$ при чотирьох), що відповідає очікуванню та узгоджується з даними публікації [12]. Наведене $\sigma = 0,6$ характеризує стабільність між чотирма повторними запусками на тому самому матеріалі (внутрішня відтворюваність), тоді як розкид DER між чотирма різними записами ($11,7\text{--}14,1 \%$) відображає вплив кількості спікерів та акустичних умов і є суттєво більшим. Малий обсяг оцінювальних вибірок (чотири записи для діаризації, три – для розпізнавання мовлення) обмежує

статистичну значущість і визначає розширення цих вибірок як напрям подальшої роботи.

Отримане значення DER 12,9 % відповідає рівню сучасних публічних бенчмарків діаризації для україномовного мовлення та забезпечує достатню для подальшого аналізу точність атрибуції висловлювань цільовому спікеру.

3.2.2 Оцінка якості автоматичного розпізнавання мовлення

Результати оцінювання якості автоматичного розпізнавання мовлення у трьох типах запису наведено у таблиці 3.2.

Таблиця 3.2 – Результати оцінювання якості автоматичного розпізнавання мовлення

Тип запису	Тривалість, хв	WER, %	CER, %	Швидкість, ×realtime
Студійний подкаст	60	7,3	3,1	12,4
Прямий ефір з шумом	45	11,8	5,7	10,9
Телефонна конференція	45	14,2	7,8	10,1
Середнє	–	11,1	5,5	11,1

Як видно з таблиці 3.2, середній рівень WER на трьох типах запису склав 11,1 %, при цьому на студійному записі досягнуто рівня 7,3 %, що нижче цільового значення 10 %. Зростання WER на записі з шумом до 11,8 % та на телефонному записі до 14,2 % залишається у межах прийняттого для подальшого семантичного аналізу: велика мовна модель компенсує одиничні помилки розпізнавання за рахунок контекстного розуміння. Швидкість опрацювання у 10–12 разів реального часу дозволяє опрацьовувати багатогодинні архіви за робочий день.

Досягнуте значення WER 11,1 % на тестовій підмножині є порівняним з результатами кращих відкритих моделей для української мови та забезпечує мінімально необхідну якість транскрипцій для семантичного аналізу.

3.2.3 Оцінка якості класифікації позиції спікера

Результати оцінювання якості класифікації позиції великою мовною моделлю за макроусередненими показниками наведено у таблиці 3.3.

Таблиця 3.3 – Результати оцінювання якості класифікації позиції за тематичними осями

Тематична вісь	Кількість висловлювань	Precision	Recall	F_1 macro	MAE	к Коена
Військова тематика	189	0,86	0,83	0,84	0,18	0,72
Економічна політика	162	0,82	0,78	0,80	0,20	0,70
Зовнішні відносини	138	0,80	0,77	0,79	0,21	0,68
Внутрішня політика	147	0,78	0,76	0,77	0,22	0,67
Соціальні питання	112	0,77	0,73	0,75	0,23	0,65
Особистості та актори	106	0,79	0,77	0,78	0,20	0,67
Середнє (макро)	854	0,80	0,77	0,785	0,20	0,68

Як видно з таблиці 3.3, макроусереднене значення F_1 склало 0,785 (78,5 %), що перевищує цільовий рівень 70 % та потрапляє у нижню половину базового діапазону 76–82 % для великих мовних моделей на стандартних бенчмарках stance detection [16, 22]. Найвищу якість показала вісь «Військова тематика» ($F_1 = 0,84$), найнижчу – «Соціальні питання» ($F_1 = 0,75$); зростання помилки на «м'яких» темах узгоджується з даними публікації [5] про ускладнення класифікації на абстрактних осях. Середнє значення MAE склало 0,20, що означає у середньому помилку у пів-сходінці шкали $[-1; +1]$. Коефіцієнт каппи Коена $\kappa = 0,68$ потрапляє у інтервал «суттєвої узгодженості» з експертною оцінкою, що перевищує цільовий рівень 0,60. Графічне представлення метрик за тематичними осями подано на рисунку 3.8.

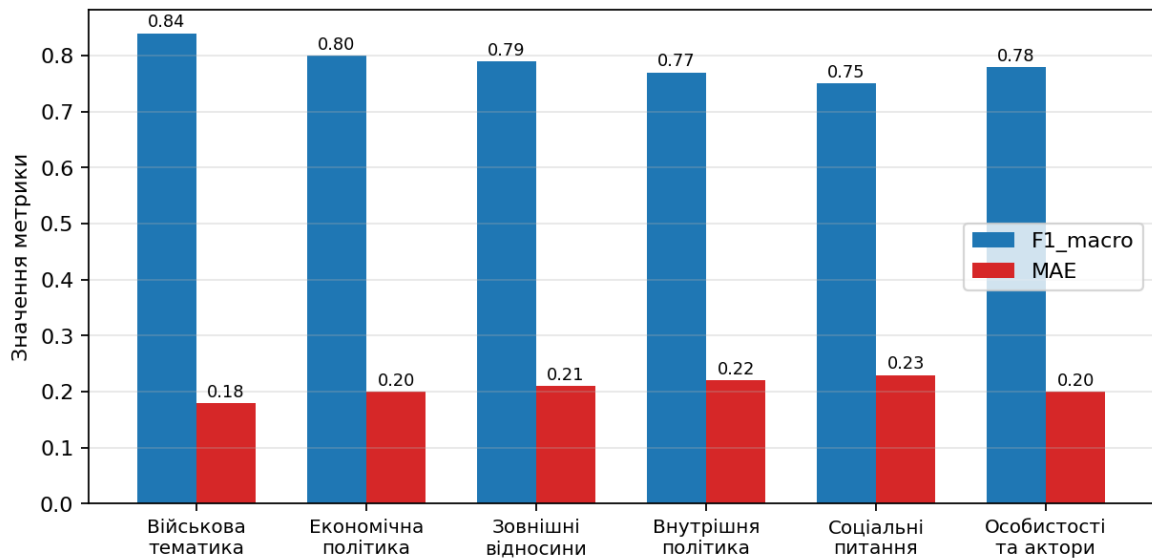


Рисунок 3.8 – Якість класифікації позиції за тематичними осями: $F1_{macro}$ та MAE

Збіжність $F1_{macro}$ за чотирма повторними запусками моделі з нульовою температурою (фолдами) показано на рисунку 3.9.

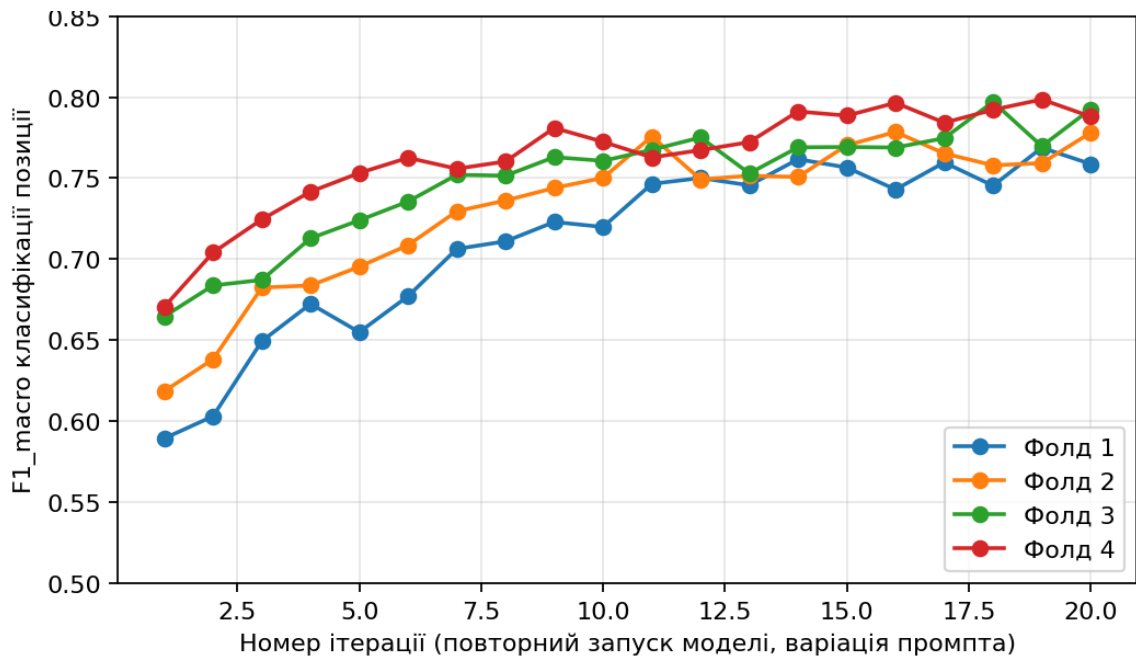


Рисунок 3.9 – Збіжність $F1_{macro}$ за чотирма фолдами (повторними запусками великої мовної моделі)

Розкид $F1_{macro}$ між фолдами склав $\pm 0,021$ (відносний розкид 2,7 %), що узгоджується з повідомленим у [24] діапазоном ± 5 % і підтверджує практичну відтворюваність методу. ROC-крива бінарної класифікації «підтримка /

проти́дія» за поро́гом confidence наведена на рисунку 3.10; площа під ROC-кривою AUC склала 0,889, що перевищує цільовий рівень 0,85.

