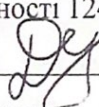


Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій

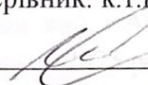
Комплексна бакалаврська кваліфікаційна робота на тему:

«СЕНТИМЕНТ-АНАЛІЗ НОВИН ДЛЯ ОЦІНЮВАННЯ ЗМІН У РИТОРИЦІ
ПУБЛІЧНИХ ОСІБ ТА ВИЯВЛЕННЯ СУПЕРЕЧНОСТЕЙ ЗА ДОПОМОГОЮ
ШТУЧНОГО ІНТЕЛЕКТУ. ПІДСИСТЕМА ВИЗНАЧЕННЯ ТА ВІЗУАЛІЗАЦІЇ
СЕНТИМЕНТУ»

Виконав: студент 4 курсу, групи СА-216
спеціальності 124 – «Системний аналіз»

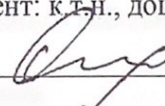
 Іван ДЕМКОВ

Керівник: к.т.н., доц. каф. САІТ

 Олексій КОЗАЧКО

« 08 » 01 2025 р.

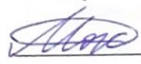
Рецензент: к.т.н., доц. каф. КН

 Олексій СІЛАГІН

« 12 » 06 2025 р.

Допущено до захисту

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

« 09 » 06 2025 р.

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра системного аналізу та інформаційних технологій
Рівень вищої освіти – перший (бакалаврський)
Галузь знань – 12 Інформаційні технології
Спеціальність 124 – «Системний аналіз»
Освітньо-професійна програма – Системний аналіз

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

“28” 05 2025 року

ЗАВДАННЯ НА БАКАЛАВРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Демкову Івану Владиславовичу

1. Тема роботи: «Сентимент-аналіз новин для оцінювання змін у риториці публічних осіб та виявлення суперечностей за допомогою штучного інтелекту. Підсистема визначення та візуалізації сентименту»
керівник роботи: Козачко О.М., к.т.н., доц. каф. САІТ
затверджені наказом ВНТУ від «20» 05 2025 року № 97
2. Строк подання студентом роботи «31» 05 2025 року
3. Вихідні дані до роботи:
 - 1) Дані висловлювань публічних осіб із новинного веб-сайту CNN;
4. Зміст текстової частини:
 - 1) Загальна характеристика об'єкту дослідження;
 - 2) Вибір оптимальних методів для розв'язання поставленої задачі;
 - 3) Моделювання та реалізація інтелектуальної системи оцінювання змін у риториці публічних осіб;
 - 4) Моделювання й реалізація системи візуалізації та аналіз результатів роботи системи
5. Перелік ілюстративного матеріалу:
 - 1) Діаграма потоку даних конвеєру аналізу сентименту публічних осіб;
 - 2) Прототип інтерфейсу віджету представлення результатів аналізу;
 - 3) Візуальне представлення результатів повного аналізу сентименту за обраними темами;

6. Консультанти розділів роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання «28» 03 2025 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання		Примітка
		початок	закінчення	
1	Аналіз предметної області	01.04.2025	20.04.2025	вик
2	Порівняльний аналіз проблематики	20.04.2025	30.04.2025	вик
3	Вибір оптимальних інформаційних технологій	04.05.2025	10.05.2025	вик
4	Розроблення моделі аналізу риторики публічних осіб	10.05.2025	20.05.2025	вик
5	Оформлення матеріалів до захисту БКР	20.05.2025	30.05.2025	вик

Студент



Іван ДЕМКОВ

Керівник роботи



Олексій КОЗАЧКО

АНОТАЦІЯ

Бакалаврська кваліфікаційна робота складається з 87 сторінок формату А4, на яких є 39 рисунків, список використаних джерел містить 25 найменувань.

Метою роботи є підвищення точності оцінювання настрою публічних осіб до тематики з аналізу новинних публікацій за рахунок використання штучного інтелекту, що дозволяє виявити суперечності у позиціях для використання у складних системах підтримки прийняття рішень.

В роботі наведено моделювання та розробку моделі інформаційної системи інтелектуального системного аналізу настрою публічних осіб до теми з визначенням протиріч, оптимізацію параметрів моделі для збільшення точності отриманої оцінки. Досліджено способи представлення отриманих даних аналізу, підготовлено макет візуалізації та реалізовано у виді віджету на базі React Native. Роботу розробленої інформаційної системи аналізу настрою публічних осіб до теми із виявленням протиріч було перевірено та визначено, що вона дотримується поставлених вимог і спрощує процес аналізу публічного ставлення.

Ключові слова: інформаційна система, штучний інтелект, аналіз ставлення, настрой-аналіз відстежування протиріч у висловлюваннях, аналіз медіа, Text mining, LLM, LangGraph, React Native.

ABSTRACT

The bachelor's thesis consists of 87 A4 pages with 39 figures and a list of references containing 25 titles.

The aim of this paper is to improve the accuracy of assessing the sentiment of public figures towards the topic of news analysis by using artificial intelligence to identify contradictions in positions for use in complex decision support systems.

The paper presents the modelling and development of a model of an information system for intelligent system analysis of public sentiment on a topic with the identification of contradictions, optimisation of model parameters to increase the accuracy of the obtained assessment. The ways of presenting the obtained analysis data are investigated, a visualisation layout is prepared and implemented as a widget based on React Native. The operation of the developed information system for analysing the sentiment of public figures towards a topic with the detection of contradictions was tested and it was determined that it meets the requirements and simplifies the process of analysing public attitudes.

Keywords: information system, artificial intelligence, attitude analysis, sentiment analysis, tracking contradictions in statements, media analysis, Text mining, LLM, LangGraph, React Native.

ЗМІСТ

ВСТУП	4
1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ	6
1.1 Проблематика аналізу ставлення публічних осіб	6
1.2 Опис об'єкта досліджень	7
1.3 Аналіз аналогів та інформаційних технологій	9
1.4 Області практичного використання інформаційної системи аналізу сентименту	12
1.5 Постановка задачі	12
1.6 Висновки	14
2 ВИБІР ОПТИМАЛЬНИХ МЕТОДІВ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ	15
2.1 Вибір підходу до інтелектуального аналізу тексту	15
2.2 Методи вирішення поставленої задачі	18
2.3 Вибір інформаційних технологій для роботи	20
2.4 Особливості налаштування моделі аналізу сентименту	23
2.5 Висновки	25
3 МОДЕЛЮВАННЯ ТА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ДЛЯ ОЦІНЮВАННЯ ЗМІН У РИТОРИЦІ ПУБЛІЧНИХ ОСІБ	26
3.1 Моделювання змін у риторичності публічних осіб	26
3.2 Підбір оптимальних налаштувань моделі аналізу сентименту	33
3.3 Вибір оптимальної великої мовної моделі для вирішення поставленої задачі	36
3.4 Висновки	44
4 МОДЕЛЮВАННЯ Й РЕАЛІЗАЦІЯ СИСТЕМИ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ СИСТЕМИ	46
4.1 Підбір оптимального способу представлення результатів аналізу	46
4.2 Прототипування інтерфейсу віджету	49
4.3 Реалізація системи візуалізації сентименту	51
4.4 Інтерпретація результатів аналізу сентименту із визначенням суперечностей у риторичності публічних осіб	54
4.5 Висновки	59
ВИСНОВКИ	60

	3
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	63
Додаток А (обов'язковий) Технічне завдання на бакалаврську кваліфікаційну роботу	66
Додаток Б (обов'язковий) Протокол перевірки бакалаврської кваліфікаційної роботи	69
Додаток В (довідниковий) Фрагмент лістингу програми	70
Додаток Г (обов'язковий) Ілюстративна частина	83
Додаток Д (довідниковий) Акт впровадження системи	89

ВСТУП

Актуальність теми. Сучасний інформаційний простір характеризується надзвичайно великим потоком великої кількості новинних матеріалів та заяв публічних осіб, що ускладнює процес їх опрацювання. Особа, що приймає рішення, змушена витратити значний час на ручний аналіз ставлення особи до теми та пошук протиріч у позиціях. У той же час розвитку набувають системи підтримки прийняття рішень із застосуванням штучного інтелекту, здатні автоматизувати класифікацію сентименту й виявляти суперечності. Це особливо актуально для складних систем аналізу публічного ставлення – від фінансових ринків до державних структур і соціальних медіа, де своєчасність і якість прийнятих рішень прямо залежать від швидкості й точності обробки інформації.

Мета і задачі дослідження. Метою дослідження є підвищення точності оцінювання сентименту публічних осіб до тематики з аналізу новинних публікацій за рахунок використання штучного інтелекту, що дозволяє виявити суперечності у позиціях для використання у складних системах підтримки прийняття рішень.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- проаналізувати предметну область моделювання сентименту публічних осіб до теми з відстежуванням протиріч;
- обрати оптимальні методи розв'язання задачі аналізу сентименту риторики публічних осіб;
- розробити моделі аналізу сентименту до теми у риториці публічних осіб;
- розробити підсистему аналізу сентименту до теми у риториці публічних осіб;
- оптимізувати параметри моделі аналізу сентименту до теми у риториці публічних осіб;

- розробити підсистему візуалізації результатів моделювання аналізу настрою публічних осіб до теми.

Об’єктом дослідження є процес оцінювання настрою публічних осіб з новинних публікацій для виявлення суперечностей у позиціях.

Предметом дослідження є методи та інформаційні технології оцінювання настрою публічних осіб з новинних публікацій для виявлення суперечностей у позиціях за допомогою штучного інтелекту.

Публікації. За результатами даної роботи була зроблена доповідь на тему «Аналіз змін у позиціях публічних осіб на основі новинних публікацій» на LIV Всеукраїнській науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету (2025) з публікацією тез [1].

1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ

1.1 Проблематика аналізу ставлення публічних осіб

У сучасному інформаційному просторі публічні висловлювання лідерів думок, політиків та експертів поширюються через численні онлайн-ЗМІ, створюючи величезний масив текстових даних, які відображають позиції ключових діячів суспільно-політичного дискурсу. Це створює надмірно перевантажену площину дискурсу, що породжує можливість маніпуляції фактами та зловживанню публічністю задля персональних репутаційних здобутків, оскільки новинні джерела відіграють переважну роль у формування політичної свідомості читача [2]. Для спрощення споживання інформації про події користувачами була винайдена концепція новинного агрегатора, що за допомогою штучного інтелекту комбінує статті із різних джерел та формує короткий підсумок подій [3]. Проте навіть такі рішення не дозволяють бути впевненими у достовірності написаного, оскільки на фінальну версію новини впливають фейкові статті [4,5].