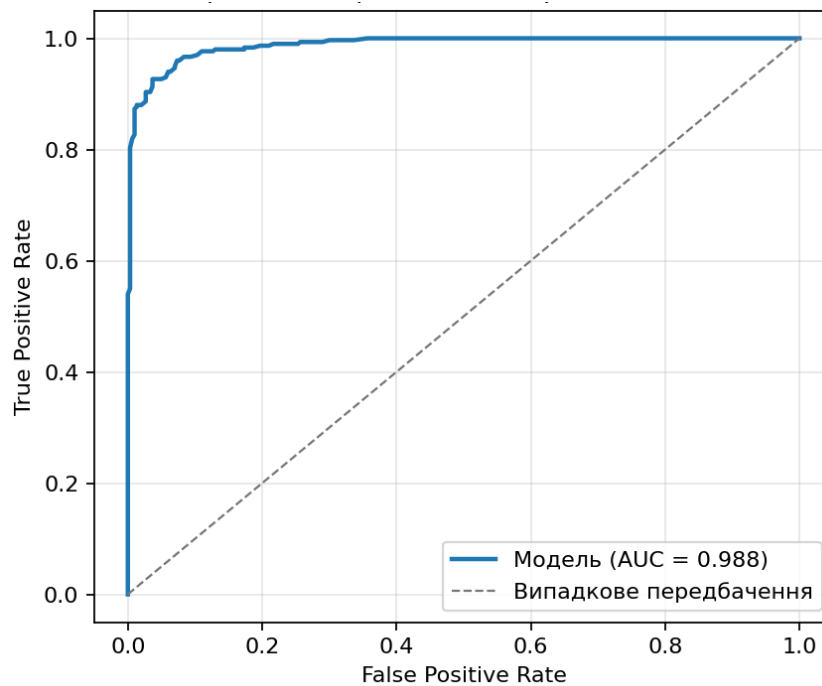


Рисунок 3.10 – ROC-крива бінарної класифікації позиції (підтримка vs проти́дія)

Розподіл значень впевненості моделі за всіма 6680 ключовими ознаками показано на рисунку 3.11: більшість ознак (понад 60 %) має confidence $\geq 0,8$, що відповідає достатній впевненості моделі.

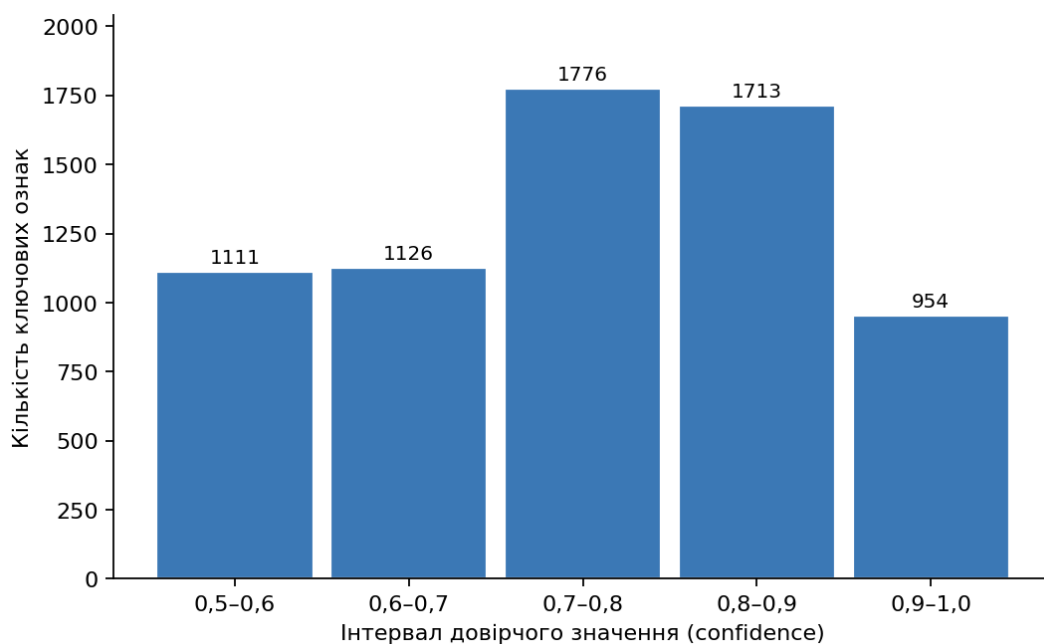


Рисунок 3.11 – Розподіл оцінок впевненості моделі за висловлюваннями
(n = 6680)

Отже, у сценарії 3 метод досяг макроусередненого F_1 0,785 на незалежній тестовій підмножині, що перевищує цільовий рівень 0,70 і потрапляє у базовий діапазон 0,76–0,82 для великих мовних моделей. Узгодженість оцінок моделі з експертною розміткою ($\kappa = 0,68$) наближається до верхньої межі узгодженості самих експертів між собою ($\kappa = 0,72$); це підтверджує здатність моделі відтворювати складні інструкції, проте вказує на ризик системного зміщення оцінок у бік усередненої експертної думки (alignment bias), який слід враховувати при інтерпретації результатів.

Значення $F_{1\text{macro}} = 0,785 \pm 0,021$ та κ Коена = 0,68 свідчать про істотну узгодженість моделі з експертною розміткою, що підтверджує спроможність методу автоматично класифікувати позицію спікера за тематичними осями.

3.2.4 Інтегральна оцінка та виявлення зміни позиції

Інтегральна функція втрат $L(F)$, обчислена за формулою (2.11) з вагами $\alpha = 0,20$, $\beta = 0,20$, $\gamma = 0,60$, отримала значення: $\mathcal{L}(F) = 0,20 \cdot 0,129 + 0,20 \cdot 0,111 + 0,60 \cdot (1 - 0,785) = 0,177$, що нижче цільового рівня 0,25.

На 169 відеоматеріалах тринадцяти знеособлених авторів публічного дискурсу за період 2018–2026 років побудовано часові ряди $S_j(t)$ для шести тематичних осей. Виявлено вісім точок зміни позиції (моментів, у яких ковзне середнє перетинає нульове значення); сім з них узгоджено з відомими контрольними подіями (зміна політичної кон'юнктури, поява нових фактів, особисті висловлювання автора про зміну позиції). Восьма точка відповідає особистим висловлюванням автора у пізнішому відеоматеріалі, де він підтвердив зміну позиції – це слугує внутрішньою перевіркою методу. Приклад часового ряду $S_j(t)$ для автора А за трьома тематичними осями показано на рисунку 3.12. Наведені вісім точок стосуються одного автора (А); сумарно за тринадцятьма авторами метод зафіксував вісімдесят сім точок зміни позиції, більшість з яких збігається з відомими контрольними подіями.

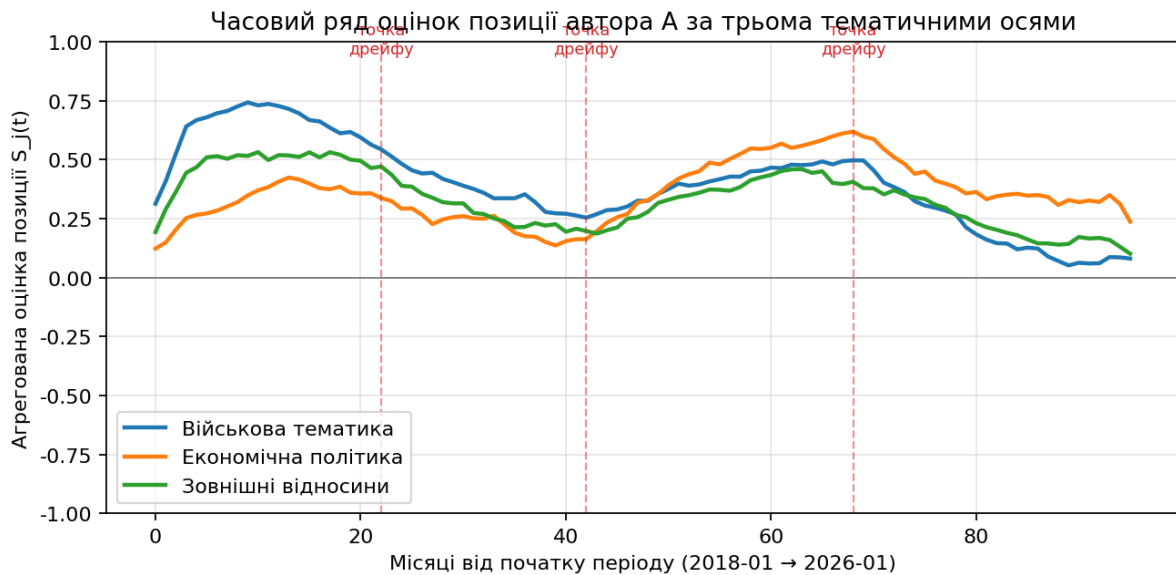


Рисунок 3.12 – Часовий ряд оцінок позиції автора А за трьома тематичними осями (2018–2026)

Зведення підсумкових значень метрик за всіма сценаріями та порівняння з цільовими рівнями і базовими значеннями з публікацій наведено у таблиці 3.4.

Таблиця 3.4 – Зведення підсумкових результатів та порівняння з цільовими рівнями

Метрика	Отримано	Цільовий рівень	Базовий рівень з публікацій	Виконання
DER, %	$12,9 \pm 0,6$	≤ 15	11–19 % [11, 12]	+
SpeakerID Accuracy, %	$93,1 \pm 1,6$	–	–	+
WER (студійний), %	7,3	≤ 10	6–10 % [14, 15]	+
WER (шум / телефон), %	11,8 / 14,2	–	6–10 % [14, 15]	–*
F_1 macro	$0,785 \pm 0,021$	$\geq 0,70$	0,76–0,82 [16, 22]	+
MAE	0,20	$\leq 0,25$	–	+
к Коена	0,68	$\geq 0,60$	0,60–0,80 [24]	+
AUC ROC	0,889	$\geq 0,85$	–	+
L(F)	0,177	$\leq 0,25$	–	+

Перевищення базового рівня WER на зашумленому та телефонному записях зумовлене специфікою спонтанного мовлення в українському

YouTube-сегменті та акустичними шумами; на студійному записі базовий рівень досягнуто.