Звичайні підходи до аналізу контенту, що ґрунтуються на ручній обробці та статичних вибірках, демонструють суттєві обмеження: вони не можуть гарантувати своєчасність оцінок, глибину охоплення та виявлення непослідовних заяв у режимі реального часу. Внаслідок цього аналітики стикаються з неможливістю відстежувати еволюцію ставлення окремих публічних осіб до актуальних тем у динамічному середовищі новин, де кожна хвиля повідомлень може містити суперечливі дані. Ця задача є частиною більш широкої теми дослідження громадської думки, модернізація якої активно досліджується [6,7].

Особливо складним є завдання виявлення протиріч у висловленнях однієї й тієї самої особи: через великий обсяг публікацій та швидку зміну контексту вручну порівняти історичні і сучасні заяви практично неможливо. Швидкість виходу нових матеріалів і фрагментарність інформації призводять до того, що аналітичні висновки часто базуються на уривках тексту, які не відображають

повної картини змін у риторичі. Це збільшує ризик хибних інтерпретацій і затримок у формуванні обґрунтованих рекомендацій для прийняття рішень.

Отже, постає нагальна потреба в автоматизованій системі, здатній цілодобово збирати новинні стрічки, структурно аналізувати ставлення конкретної особи до заданої теми та виявляти внутрішні суперечності у її висловлюваннях. Така підсистема має забезпечити поєднання швидкості обробки великих обсягів даних із достатньою точністю оцінок, що дозволить мінімізувати час, необхідний для підготовки аналітичного огляду, та підвищить корисність висновків для кінцевого користувача.

Для досягнення цієї мети необхідно передусім відстежувати три ключові компоненти: поточну тональність заяв публічної особи щодо теми, невідповідність між новими та попередніми висловами тієї самої особи та формування інтегрального підсумку її позиції. Об'єднуючи ці елементи, передбачається створити інструмент, який спростить процес аналізу, забезпечить швидке візуальне відображення динаміки громадської думки та сприятиме оперативному й обґрунтованому прийняттю рішень.

1.2 Опис об'єкта досліджень

Із попереднього аналізу проблематики окреслюється потреба в системі, яка не обмежується простим агрегуванням новинних повідомлень, а здатна послідовно виокремлювати з них ключові події та висловлювання публічних осіб, оцінювати тональність цих висловлювань і водночас фіксувати внутрішні суперечності. Саме таке комплексне рішення забезпечує аналітику необхідними параметрами – від зміни ставлення до теми до наявності неспівпадінь у позиціях однієї й тієї самої особи, що дозволяє глибше зрозуміти динаміку суспільної думки. Ця ідея формалізує об'єкт дослідження – процес автоматизованої обробки новинних текстів із метою визначення настрою та виявлення протиріч у висловлюваннях публічних осіб.

Потреба в такому комплексному підході зумовлена зростанням складності систем прийняття рішень, де аналітик повинен враховувати не лише факт появи

нової заяви, а й її контекст, історію попередніх висловлювань тієї самої особи та потенційні внутрішні суперечності. Таким чином, візуалізація стає ключовим елементом підтримки прийняття рішень, оскільки вона знижує когнітивне навантаження, демонструючи еволюцію позицій і допомагаючи швидко ідентифікувати критичні точки зміни настроїв або наявність несумісних заяв. Умовою оптимальності результату прогнозування у даному випадку буде досягнення високого рівня точності проведеного аналізу порівняно із експертним оцінюванням. Це у свою чергу дозволить зменшити час, витрачений на визначення позиції особи, що суттєво спростить прийняття рішень на основі цієї інформації.

Тому мета проведеного дослідження полягає в оптимізації процесу прийняття рішень користувачем шляхом створення інструментарію, який дозволяє скоротити час на аналіз новин і спростити отримання висновків. Предметом дослідження виступають методи та інформаційні технології комплексного аналізу та візуалізації сентименту до теми з подальшим виявленням суперечностей у позиціях публічних осіб. Такий підхід передбачає формалізацію умов оптимальності результату, де головним критерієм є досягнення достатнього рівня точності інтерпретації тональності та суперечностей порівняно з експертним оцінюванням.

Концептуально система складається з аналітичного модуля, який здійснює класифікацію сентименту, виявлення логічних невідповідностей і оцінювання «цікавості» висловлювань, та візуалізаційного модуля, що відтворює часові шкали та графічні позначки змін настрою і суперечностей. Модуль виведення результатів оцінювання ставлення публічної особи до теми повинен представляти у зручній та зрозумілій формі зміну ставлення особи до теми на хронологічній шкалі, відображаючи при цьому довідникову інформацію у підказках. Концептуально дана система повинна формувати дашборд подібно тому, як представлено на рисунку 1.1.



Рисунок 1.1 – Приклад візуалізації настрою відносно часу події

Інші компоненти – збір, попередня обробка та зберігання даних – на даному етапі розглядаються як необхідна підґрунтя для забезпечення безперебійного функціонування аналітики та відображення результату в зрозумілій формі. Такий логічний перехід від об'єкта до реалізації підкреслює наукову новизну дослідження й забезпечує обґрунтоване формування подальших технічних рішень.