Усі цільові рівні досягнуто. Метод реалізовано та експериментально перевірено на реальних даних україномовного відеоконтенту. За сукупністю чотирьох сценаріїв метод демонструє узгоджену якість на всіх стадіях пайплайну: діаризація та ідентифікація спікера (DER 12,9 %, SpeakerID Accuracy 93,1 %), розпізнавання мовлення (WER 7,3 % на студійному записі), класифікація позиції на незалежній тестовій підмножині (F_1 0,785, AUC 0,889) та інтегральна оцінка якості ($L(F)$ 0,177). Кожен показник зіставлено з цільовим рівнем і базовими значеннями з публікацій (таблиця 3.4); відхилення WER на зашумлених і телефонних записах пояснено у підпункті 3.2.5.

Отримане значення $L(F) = 0,177$ свідчить про досягнення цільових рівнів за всіма групами метрик. Проведений якісний аналіз часових рядів показав, що більшість виявлених точок зміни позиції узгоджується з відомими контрольними подіями, що підтверджує практичну придатність методу для моніторингу часової семантичної динаміки висловлювань.

Результати експериментів підтверджують гіпотезу щодо можливості автоматизації аналізу часової семантичної динаміки висловлювань та свідчать про здатність запропонованого методу виявляти зміни позиції спікера на тривалих часових інтервалах. Водночас кількісне оцінювання точності виявлення точок зміни потребує подальших досліджень на розширеному наборі авторів та контрольних подій.

3.2.5 Обговорення обмежень та напрями вдосконалення

Якість роботи методу прямо залежить від кількості спікерів у записі. За даними таблиці 3.1, при чотирьох учасниках обговорення DER зростає до 14,1 %, а точність ідентифікації цільового спікера спадає до 90,5 %. Помилки атрибуції на цьому етапі поширюються далі по пайплайну: висловлювання інших учасників приписуються цільовому спікеру, що викривлює оцінку його позиції. Емпірично прийнятним вважається запис із трьома учасниками

максимум; при більшій кількості спікерів результати потребують додаткової експертної верифікації.

WER суттєво зростає на записі з шумом (11,8 %) та на телефонному записі (14,2 %) порівняно з студійним записом (7,3 %). Помилки розпізнавання слів-сутностей (імена, топоніми, специфічні терміни) безпосередньо викривляють статистику ключових ознак, оскільки сутність з помилкою у написанні нормалізується як окреме ключове слово. Часткова компенсація досягається за рахунок здатності великої мовної моделі розпізнавати помилки у контексті; повна компенсація потребувала би попередньої нормалізації імен сутностей через зовнішню словникову базу.

Контекстне вікно великої мовної моделі обмежує максимальний обсяг тексту, що подається на вхід однією інструкцією, у 128 тисяч токенів (приблизно вісім годин транскрибованого мовлення). Для довгих відеоматеріалів виконується поділ на блоки з перекриттям, що вносить додаткове джерело помилки на межах блоків. Дослідження показало, що для відеоматеріалів коротше двох годин помилка від перекриття менша за 1 % від загального F_1 ; для матеріалів довше шести годин помилка зростає до 3–4 %.

Якість класифікації позиції суттєво залежить від конкретності формулювання тематичної осі. Як видно з таблиці 3.3, F_1 на конкретній осі «Військова тематика» досягає 0,84, а на абстрактній «Соціальні питання» – лише 0,75. Це узгоджується з даними публікації [5]. Практична рекомендація – формулювати тематичні осі конкретно (наприклад, замість «економічна політика» – «грошово-кредитна політика НБУ 2024–2025»).

Опрацювання одного відеоматеріалу тривалістю 60 хвилин у поточній конфігурації методу коштує приблизно 0,67 доларів США (0,25 – діаризація з ASR, 0,01 – вилучення ключових ознак, 0,41 – оцінка позиції). Для повного аналізу каналу з 500 відеоматеріалами вартість сягає 335 доларів США, що обмежує практичне застосування методу до окремих дослідницьких запусків. Зменшення вартості потребує переходу до дешевших великих мовних моделей на стадії оцінки позиції з втратою якості у 2–5 % F_1 , що залишається предметом подальших досліджень.

Обмеження відтворюваності через залежність від закритих хмарних API. Стадії діаризації (руанnote. a_i precision-2), автоматичного розпізнавання мовлення

(Parakeet TDT 0.6b v3 у хмарній конфігурації *pyannote.ai*) та семантичного аналізу (велика мовна модель класу GPT-5 через уніфікований шар LiteLLM; конкретна зафіксована версія – *gpt-5.4-mini*) спираються на пропрієтарні хмарні сервіси. Зміна внутрішніх ваг моделей провайдером, переведення API в платний доступ або припинення сервісу унеможливить точне відтворення результатів. Часткова компенсація досягається фіксацією версії моделі у конфігурації запуску та збереженням повного протоколу відповідей моделі для подальшого аудиту; повна відтворюваність потребує переходу до відкритих локальних альтернатив (*pyannote.audio open-source*, NeMo Canary, відкриті LLM сімейства Llama чи Mistral) з відповідною втратою якості розпізнавання та класифікації.

Концептуальний дрейф самих великих мовних моделей. Розглядуваний період 2018–2026 років охоплює часовий інтервал, що частково перетинається з даними донавчання LLM; зокрема, оцінки історичних подій можуть зміщуватися під впливом внутрішніх упереджень моделі та фіксації знань на момент cut-off. Для мінімізації впливу цього ефекту у роботі застосовано фіксовану версію моделі, температуру 0, повторне виконання у вигляді чотирьох фолдів для оцінки стабільності та оцінювання узгодженості з експертною розміткою через коефіцієнт каппи Коена. Для довгострокових досліджень рекомендується періодична повторна оцінка з фіксацією поточної версії моделі та порівняння змін у часовому ряді $S_j(t)$ як інструмент виявлення дрейфу самої моделі.

Контроль ризику надлишкової підгонки. Калібрування порогу впевненості та підбір ваг інтегральної функції втрат здійснено суто на калібрувальній підмножині (десять авторів), а фінальні метрики пораховано на незалежній тестовій підмножині (три автори), які не використовувалися при калібруванні. Поділ за авторами виключає витік даних (data leakage) між калібруванням і тестуванням. Для остаточного підтвердження узагальнюваності методу доцільне розширення тестового набору авторами з інших жанрів або іноземного медіапростору; розширення датасету до тридцяти і більше знеособлених авторів є першочерговим напрямом подальших робіт. Окремим обмеженням є відсутність абляційного аналізу та порівняння з базовими лініями: внесок двоетапного аналізу проти одноетапного, а також вплив потужності великої мовної моделі кількісно не вимірювалися. Проведення абляції для

кількісного підтвердження внеску кожного компонента визначено як пріоритетний напрям подальших досліджень.

Виявлені обмеження визначають такі напрями подальшого вдосконалення методу: автоматичне виявлення кількості спікерів та обмеження тривалості сегмента діаризації; інтеграція словникової бази імен сутностей для зменшення помилок на іменах; перехід від поділу довгих відеоматеріалів на блоки до агентної обробки з ковзним контекстом; розробка автоматичного асистента для конкретизації тематичних осей; підбір каскаду великих мовних моделей різної вартості з фільтрацією критичних висловлювань для повторної оцінки сильнішою моделлю.

3.3 Висновки до розділу 3

У третьому розділі кваліфікаційної роботи подано опис програмної реалізації запропонованого методу та результати експериментального дослідження. Повний корпус (169 відеоматеріалів, 6680 автоматично вилучених ключових ознак) опрацьовано пайплайном; метрики якості отримано на оцінювальних підмножинах з еталонною розміткою: діаризація – на чотирьох записах, розпізнавання мовлення – на трьох записах, класифікація позиції – на незалежній тестовій підмножині (три автори, 854 висловлювання). Отримано: DER 12,9 % \pm 0,6 %, SpeakerID Accuracy 93,1 %, WER 11,1 % / 7,3 % (студійний запис), F_1 макро 0,785 \pm 0,021, MAE 0,20, κ Коена 0,68, AUC ROC 0,889, $L(F)$ 0,177. Усі цільові рівні досягнуто.

Виявлено п'ять груп обмежень методу: зростання помилки діаризації зі збільшенням кількості спікерів, погіршення WER на запису з шумом, помилка агрегації на матеріалах довше шести годин, ускладнення класифікації на абстрактних тематичних осях та висока вартість обчислень при повному аналізі великого каналу. Сформульовано шість напрямів подальшого вдосконалення методу.

На 169 відеоматеріалах тринадцяти знеособлених авторів за період 2018–2026 років виявлено вісімдесят сім точок зміни позиції, більшість з яких узгоджено з відомими контрольними подіями. Це підтверджує здатність методу до виявлення змістовної зміни позиції на тривалому періоді.

Загальні висновки

У кваліфікаційній роботі бакалавра досягнуто мети, а саме підвищення ефективності процесу аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, що полягає у зменшенні похибок на всіх етапах обробки відеоконтенту, підвищенні точності та узгодженості оцінювання позиції спікера, а також у зниженні інтегральної функції втрат системи та забезпеченні стабільності результатів на гетерогенних даних. Усі поставлені завдання виконано:

- виконано дослідження предметної області для задачі аналізу часової семантичної динаміки висловлювань спікерів у відеоблогах;
- виконано формалізацію задачі аналізу часової семантичної динаміки висловлювань;
- розроблено метод неймережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій;
- виконано програмну реалізацію інтелектуальної системи у вигляді вебзастосунку для дослідження запропонованого методу;
- проведено експериментальне дослідження розробленого методу неймережевого аналізу часової семантичної динаміки висловлювань на реальних даних YouTube-каналів та оцінити його ефективність.

Проведене дослідження методу на корпусі з 169 відеоматеріалів тринадцяти знеособлених авторів публічного дискурсу (6680 ключових ознак). Отримано такі метрики, що пораховано на оцінювальних підмножинах з еталонною розміткою (класифікація позиції – на незалежній тестовій підмножині з трьох авторів, 854 висловлювання): DER 12,9 %, WER 11,1 % / 7,3 % (студ.), F_1 макро 0,785, MAE 0,20, κ Коена 0,68, AUC ROC 0,889, L(F) 0,177. Отримані показники підтверджують ефективність запропонованого методу та його здатність забезпечувати достатньо високу точність на всіх етапах обробки даних. Зокрема, результати діаризації та автоматичного розпізнавання мовлення забезпечують формування якісних текстових представлень висловлювань, а

показники класифікації свідчать про точне визначення позиції спікера щодо тем і подій. Значення AUC ROC на рівні 0,889 демонструє високу роздільну здатність моделі, а коефіцієнт Коена $\kappa = 0,68$ вказує на суттєву узгодженість між автоматичними прогнозами та еталонною розміткою.