1.3 Аналіз аналогів та інформаційних технологій

У сучасному полі автоматизованого аналізу новин виділяються кілька провідних платформ і дослідницьких ініціатив, які застосовують техніки оброблення природної мови для класифікації настрою та групування матеріалів у події. Зокрема, система Event Registry пропонує інтелектуальне кластерування новинних статей у єдині події з урахуванням семантичної близькості й тимчасового контексту, водночас здійснюючи базовий настроєвий аналіз для відстеження настроїв публіки щодо конкретних тем чи осіб (рис. 1.2). Завдяки функціям візуалізації, що підтримують лінійні та стовпчикові діаграми

зміни тональності в часі, користувачі отримують інструменти для моніторингу еволюції ставлення до ключових об'єктів дослідження, проте ця платформа не фокусується на виявленні суперечностей у висловлюваннях окремих персоналій та не забезпечує гнучкого відстеження «цікавості» окремих подій чи думок через інтегровану систему оцінок [8].

Аналогічно, проєкт GDELT у відкритому доступі надає глобальний граф новин із автоматично згенерованими тональними оцінками та можливістю візуалізації часових рядів емоційних реакцій у понад 100 мовах, однак його інтерфейс і аналітичні модулі орієнтовані здебільшого на макрорівень подій і не підтримують виявлення індивідуальних суперечностей у висловлюваннях публічних осіб (рис. 1.3) [9].