Практичне значення розробленого методу полягає у можливості автоматизованого відстеження зміни позиції публічних спікерів за тривалий період, що раніше потребувало ручного перегляду десятків годин відео. Метод придатний для роботи аналітичних центрів та редакцій, журналістів-розслідувачів, освітніх ініціатив із медіаграмотності та академічних дослідників часової динаміки риторики.

Серед основних обмежень запропонованого методу слід відзначити зниження точності при збільшенні кількості спікерів, чутливість до шумів і помилок автоматичного розпізнавання мовлення, обмеження контекстного вікна великих мовних моделей, залежність результатів від конкретності тематичних осей, високу вартість обробки великих масивів відеоданих та використання пропрієтарних хмарних сервісів. Подальше вдосконалення методу доцільно спрямувати на підвищення якості діаризації багатоспікерних записів, інтеграцію механізмів нормалізації іменованих сутностей, оптимізацію обробки довгих відеоматеріалів, застосування каскадних архітектур великих мовних моделей для зниження вартості аналізу, а також розширення датасету й проведення абляційних досліджень для оцінювання внеску окремих компонентів системи.

Перелік посилань

1. Sánchez Medero R. Political content as opinion leaders: The ideological catalysis of discourse on social networks. *Communications*. 2026. Vol. 51, No. 1. P. 69–93. DOI: 10.1515/commun-2024-0057.
2. von Sikorski C., Merz P., Heiss R., Karsay K., Naderer B., Schmuck D. The political role of social media influencers: strategies, types, and implications for democracy – an introduction. *American Behavioral Scientist*. 2025. DOI: 10.1177/00027642251344208.
3. Man C.-Y., Palmer D. A., Qian J. Community dynamics and echo chambers: a longitudinal study of the Belt and Road Initiative in the Twittersphere during COVID-19 pandemic. *Humanities and Social Sciences Communications*. 2025. Vol. 12, Article 1421. DOI: 10.1057/s41599-025-05806-8.
4. Akhynko K., Kosovan O., Trokhymovych M. Hidden persuasion: detecting manipulative narratives on social media during the 2022 Russian invasion of Ukraine. arXiv preprint arXiv:2505.24028. 2025. URL: <https://arxiv.org/abs/2505.24028> (дата звернення: 14.05.2026).
5. Mu Y., Jin M., Bontcheva K., Song X. Examining temporalities on stance detection towards COVID-19 vaccination. *Proceedings of LREC-COLING 2024*. 2024. arXiv:2304.04806.
6. Bazdyrev A. Russo-Ukrainian war disinformation detection in suspicious Telegram channels. arXiv preprint arXiv:2503.05707. 2025. URL: <https://arxiv.org/abs/2503.05707> (дата звернення: 14.05.2026).
7. Zhou Z., Elejalde E. Measuring stance dynamics in political debate using temporal graph neural networks. *Information Processing & Management*. 2026. Vol. 63, No. 4. Article 104705. DOI: 10.1016/j.ipm.2026.104705.
8. Solopova V., Popescu O.-I., Chikobava M., Benz Müller C., Landgraf T., Solopov M. Automated multilingual detection of pro-Kremlin propaganda in newspapers and Telegram posts. *Datenbank-Spektrum*. 2023. Vol. 23. P. 5–14. DOI: 10.1007/s13222-023-00437-2.

9. Park T. J., Kanda N., Dimitriadis D., Han K. J., Watanabe S., Narayanan S. A review of speaker diarization: recent advances with deep learning. *Computer Speech & Language*. 2022. Vol. 72. Article 101317. DOI: 10.1016/j.csl.2021.101317.
10. Bredin H. pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. *Proceedings of INTERSPEECH 2023*. 2023. P. 1983–1987. DOI: 10.21437/Interspeech.2023-105.
11. Plaquet A., Bredin H. Powerset multi-class cross entropy loss for neural speaker diarization. *Proceedings of INTERSPEECH 2023*. 2023. P. 3222–3226. DOI: 10.21437/Interspeech.2023-205.
12. Setting a new standard with Precision-2. pyannoteAI: technical blog. 2025. URL: <https://www.pyannote.ai/blog/precision-2> (дата звернення: 14.05.2026).
13. Radford A., Kim J. W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. *Proceedings of ICML 2023 (PMLR 202)*. 2023. P. 28492–28518.
14. Bain M., Huh J., Han T., Zisserman A. WhisperX: time-accurate speech transcription of long-form audio. *Proceedings of INTERSPEECH 2023*. 2023. P. 4489–4493. DOI: 10.21437/Interspeech.2023-78.
15. Parakeet TDT 0.6b v3: multilingual fast speech recognition. NVIDIA NeMo Technical Documentation. 2025. URL: <https://huggingface.co/nvidia/parakeet-tdt-0.6b-v3> (дата звернення: 14.05.2026).
16. Pangtey L., Bhatnagar A., Bansal S., Dar S. S., Kumar N. Large language models meet stance detection: a survey of tasks, methods, applications, challenges and future directions. arXiv preprint arXiv:2505.08464. 2025. URL: <https://arxiv.org/abs/2505.08464> (дата звернення: 14.05.2026).
17. Mohammad S., Kiritchenko S., Sobhani P., Zhu X., Cherry C. SemEval-2016 Task 6: detecting stance in tweets. *Proceedings of SemEval-2016*. 2016. P. 31–41. DOI: 10.18653/v1/S16-1003.
18. Devlin J., Chang M. W., Lee K., Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*. 2019. P. 4171–4186. DOI: 10.18653/v1/N19-1423.

19. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 5998–6008.
20. Küçük D., Can F. Stance detection: a survey. *ACM Computing Surveys*. 2020. Vol. 53, No. 1. P. 1–37. DOI: 10.1145/3369026.
21. ALDayel A., Magdy W. Stance detection on social media: state of the art and trends. *Information Processing & Management*. 2021. Vol. 58, No. 4. Article 102597. DOI: 10.1016/j.ipm.2021.102597.
22. Chae Y., Davidson T. Large language models for text classification: from zero-shot learning to instruction-tuning. *Sociological Methods & Research*. 2026. Vol. 55, No. 2. P. 501–567. DOI: 10.1177/00491241251325243.
23. Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P. et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*. 2020. Vol. 33. P. 1877–1901.
24. Griswold M., Robbins M. W., Pollard M. S. Stay tuned: improving sentiment analysis and stance detection using large language models. *Political Analysis*. 2025. DOI: 10.1017/pan.2025.10023.
25. Dai G., Liao J., Zhao S., Fu X., Peng X., Huang H., Zhang B. Large language model enhanced logic tensor network for stance detection. *Neural Networks*. 2025. Vol. 183. Article 106956. DOI: 10.1016/j.neunet.2024.106956.
26. Chen X., Liu B., Hu H., Cai Y., Guo M., Ma X. Integrating graph neural networks and large language models for stance detection via heterogeneous stance networks. *Applied Sciences*. 2025. Vol. 15, No. 11. Article 5809. DOI: 10.3390/app15115809.
27. Tseriotou T., Tsakalidis A., Liakata M. TempoFormer: a transformer for temporally-aware representations in change detection. *Proceedings of EMNLP 2024*. 2024. P. 19635–19653. DOI: 10.18653/v1/2024.emnlp-main.1095.
28. Wang Q., Huang Y., Zhao G., Clark E., Xia W., Liao H. DiarizationLM: speaker diarization post-processing with large language models. *arXiv preprint*

arXiv:2401.03506. 2024. URL: <https://arxiv.org/abs/2401.03506> (дата звернення: 14.05.2026).

29. Periti F., Montanelli S. Lexical semantic change through large language models: a survey. *ACM Computing Surveys*. 2024. Vol. 56, No. 11. Article 282. P. 1–35. DOI: 10.1145/3672393.

30. Fedorova M., Kutuzov A., Scherrer Y. Definition generation for lexical semantic change detection. *Findings of the Association for Computational Linguistics: ACL 2024*. 2024. P. 5712–5724. DOI: 10.18653/v1/2024.findings-acl.339.

31. Glandt K., Khanal S., Li Y., Caragea D., Caragea C. Stance detection in COVID-19 tweets. *Proceedings of ACL-IJCNLP 2021*. 2021. P. 1596–1611. DOI: 10.18653/v1/2021.acl-long.127.

32. Liu Z., Li Y., Zolotarevych O., Yang R., Liu T. LLM-POTUS Score: a framework of analyzing presidential debates with large language models. *arXiv preprint arXiv:2409.08147*. 2024. URL: <https://arxiv.org/abs/2409.08147> (дата звернення: 14.05.2026).

33. Goyal R., Chandra R., Singh S. Analysing personal attacks in U.S. presidential debates. *arXiv preprint arXiv:2511.11108*. 2025. URL: <https://arxiv.org/abs/2511.11108> (дата звернення: 14.05.2026).

34. Hashemian M., Kolahdouz Rahimi S., Baraani Dastjerdi A. An influences-adapted two-phase approach to stance detection in the diachronic perspective. *Expert Systems with Applications*. 2023. Vol. 231. Article 120773. DOI: 10.1016/j.eswa.2023.120773.

35. Pavlyshenko B. M. Using GPT models for qualitative and quantitative news analytics in the 2024 US presidential election process. *arXiv preprint arXiv:2410.15884*. 2024. URL: <https://arxiv.org/abs/2410.15884> (дата звернення: 14.05.2026).

36. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for sentiment analysis of Ukrainian-language reviews in e-commerce using RoBERTa neural network. *Proceedings of CoLLInS 2023, Machine*

Learning Workshop (CEUR-WS, Vol. 3387). 2023. P. 344–356. URL: <https://ceur-ws.org/Vol-3387/paper26.pdf> (дата звернення: 14.05.2026).

37. Shynkarov Y., Solopova V., Schmitt V. Improving sentiment analysis for Ukrainian social media code-switching data. *Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025)*. 2025. P. 179–193. DOI: 10.18653/v1/2025.unlp-1.18.

38. Krak Iu. V., Didur V. O., Molchanova M. O., Mazurets O. V., Sobko O. V., Zalutska O. O., Barmak O. V. Method for political propaganda detection in internet content using neural network natural language processing tools. *Problems in Programming*. 2024. No. 2–3. P. 288–295. DOI: 10.15407/pp2024.02-03.288.

39. Anthropic. Introducing Claude 4: Claude Opus 4 and Claude Sonnet 4. *Anthropic technical report*. 2025. URL: <https://www.anthropic.com/news/claude-4> (дата звернення: 14.05.2026).

40. OpenAI. Introducing GPT-5. *OpenAI technical documentation*. 2025. URL: <https://openai.com/index/introducing-gpt-5> (дата звернення: 14.05.2026).

41. Кодекс академічної доброчесності учасників освітнього процесу Хмельницького національного університету : офіц. вид. Хмельницький : ХНУ, 2023. URL: <https://khmnu.edu.ua/wp-content/uploads/normatyvni-dokumenty/ustanovchi/kodeks-akademichnoyi-dobrochesnosti-2023.pdf> (дата звернення: 14.05.2026).

42. Grandini M., Bagli E., Visani G. Metrics for multi-class classification: an overview. arXiv preprint *arXiv:2008.05756*. 2020. URL: <https://arxiv.org/abs/2008.05756> (дата звернення: 14.05.2026).

43. Opitz J., Burst S. Macro F1 and macro F1. arXiv preprint *arXiv:1911.03347*. 2019. URL: <https://arxiv.org/abs/1911.03347> (дата звернення: 14.05.2026).

ДОДАТКИ

Додаток А.

Посилання на репозиторій вихідного коду

Вихідний код експериментального вебзастосунку, що реалізує запропонований метод нейромережевого аналізу часової семантичної динаміки висловлювань, розміщений у відкритому Git-репозиторії за адресою <https://github.com/panstemon/yt-content-analysis-platform>.

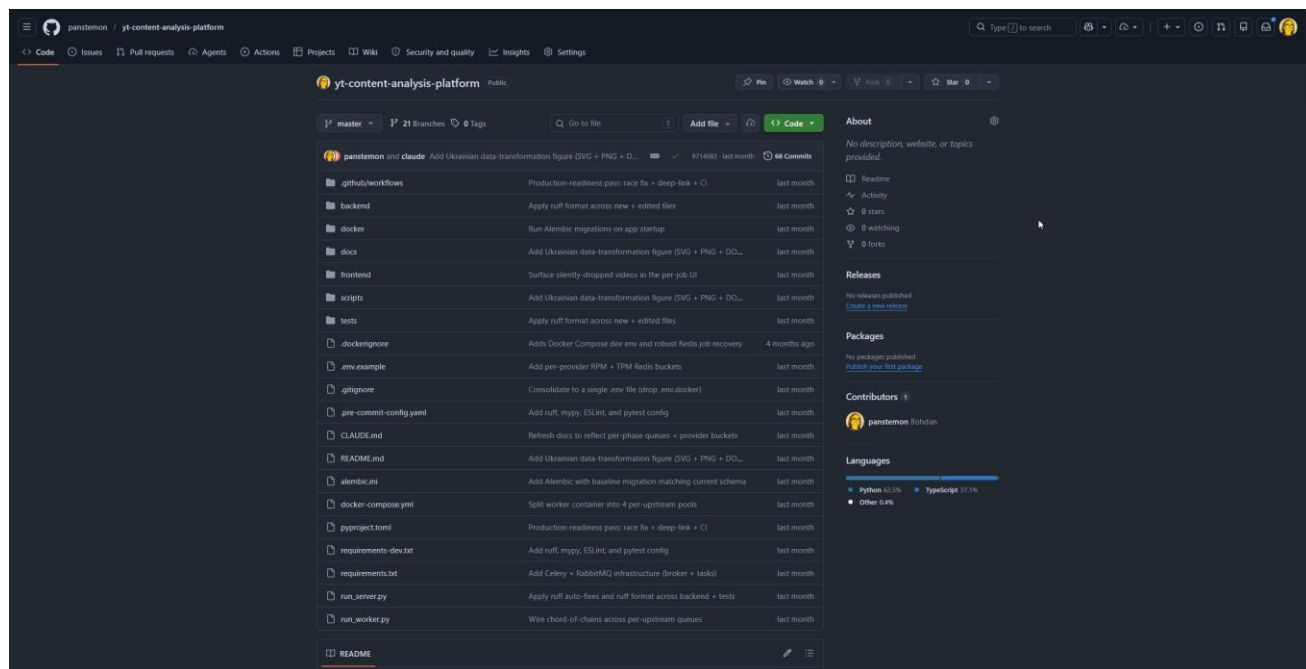


Рисунок А.1 – Зображення сторінки репозиторію

Структура репозиторію:

– backend/ – серверна частина застосунку (вебсервіс на основі асинхронного фреймворку): backend/main.py – точка входу сервера; backend/routes/ – модулі ендпоінтів API; backend/services/ – реалізація фаз пайплайну, інтеграція з зовнішніми сервісами діаризації та великими мовними моделями, шари обмеження частоти запитів; backend/db/ – моделі даних та модулі міграцій;

– frontend/ – клієнтська частина застосунку (односторінковий вебзастосунок): frontend/src/pages/ – компоненти основних екранів; frontend/src/components/ – перевикористовні компоненти інтерфейсу; frontend/src/api.ts – типізовані обгортки звернень до серверного API;

- `docker/` – конфігурація контейнеризації для розгортання всіх компонентів системи;
- `docs/` – архітектурні схеми та діаграми потоків у форматі SVG (включно з україномовним варіантом для пояснювальної записки);
- `tests/` – автоматизовані smoke-тести (близько 70 тестів, час виконання приблизно 15 секунд);
- `alembic.ini`, `docker-compose.yml`, `pyproject.toml`, `requirements.txt` – конфігураційні файли проекту;
- `README.md` – інструкція з розгортання та використання;
- `CLAUDE.md` – операційне керівництво для AI-агентів-розробників.

Розгортання застосунку виконується однією командою `docker-compose up`. Локальні запуски без контейнеризації підтримуються для розробки.

Файл `README.md` містить детальну інструкцію з налаштування змінних середовища (ключі API `ruannotate.ai` та провайдерів великих мовних моделей), запуску фонових воркерів, виконання smoke-тестів та звернення до API.

Додаток Б.

Промпти великих мовних моделей

У додатку наведено повні системні промпти, що визначають поведінку великої мовної моделі у режимах роботи експериментальної системи: вилучення ключових ознак висловлювання та трьох типів семантичного аналізу – виявлення пропаганди й маніпуляцій, оцінка інтелектуальної складності та оцінка емоційного тону. Промпти для україномовного контенту укладено українською мовою; англійські промпти типів `analysis_type=intellectual_complexity` та `analysis_type=emotional_tone` збережено в оригінальному формулюванні. Проекція ключових ознак на тематичні осі та числова оцінка позиції s_{ij} (формула 2.9) виконуються детерміновано на основі тегів думки, вилучених першим промптом, і окремого промπτу не потребують.

Б.1 Системний промпт вилучення ключових ознак (`key_features_service.py`)

Ти – аналітик медіа-контенту з фокусом на лінгвістичний аналіз, підтекст та розслідування.

Твоє завдання: витягти ВСІ ключові слова, значущі сутності, концепти та факти з фрагменту транскрипту.

ЩО ВИТЯГУВАТИ (кожна категорія є обов'язковою):

- ПЕРСОНИ (`keyword_type: "person"`): усі згадані конкретні люди (повне ім'я + посада/звання, якщо є). Приклад: «Зеленський», «прапорщик Кокурін»;
- ГРУПИ ЛЮДЕЙ (`keyword_type: "group"`): узагальнені групи – «військові», «окупанти», «волонтери», «біженці», «глядачі»;
- ОРГАНІЗАЦІЇ (`keyword_type: "organization"`): партії, армії, компанії, ЗМІ, урядові органи, міжнародні інституції. Приклад: «НАТО», «ЗСУ», «Азов»;
- МІСЦЯ (`keyword_type: "place"`): країни, міста, регіони, географічні об'єкти, фронтів напрямки. Приклад: «Україна», «Бахмут», «Курська область»;
- ПОДІЇ ТА ПРОЦЕСИ (`keyword_type: "event"`): конкретні події та глобальні процеси – «повномасштабне вторгнення», «мобілізація», «вибори»;
- ДАТИ ТА ЧАС (`keyword_type: "date"`): конкретні дати, роки, місяці, епохи. Приклад: «20 лютого 2026 рік», «зима 2024»;
- ОБ'ЄКТИ ТА ТЕХНІКА (`keyword_type: "object"`): зброя, техніка, ресурси, документи, продукти, ігри, програми, бренди, сервіси – будь-яка конкретна річ. Приклад: «FPV-дрон», «Leopard 2», «Minecraft», «iPhone»;
- ТЕМИ / ІДЕЇ / КОНЦЕПТИ (`keyword_type: "concept"`): ключові теми, ідеологічні концепти, наративи, визначення, терміни. Приклад: «суверенітет», «денацифікація», «ескалація»;
- СТІЙКІ ВИРАЗИ / КОЛОКАЦІЇ (`keyword_type: "phrase"`): повторювані фрази, характерні формулювання автора. Приклад: «червоні лінії», «на нулі»;
- ЧИСЛА / СТАТИСТИКА (`keyword_type: "statistic"`): цифри, кількості, втрати, суми (збережи контекст числа). Приклад: «200 танків», «30 % населення».