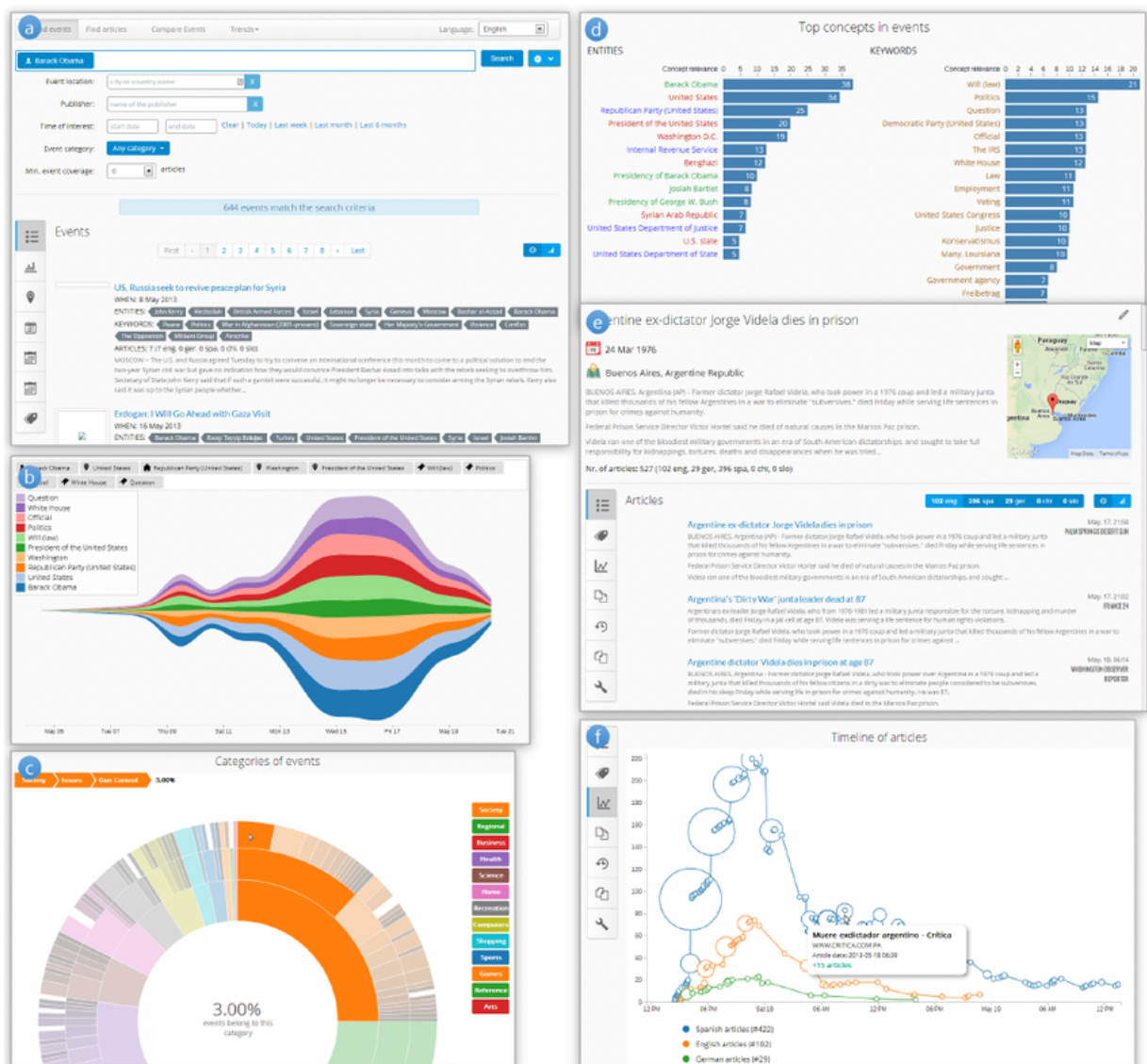


Рисунок 1.2 – Інтерфейс системи EventRegistry

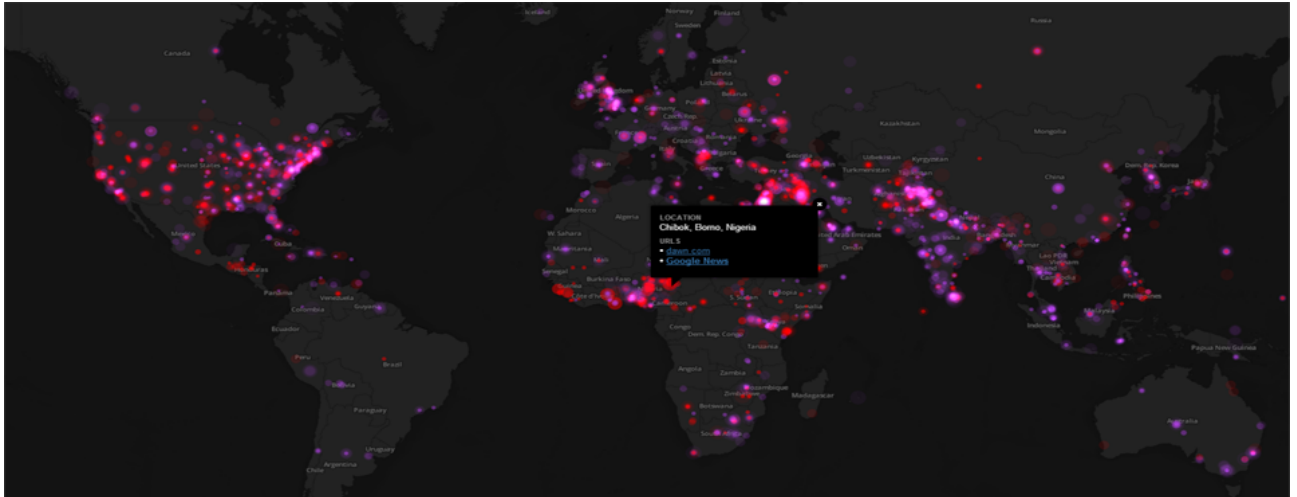


Рисунок 1.3 – Інтерфейс проєкту GDELT

Серед переваг розглянутих рішень можна назвати масштабованість (GDELT), можливість оперативного оновлення даних у майже у реальному часі (Event Registry) та високу якість тональної класифікації за рахунок великих наперед натренованих моделей. Натомість відсутність уніфікованої системи для виявлення суперечностей, обмежена гнучкість у налаштуванні візуальних елементів і відсутність комплексного підходу до вимірювання «цікавості» подій та думок у рамках єдиної конвеєрної архітектури суттєво звужують аналітичний потенціал цих рішень.

На основі вивчених аналогів формується теоретичний концепт моделі проєкту як модульної системи, що поєднує сильні сторони: масштабоване збирання та сегментацію текстів, використання аналітичного конвеєра для багаторівневого оцінювання настроїв, реалізацію модуля виявлення суперечностей шляхом порівняння з історичними даними, а також інтерактивну візуалізацію з хронологічними шкалами для інтерпретації стану публічної думки. Такий підхід дозволить не лише відслідковувати динаміку публічних настроїв, але й оперативно ідентифікувати несумісні позиції окремих осіб, що є критично важливим для підтримки прийняття рішень у складних інформаційних екосистемах.

1.4 Області практичного використання інформаційної системи аналізу настроїв

Інформаційна система аналізу настроїв має широкий спектр практичних застосувань, насамперед у вигляді інтерактивного віджету, який дозволяє як професійним аналітикам громадської думки, так і пересічним читачам відстежувати динаміку ставлення публічних осіб до обраної теми. У фінансових установах це рішення може допомогти трейдерам та інвесторам оперативно оцінювати настрої керівників компаній та новинні сигнали ринку, у державних системах раннього попередження – прогнозувати соціально-політичні ризики на основі змін у риторичній політиці. Соціальні медіа-платформи зможуть інтегрувати віджет для модерації контенту й рекомендацій, а екологічні служби – оперативно реагувати на згадки про екологічні катастрофи чи загрози. У корпоративній аналітиці він слугуватиме єдиним інструментом для поєднання медіа- та внутрішніх звітів, а новинні ресурси – збільшуватимуть охоплення аудиторії за рахунок залучення користувачів до інтуїтивного дослідження хронології оцінок і перевірки стабільності політичних або публічних позицій.

1.5 Постановка задачі

Виходячи з аналізу сучасних підходів до автоматизованого аналізу новинних текстів, постає потреба у створенні комплексної системи, що поєднує в собі модулі інтелектуального аналізу природної мови та інтерактивної візуалізації результатів. Існуючі рішення, хоч і забезпечують базову класифікацію тональності чи просте відображення хронології подій, але не приділяють достатньої уваги виявленню внутрішніх суперечностей у позиціях публічних осіб та зручному перегляду динаміки їхніх висловлювань. Часто у таких системах відсутній механізм накопичення історії заяв конкретної особи з автоматичним порівнянням нових даних із вже наявними, що значною мірою обмежує їхню корисність та загальний контекст прийняття рішень у складних інформаційних екосистемах.

На основі виокремлених сильних і слабких сторін аналогічних розробок сформовано концепцію модульного підходу, у якому процеси збору, попередньої обробки та зберігання новинних матеріалів розглядаються лише як базова інфраструктура. Головний фокус такої моделі покладено на послідовний семантичний аналіз фрагментів тексту за допомогою багаторівневого конвеєра обробки та на реалізацію механізму виявлення суперечностей шляхом порівняння отриманих тональних оцінок із історичними записами. У результаті з'являється можливість не тільки вимірювати тональність висловлювань, а й автоматично маркувати фрагменти, котрі суперечать попереднім позиціям, що є критично важливим для аналітика під час прийняття рішення.