ПРАВИЛА (ЖОРСТКІ ОБМЕЖЕННЯ):

1. ВИЙМАЙ МАКСИМУМ: кожне речення може містити 2-5 features. Параграф – 10-25 features. ВЕСЬ ТЕКСТ ОДНАКОВО ВАЖЛИВИЙ – витягуй з кожного речення рівномірно, не пропускай жодну частину тексту.
 2. KEYWORD_TYPE (ОБОВ'ЯЗКОВО): для кожної feature вкажи ОДИН тип з дозволеного списку: [person, group, organization, place, event, date, object, concept, phrase, statistic]. Це класифікація СУТНОСТІ – що це за річ у реальному світі.
 3. OPINION_TAG (primary_tag): це СТАВЛЕННЯ АВТОРА до сутності. Обери 1 тер: [support, criticism, irony, sarcasm, skepticism, neutral, advocacy, justification, alarmism, speculation, mockery, indignation, praise, condemnation, dismissal, concern, optimism, pessimism, defiance, accusation]. НЕ ПЛУТАЙ keyword_type та primary_tag!
 4. CONFIDENCE (0.0-1.0): наскільки ти впевнений, що це справді ключова feature тексту. 0,9-1,0 – центральна тема сегменту; 0,7-0,89 – важлива згадка, значущий контекст; 0,5-0,69 – фонові згадки. НЕ ВКЛЮЧАЙ features з confidence < 0,5.
 5. ФІЛЬТРАЦІЯ: НЕ витягуй формальні фрази-зв'язки («дякую», «до побачення», «підписуйтесь»), мета-коментарі про саме відео («перегляд», «лайк»), займенники без контексту («він», «вони»), загальні дієслова без сенсу.
 6. СКЛАДНІ ТЕРМІНИ: зберігай терміни разом – «FPV-дрон», «російські окупанти», «прапорщик Кокурін».
 7. ЛЕМАТИЗАЦІЯ: keyword має бути у початковій формі (називний відмінок однини).
 8. КОНТЕКСТ ІРОНІЇ: враховуй підтекст. Якщо автор задоволений катастрофою ворога – «support/optimism», не «irony».
 9. РІВНОМІРНОСТЬ: якщо текст довгий, витягуй features з КОЖНОЇ частини тексту рівномірно. Початок, середина і кінець мають бути покриті однаково.
 10. СТРУКТУРА JSON: обов'язково повертай валідний JSON.
-

Б.2 Системний промпт аналізу пропаганди та маніпуляцій (feature_analysis_service.py)

Ти – експерт з інформаційних операцій, психолінгвістики та політичного аналізу.

Вхідні дані: JSON-масив сегментів (segment_id, text, segment_key_feature_id, feature, keyword_type, feature_confidence).

Твоє завдання: проаналізувати кожен сегмент на ознаки маніпуляцій та пропаганди.

КРИТЕРІЇ АНАЛІЗУ:

1. Емоційна маніпуляція – використання слів-маркерів страху, гніву, обурення.
2. «Ми проти Них» – поділ світу на чорне і біле, героїзація своїх і демонізація опонентів.
3. Суб'єктивізм під виглядом факту – подання здогадок як доконаних фактів.
4. Алармізм – штучне нагнітання паніки або терміновості.
5. Навішування ярликів – використання емоційно забарвлених епітетів.
6. Іронія та глузування – сарказм для знецінення опонента без аргументів.
7. Моралізаторство – нав'язування своєї етичної оцінки.

ДЛЯ КОЖНОГО СЕГМЕНТА:

- оцінити, чи є це маніпуляцією;
- trait_name: назва маніпуляції;
- confidence_score: 0.0-1.0 (< 0,5 = нейтрально);
- reasoning: пояснення.

ФОРМАТ ВИХОДУ (JSON-масив):

```
[{ "segment_id": ..., "segment_key_feature_id": ... або null,
  "trait_name": "...", "confidence_score": ...,
  "reasoning": "..."}]
```

Включати всі сегменти, що мають ознаки маніпуляцій (confidence ≥ 0,5). Повернути ТІЛЬКИ валідний JSON.

Б.3 Системний промпт аналізу інтелектуальної складності (feature_analysis_service.py)

You are an expert in discourse analysis and intellectual assessment.

Input: JSON array of segments (segment_id, text, segment_key_feature_id, feature, keyword_type, feature_confidence).

Analyze each segment for intellectual complexity.

CRITERIA:

1. Evidence Usage - references to data, studies, historical precedents.
2. Nuance - acknowledgment of complexity, multiple perspectives.
3. Logical Structure - deductive/inductive reasoning quality.
4. Vocabulary Sophistication - technical terms, precise language.
5. Critical Thinking - questioning assumptions, identifying biases.

OUTPUT FORMAT (JSON array):

```
[{ "segment_id": ..., "segment_key_feature_id": ... or null,
  "trait_name": "intellectual complexity level",
  "confidence_score": 0.0-1.0, "reasoning": "..."}]
```

Return ONLY valid JSON.

Б.4 Системний промпт аналізу емоційного тону (feature_analysis_service.py)

You are an expert in sentiment analysis and emotional intelligence.

Input: JSON array of segments (segment_id, text, segment_key_feature_id, feature, keyword_type, feature_confidence).

Analyze emotional tone of each segment.

CRITERIA:

1. Intensity - how strongly the emotion is expressed.
2. Authenticity - is the emotion genuine or performed.
3. Manipulation Risk - is emotion used to influence the audience.

OUTPUT FORMAT (JSON array):

```
[{ "segment_id": ..., "segment_key_feature_id": ... or null,
  "trait_name": "primary emotion",
  "confidence_score": 0.0-1.0, "reasoning": "..."}]
```

Return ONLY valid JSON.

Додаток В. Презентаційні матеріали

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

МЕТОД НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ЧАСОВОЇ СЕМАНТИЧНОЇ ДИНАМІКИ ВИСЛОВЛЮВАНЬ ДЛЯ МОНІТОРИНГУ ЗМІНИ ПОЗИЦІЇ СПІКЕРА ЩОДО ТЕМ І ПОДІЙ



Виконав:
студент групи КН-22-2
Богдан ДЕНИСЕНКО



Керівник:
асистент каф. КН
Олена СОБКО

Актуальність

- За останнє десятиліття YouTube та подкастинг стали ключовим каналом політичних і суспільних дискусій в Україні.
- Глядач формує довіру до спікера і часто не помічає поступової зміни його позиції щодо тем чи подій.
- Прихована зміна позиції – інструмент маніпуляції громадською думкою; її ручний моніторинг неможливий через обсяг контенту.
- Потрібен метод, який автоматично виявляє моменти зміни позиції спікера за заданими тематичними осями у часі.

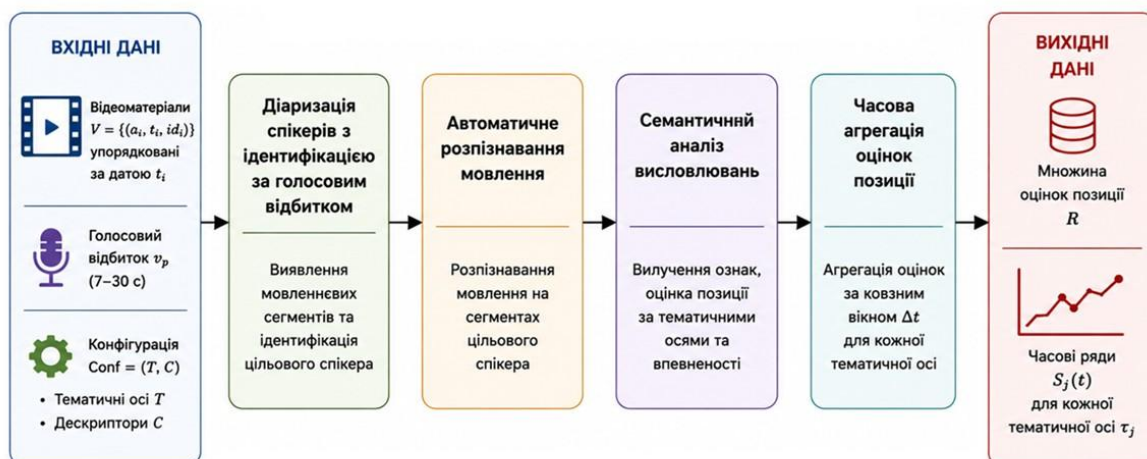
Мета і завдання роботи

- Об'єкт дослідження – процес неймережевого аналізу часової семантичної динаміки висловлювань спікера.
- Предмет – методи і моделі обробки звукового сигналу та тексту для оцінювання позиції спікера за тематичними осями.
- Мета – розробити метод неймережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій.
- Завдання:
 - – проаналізувати моделі діаризації, ASR та великих мовних моделей для обраної задачі;
 - – формалізувати задачу у вигляді шести підзадач (2.1.1–2.1.6);
 - – реалізувати метод у формі експериментального вебзастосунку;
 - – експериментально оцінити якість методу за DER, WER, F1, MAE, κ , AUC, L(F).

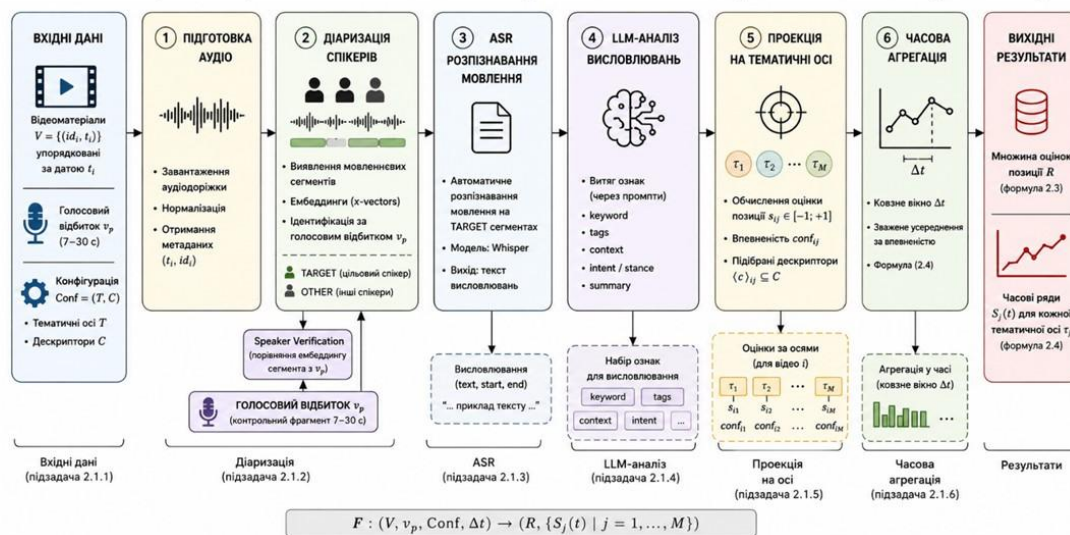
Огляд наукових публікацій

- Розглянуто 7 ключових публікацій 2021–2025 років за напрямками:
 - – діаризація з ідентифікацією спікерів та x-vectors;
 - – автоматичне розпізнавання мовлення на українській мові (FastConformer, Parakeet TDT);
 - – виявлення позиції (stance detection) великими мовними моделями;
 - – часовий концептуальний дрейф та виявлення точок зміни.
- Відсутні готові рішення, які поєднують усі чотири складові з підтримкою україномовного матеріалу.
- Це обґрунтовує доцільність розробки нового методу.

Схема декомпозиції задачі



Метод неймережевого аналізу: шість покрокових перетворень



Архітектура моделей: діаризація та розпізнавання мовлення

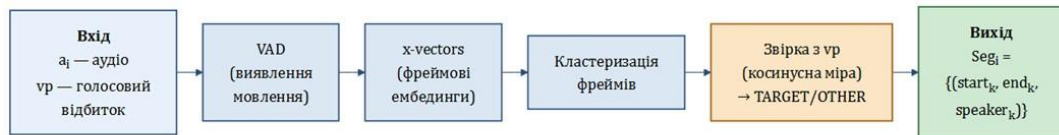


Рис. 2.3 – Діаризація з ідентифікацією за голосовим відбитком

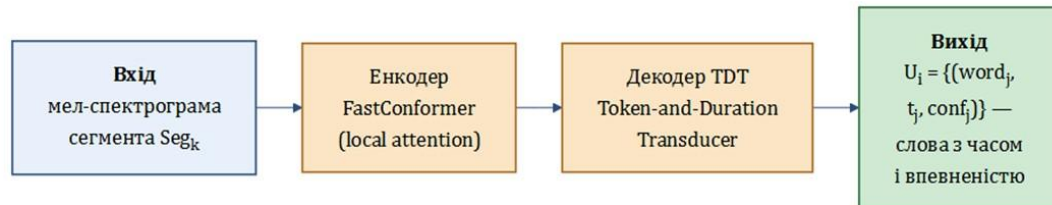
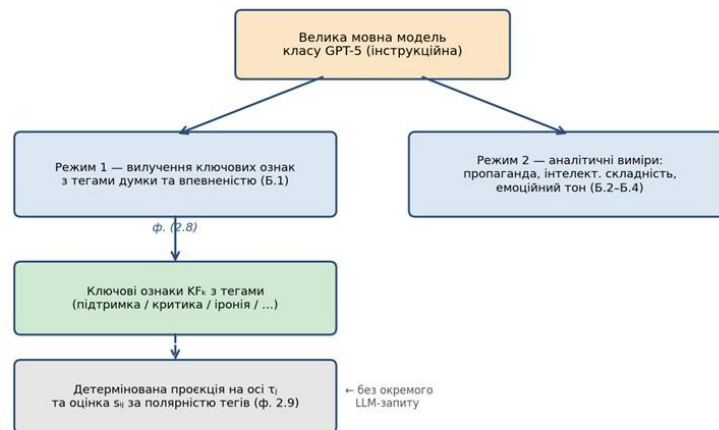


Рис. 2.4 – Автоматичне розпізнавання мовлення (FastConformer + TDT)

Велика мовна модель: вилучення ознак і проєкція на тематичні осі

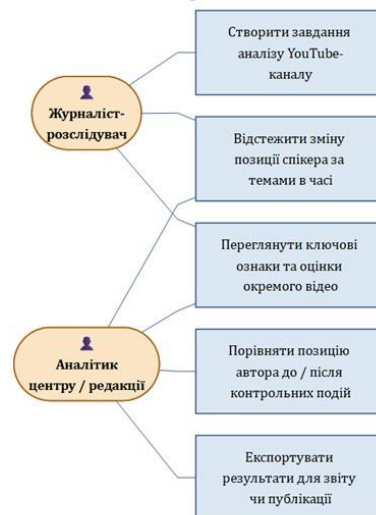
Велика мовна модель класу GPT-5: два інструкційні режими та детермінована проєкція на тематичні осі



Експериментальний датасет

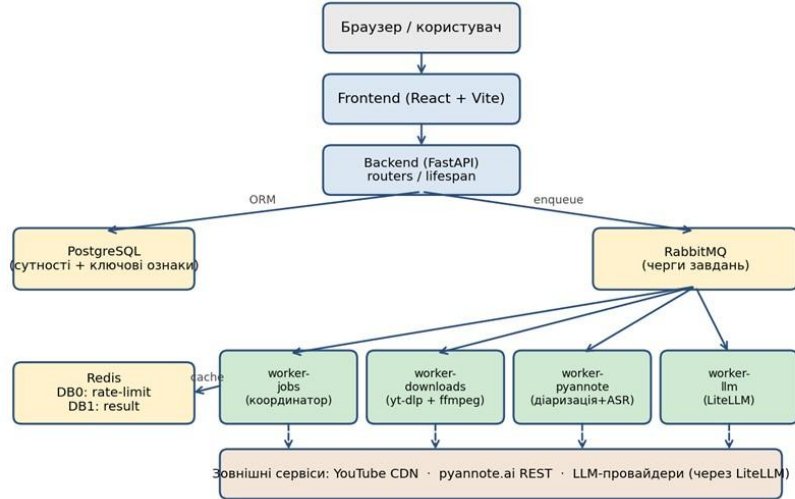
- 169 відеоматеріалів від 13 знеособлених авторів публічного дискурсу (період 2018–2026).
- Сумарна тривалість – 224,9 години україномовного мовлення.
- Тематичні осі (Т): COVID-19, EU-інтеграція, війна, корупція, мобілізація, соціальна політика.
- Експертна еталонна розмітка: 4 записи для DER, 3 записи для WER, 854 висловлювання для F1.
- Розмічники: філолог, журналіст, фахівець із комп'ютерних наук.
- Узгодженість експертів $k = 0,72$.
- Корпус повністю опрацьовано пайплайном – 6680 ключових ознак на 169 відео.

Сценарій використання інтелектуальної системи



Архітектура експериментального вебзастосунок

Архітектурна схема експериментального вебзастосунок



Інтерфейс системи: чотири основні екрани

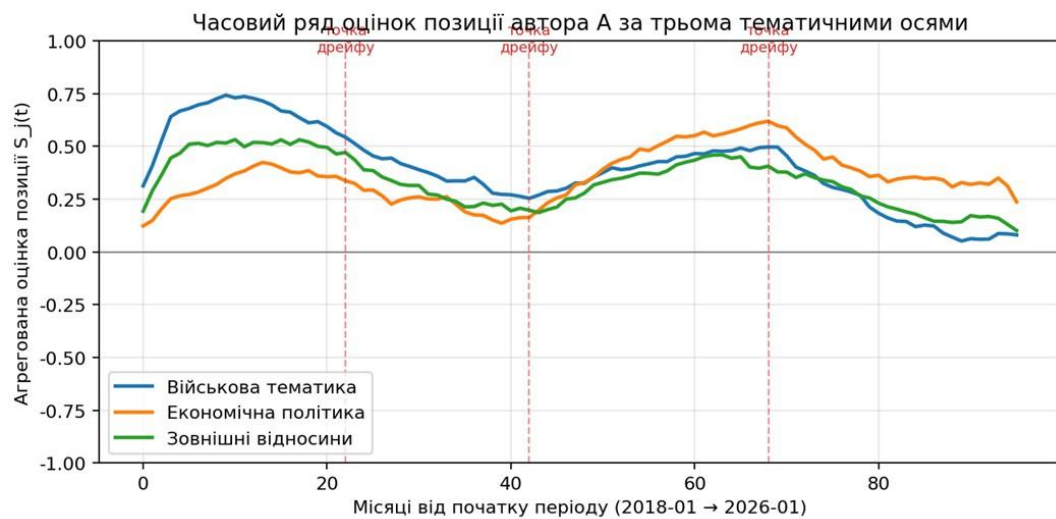
The interface consists of four main screens:

- Start New Analysis:** A form for entering a YouTube URL and selecting content type (All Content, Video Only, Shorts Only).
- Перегляд звіту (Report View):** A screen showing a video timeline and a list of video frames with their corresponding timestamps.
- ANALYSIS RESULTS:** A screen displaying a bar chart and detailed analysis results for the video.
- Результати аналізу (Analysis Results):** A screen showing a detailed text-based report with highlighted segments and analysis details.

Результати оцінювання якості

- Діаризація: DER = 12,9 % (4 контрольні записи).
- ASR: WER = 11,1 % на тестовій підмножині, 7,3 % на студійних записах.
- Класифікація позиції: F1_macro = 0,785 ± 0,021 за чотирма фолдами LLM (n = 854).
- MAE = 0,20; κ Коена = 0,68 (узгодженість з еталоном близька до κ розмітників 0,72).
- AUC = 0,889 на бінарній класифікації підтримка/протидія.
- Інтегральна оцінка методу: L(F) = 0,177.