Ключовим елементом системи стає інтерактивна візуалізація, що дозволяє надати користувачу зручний інструментарій для фільтрації за темою, особами та часовим інтервалом, а також чітко показувати зміну позицій публічних осіб на хронологічній шкалі з поєднанням діаграм та індикаторів суперечливості. Такий підхід значно зменшує когнітивне навантаження: замість опрацювання масиву сирих даних аналітик може оперативно оцінити вагомість кожного висловлювання та виявити потенційні розбіжності в тезах, що підвищує швидкість і якість прийняття рішень.

Таким чином, остаточна постановка задачі полягає у розробці моделі модульної інформаційної системи, яка забезпечить: по-перше, формування чітких функціональних вимог до підсистеми аналізу настрою та суперечностей на основі огляду наукових публікацій і аналізу користувацьких сценаріїв; по-друге, проєктування багаторівневого конвеєра для оцінювання тональності текстів із узгодженою схемою сегментації, формування промптів та нормалізації результатів; по-третє, створення механізму виявлення суперечностей на основі формалізованих логічних умов і порівняння з історичними даними; по-четверте, реалізацію інтерфейсу інтерактивної візуалізації з хронологічними шкалами та комбінованими бульбашковими діаграмами для ефективної підтримки прийняття рішень у складних системах аналізу публічного ставлення.

1.6 Висновки

У цьому розділі було проаналізовано предметну область моделювання ставлення публічних осіб до теми із відстежуванням протиріч, оцінено аналоги подібних інформаційних систем та конкретизовано області можливого використання програмного рішення. За отриманими вимогами сформульовано задачі із чітким зазначенням необхідних функціональних та нефункціональних вимог до розроблюваної моделі.

2 ВИБІР ОПТИМАЛЬНИХ МЕТОДІВ РОЗВ'ЯЗАННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ

2.1 Вибір підходу до інтелектуального аналізу тексту

Задля вирішення задачі аналізу позиції публічної особи щодо теми потрібно визначити спосіб отримання висновків із даних природньої мови. Методологія отримання структурованої інформації із великих обсягів тексту є досить дослідженою та має назву text mining [10,11]. Варто зазначити, що необхідним результатом екстракції у контексті даного дослідження є аналіз ставлення публічної особи відносно певної теми обговорення, а не загальне емоційне забарвлення тексту, тому варто насамперед проаналізувати існуючі підходи аналізу політичних позицій [12].

Загалом методи обробки природньої мови можна класифікувати за рівнем складності та глибини розуміння написаного, яке вони забезпечують. Ілюстрація взаємозв'язку між різними методами екстракції даних із тексту наведена на рисунку 2.1.

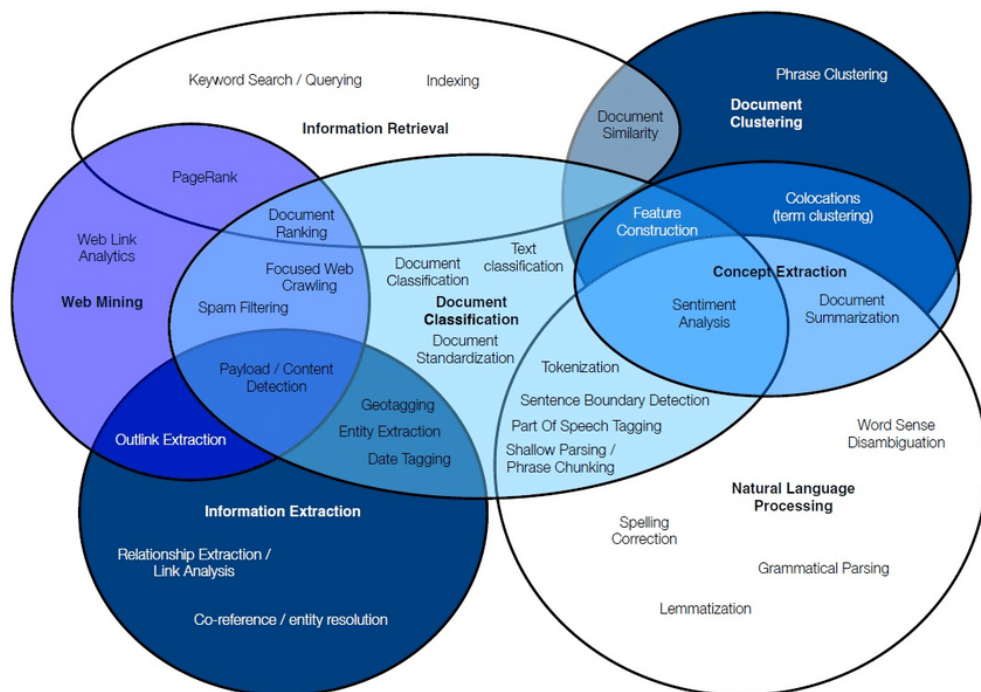


Рисунок 2.1 – Взаємозв'язки між різними методами text mining та їх функціональністю [11]

Найпростіший метод отримання позиції особи – використання лексиконно-граматичних методів, заснованих на формальних правилах, шаблонах та словниках емоційно забарвлених лексем. Це дозволяє здійснювати базову синтаксичну обробку та тональну класифікацію текстів, однак характеризувалися низькою гнучкістю та обмеженими можливостями узагальнення [13].