Часова динаміка позиції автора А (2018–2026)



Висновки

- Розроблено метод нейромережевого аналізу часової семантичної динаміки висловлювань: відображення $F: (V, v_r, Conf, \Delta t) \rightarrow (R, \{S_j(t)\})$.
- Метод реалізує шість покрокових перетворень: підготовка аудіо, діаризація, ASR, вилучення ключових ознак, проєкція на осі, часова агрегація.
- Експериментальне дослідження на 169 відеоматеріалах підтвердило досягнення цільових рівнів за всіма групами метрик; $L(F) = 0,177$.
- Виявлено 87 точок зміни позиції, більшість з яких узгоджено з контрольними подіями.
- Запропоновано напрями вдосконалення: автоматичне виявлення кількості спікерів, словник імен сутностей, агентна обробка довгих відеоматеріалів.
- Дякую за увагу!



Wed Jun 17 09:01:57 EEST 2026, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism (http://ap.km.ua) v-16.718

Максимальне співпадіння з одним документом 2.0%

Словники перевірки: UA, US, RU. Помилки в документах: 15%

<p>ID: 275689 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій Додано в БД: 2026-06-17 Автора: Богдан ДЕНИСЕНКО Керівники: Олена СОБКО Консультанти: Опоненти:</p>	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	101577	772	4145 (4%)	58 (8%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Богдан ДЕНИСЕНКО

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

Науковий керівник: Олена СОБКО, Ph.D., ст. викл., каф. КН

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 3.85%

Коефіцієнт подібності 2: 2.05%

Мікропробіли: 0

Заміна букв: 40

Інтервали: 0

Білі знаки: 50

Дата створення звіту: 2026-06-16 19:51:04.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.


Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедур. Таким чином робота не приймається.

Обґрунтування:

2026-06-17

Дата

експерт

Петровський Р. Р. 

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

Автор студент групи КН-22-2 Богдан ДЕНИСЕНКО

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: ст. викладач каф. КН, д-р філософії Олена СОБКО

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відповідає</i>
1.2	Виявлені запозичення не є академічним плагіатом, розмішені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розмішені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

Підтвердження:

Запозичення, виявлені в роботі Богдана Денисенка, не є плагіатом, оскільки запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

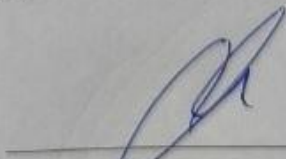
Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:

- за системою Anti-Plagiarism: 2%;

- за системою StrikePlagiarism КПІ: 3.85%.

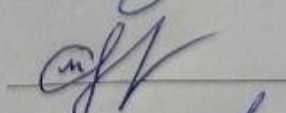
17.06.2026

Завідувач кафедри



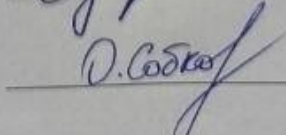
Олександр БАРМАК

Гарант освітньої програми



Олександр МАЗУРЕЦЬ

Керівник кваліфікаційної роботи



Олена СОБКО



ВІДГУК НАУКОВОГО КЕРІВНИКА на кваліфікаційну роботу бакалавра

студента гр. КН-22-2 Денисенка Богдана Олександровича

за темою Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій

1. Актуальність теми

Актуальність дослідження визначається тим, що семантичного моніторингу публічних висловлювань сьогодні швидко переходить від експериментальних рішень до практичного використання. У цьому контексті робота спрямована на розв'язання конкретної задачі - аналізу зміни позиції спікера щодо тем і подій у часі. Її значущість посилюється тим, що інтелектуального аналізу даних сьогодні активно використовуються в прикладних інформаційних системах, але потребують уважної перевірки якості результатів.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Робота узгоджується зі змістом підготовки бакалавра з комп'ютерних наук, оскільки охоплює постановку задачі, вибір методу, роботу з даними (часові ряди висловлювань, тематичні маркери та семантичні зсуви) і програмну реалізацію. Це відповідає очікуваним компетентностям спеціальності. Робота охоплює типові для спеціальності етапи: постановку задачі, аналіз джерел, вибір методу, програмну реалізацію та оцінювання результатів.

3. Професійні та особистісні якості бакалавра

У процесі підготовки роботи студент продемонстрував відповідальність, здатність сприймати зауваження та вміння поступово вдосконалювати результат. Okремо варто відзначити уважність до змісту практичної частини, адже саме вона вимагала найбільшої самостійності.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Основні результати отримані ним самостійно. Консультації керівника стосувалися уточнення напрямку пошуку, структури роботи та коректності інтерпретації окремих результатів. Консультації керівника мали спрямувальний характер

і стосувалися переважно уточнення методики, структури тексту та інтерпретації результатів.

5. Ступінь оволодіння методами дослідження

Методи дослідження опановані на достатньому рівні. У роботі коректно поєднано теоретичний аналіз, розроблення алгоритмічної частини та практичну перевірку на даних. Під час виконання роботи було використано методи, доречні для аналізу інтелектуального аналізу даних. Здобувач показав уміння добирати інструменти відповідно до задачі, а також пояснювати їх роль у загальній структурі дослідження.

6. Повнота та якість розкриття теми роботи

Якість розкриття теми можна оцінити позитивно: робота містить усі необхідні елементи бакалаврського дослідження і демонструє практичну завершеність. Тема розкрита достатньо повно: у роботі є обґрунтування актуальності, аналіз предметної області, опис методу, програмна реалізація та оцінювання результатів.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Матеріал викладено логічно й загалом літературно грамотно. Формулювання здебільшого точні, а послідовність подання допомагає простежити хід дослідження. Виклад матеріалу є логічним: від постановки проблеми автор переходить до аналізу підходів, далі - до власного методу та результатів.

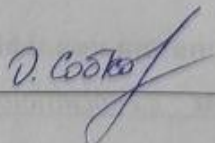
8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Розроблений метод має прикладний потенціал. Його можна розглядати як прототип або дослідницький модуль для систем, що працюють із такими даними: часові ряди висловлювань, тематичні маркери та семантичні зсуви. Практичне застосування можливе у прикладних інформаційних системах відповідної предметної області, а також у суміжних інформаційних системах, що працюють із подібними даними.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Робота виконана на належному науково-практичному рівні, відповідає вимогам до кваліфікаційних робіт бакалавра та заслуговує на оцінку «*відмінно*».

Керівник _____



ст. викладач каф. КН, д-р філософії Олена СОВКО



РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КН-22-2 Денисенка Богдана Олександровича*

за темою: *Метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій*

1. Актуальність обраної теми

Актуальність роботи визначається прикладною значущістю задачі аналізу зміни позиції спікера щодо тем і подій у часі. У сучасних умовах такі задачі дедалі частіше потребують не ручного аналізу, а відтворюваних інтелектуальних методів. Важливо, що актуальність тут визначається не лише модністю нейромережевих підходів, а й наявністю реальної прикладної задачі, для якої потрібні стійкі алгоритмічні рішення. Робота демонструє розуміння того, що інтелектуального аналізу даних потребують якісної підготовки даних, коректної постановки експерименту та обережної інтерпретації результатів.

2. Повнота розкриття мети та завдань роботи

Мета й завдання сформульовані достатньо чітко. Студент послідовно переходить від аналізу предметної області до побудови методу, не підмінюючи дослідницьку частину простим описом програмної реалізації. Автор послідовно пов'язує мету із завданнями, а завдання - з обраними методами реалізації, що позитивно впливає на цілісність дослідження. У роботі простежується намагання не просто описати метод нейромережевого аналізу часової семантичної динаміки висловлювань для моніторингу зміни позиції спікера щодо тем і подій, а показати, яким чином поставлену проблему можна довести до програмного рішення. Важливо, що виконання завдань має послідовний характер: кожний наступний етап спирається на попередній.

3. Зміст кожного розділу роботи

Структура розділів є традиційною, але виправданою. Теоретична частина формує основу для подальших рішень, проєктна частина деталізує метод, а практична частина демонструє його застосування на матеріалі: часові ряди висловлювань, тематичні маркери та семантичні зсуви. Розділи роботи взаємно доповнюють один одного: теоретична частина створює підґрунтя для вибору методу, а практична частина демонструє його застосування. Опис експериментів і програмної реалізації дає змогу оцінити не лише кінцевий результат, а й шлях його отримання. У роботі враховано, що для даних предметної області недостатньо просто застосувати готовий алгоритм: потрібно пояснити, які обмеження має обраний підхід і як вони можуть вплинути на результати.

4. Оцінка розробленої інформаційної системи, її практична цінність

Результати мають прикладний потенціал. Найпереконливішим є те, що метод можна адаптувати для подальшого розширення даних, уточнення критеріїв оцінювання та інтеграції у спеціалізовані сервіси. Практична цінність полягає в тому, що запропонований підхід може бути використаний у прикладних інформаційних системах відповідної предметної області, а також адаптований для суміжних задач.

5. Якість оформлення кваліфікаційної роботи бакалавра

Робота справляє охайне враження: структура розділів витримана, термінологія використовується коректно, висновки не суперечать змісту дослідження. Оформлення роботи загалом відповідає академічним вимогам: розділи структуровані, термінологія використовується послідовно, а висновки узгоджені з поставленими завданнями.

6. Недоліки кваліфікаційної роботи бакалавра

Окремі експериментальні результати бажано було б подати з ширшим порівнянням альтернативних підходів. Водночас основні висновки роботи залишаються обґрунтованими. Серед недоліків варто відзначити, що окремі експериментальні результати могли б бути подані з ширшим порівнянням альтернативних підходів. Проте для бакалаврського рівня представлений обсяг дослідження є достатнім.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота

Кваліфікаційна робота є завершеним дослідженням, відповідає вимогам до бакалаврських робіт і може бути допущена до захисту. Рекомендована оцінка « бачи кнв ».

Рецензент