Розвиток та вдосконалення технологій виокремило також суто статистичні методи, що включають моделі n-грам, наївний баєсівський класифікатор, логістичну регресію та метод опорних векторів. Вони забезпечують порівняно кращу адаптивність до нових текстових даних завдяки використанню ймовірнісних закономірностей розподілу слів. Проте, такі методи залишаються обмеженими у розумінні складних лінгвістичних конструкцій і контекстуальних залежностей [12,14].

Важливим етапом розвитку стало запровадження векторних подань слів (наприклад, Word2Vec, GloVe, FastText), які надали змогу відображати семантичну близькість лексем у багатовимірному просторі. Це дозволило поліпшити якість класифікації текстів та розширити можливості тематичного моделювання, хоча ці підходи й надалі працювали на рівні слів або коротких фраз без урахування повного контексту висловлювання [15].

Подолати обмеження, – розуміння контексту, дозволили рекурентні нейронні мережі (зокрема LSTM та GRU), що забезпечували обробку послідовностей змінної довжини. Однак ці моделі мали обмежену здатність масштабування до довгих текстів та були ресурсомісткими під час навчання [7].

Суттєвий прорив у сфері обробки природної мови забезпечила архітектура Transformer (рис. 2.2), яка базується на механізмі самоуваги (self-attention). Вона дала змогу паралельно обробляти текстові послідовності та ефективно враховувати довготривалі контекстуальні залежності. Це стало основою для побудови великих мовних моделей (LLM), зокрема BERT, RoBERTa, GPT, T5, FLAN тощо [16,17,18].

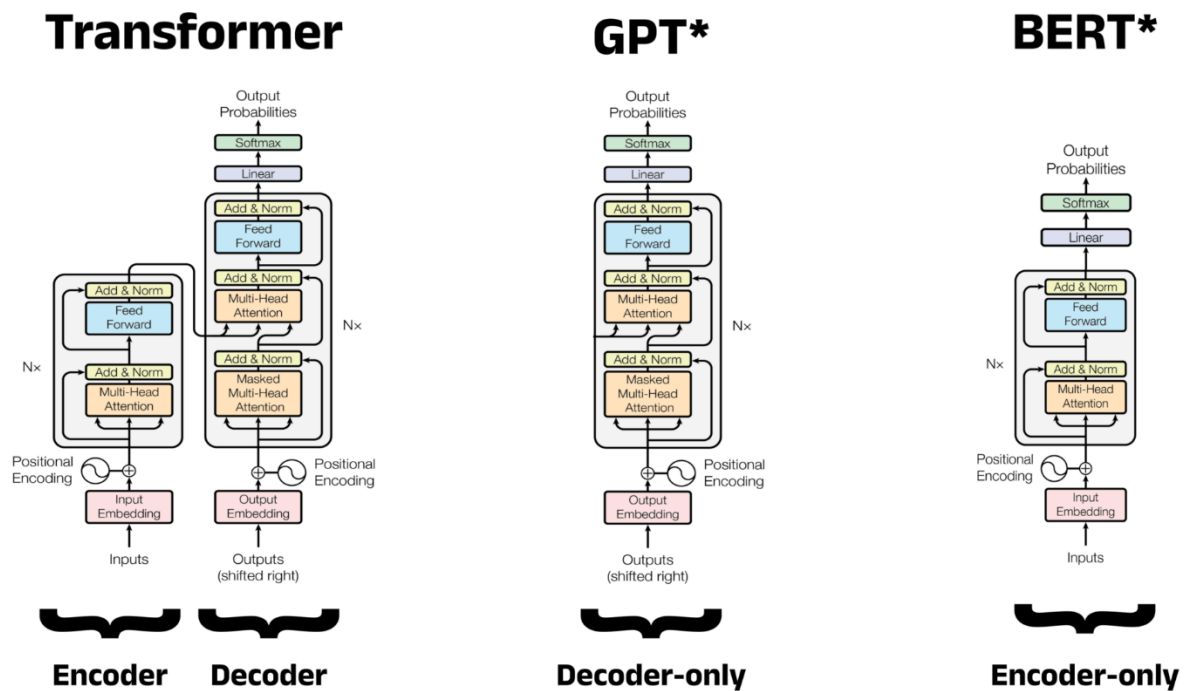


Рисунок 2.2 – Архітектура Transformer-моделі [16]

Виходячи з найбільшої інтелектуальності та сучасності застосування LLM є доцільним у контексті задачі, пов'язаної з виявленням ставлення публічних осіб до певних тем, визначенням риторичних змін, виявленням суперечностей та побудовою висновків на основі вхідних текстів. На відміну від традиційних підходів, спрямованих переважно на визначення емоційного забарвлення або тональності тексту, LLM дозволяють здійснювати глибинний контекстуальний аналіз та забезпечують здатність моделювати високорівневі логічні залежності [18].

Отже, враховуючи функціональні вимоги до інформаційної системи та порівняльні характеристики сучасних методів обробки тексту, використання великих мовних моделей на базі Transformer архітектури як засобу інтелектуального аналізу тексту є обґрунтованим та найбільш перспективним для реалізації поставленої задачі [12].

Завдяки здатності обробляти текст у контексті великих обсягів даних і аналізувати семантичні зв'язки на рівні діалогових запитів, LLM можуть «осмислити» зміст документу, сформулювати відповідь на запитання про прихильність людини до теми і навіть виявити суперечності у порівнюваних

висловлюваннях. Це у свою чергу надає можливість генеративного моделювання, коли модель не тільки декларує наявність позитивної чи негативної тональності, а створює на основі запиту (промпту) та переданого контексту чіткий висновок про позицію публічної особи із коментарем-підсумком. Це дозволяє розширити таку систему аналізу сентименту та додати необхідний довідковий матеріал. Перспективи використання LLM також полягають у високій гнучкості та масштабованості отриманих програмних рішень, що дозволяють обирати оптимальний варіант налаштування шляхом проведення тестувань.

2.2 Методи вирішення поставленої задачі

Для системного дослідження великих масивів новинних матеріалів необхідний підхід, що поєднує автоматизацію збору інформації з потужними методами обробки тексту та інтелектуального аналізу. Першим кроком є масове збирання релевантних статей за обраними темами: завдяки налаштованим запитам система обходить результати пошуку на новинному ресурсі та зберігає повні тексти статей у структурованому вигляді. Це дозволяє накопичити достатній корпус даних для подальшого аналізу без ручного втручання.

Наступний етап – підготовка текстів до глибинного розбору. Оскільки головний інструмент інтелектуального аналізу, – сучасні великі мовні моделі, мають обмеження на довжину контексту, статті розбиваються на змістові блоки з невеликим перекриттям, що забезпечує збереження логічного зв'язку між фрагментами [19]. Такий підхід дозволяє аналізувати кожен фрагмент окремо, водночас підтримуючи глобальну цілісність тексту через механізм накопичення контексту і не втрачаючи деталей через надмірне абстрагування. Модель аналітичної частини аналізу сентименту до теми у публічних висловлюваннях полягає у створенні конвеєра обробки інформації, де великі мовні моделі, промпти та конфігурація системи приймають роль гіперпараметрів. Задля досягнення оптимальності потрібно підібрати такі налаштування, за яких

критерій оптимальності, – точність відносно експертних оцінок, буде максимізовано.

Ключем до розуміння громадської думки є здатність системи виділяти дві категорії інформації: події загального характеру та висловлювання публічних осіб. Для цього впроваджується багаторівнева обробка тексту: спочатку розпізнаються сутності та події, потім окремим підходом аналізуються цитати та коментарі політиків чи експертів. За допомогою цього поділу вдається виділити не лише фактологічні події, а й позиції конкретних особистостей.

Щоб оцінити належність висловлювання до певного ставлення (підтримка, критика, нейтральність), застосовується аналіз тональності за допомогою мовної моделі, що перетворює якісні судження в числові показники. При цьому, завдяки використанню саме LLM, досягається осмисленість такого оцінювання, а не лише зазначення емоційного забарвлення фрагменту тексту. Додатковий модуль зіставляє нові висловлювання із вже наявними та виявляє суперечності – це дозволяє простежити, чи були в заявах політика раніше різкі зміни позиції або внутрішні логічні розбіжності.

Наступним етапом після виявлення ролі персональних показників у загальному аналізі було сформульовано вимогу до запровадження механізму, який би оцінював «цікавість» новини або висловлювання ще до глибокого занурення в деталі тексту. Такий підхід дозволить відсіювати малозначущі повідомлення та залишати в полі зору користувача лише ті матеріали, які можуть суттєво вплинути на сприйняття теми. У разі подій загального характеру «цікавістю» пропонується вважати поєднання релевантності до теми, потенційного впливу на розвиток обговорення та новизни інформації. Якщо ж йдеться про висловлювання публічних осіб, показником «цікавості» стає комбінація релевантності до теми, внеску в дискусію та наявності суперечливих тверджень. Завдяки цьому ранжуванню система зможе візуально розмежовувати дані за рівнем «цікавості», забезпечуючи швидке орієнтування в потоці новин, а також зменшуючи когнітивне навантаження аналітика під час прийняття рішень.

Зібрані структуровані дані (дата, тема, суб'єкт, тональність, показник цікавості події або висловлювання, маркери суперечностей) агрегуються у часові

ряди. На основі них генерується динамічна хронологічна візуалізація, яка демонструє еволюцію думок окремих осіб та загальне «суспільне відчуття» щодо теми. Такий підхід дозволяє не лише відстежувати зміни уподобань, а й виявляти непослідовність і маніпуляції в комунікації публічних фігур.

У підсумку, комбінація автоматичного збору, гнучкого розбиття тексту, багаторівневого NLP-аналізу та часової візуалізації створює потужний інструмент для масштабного моніторингу та вивчення громадської думки за обраними темами. Цей методичний ланцюжок може бути адаптований до різних інформаційних ресурсів і налаштовуваний під специфіку конкретної предметної області.

2.3 Вибір інформаційних технологій для роботи

Для реалізації моделі визначення сентименту, яка досліджується у даній роботі, обрано ряд технологій та інструментів, що найкраще відповідали вимогам до ефективності, гнучкості та масштабованості системи. Основна частина логіки була реалізована за допомогою мови програмування Python, яка є однією з найпопулярніших у сфері обробки даних, автоматизації та побудови систем зі штучним інтелектом. Python дозволив швидко розгорнути базову інфраструктуру та інтегрувати зовнішні бібліотеки, необхідні для взаємодії з веб-ресурсами, базами даних і мовними моделями [20].

Для побудови моделі системи на основі LLM було обрано фреймворк LangGraph, який забезпечує декларативний підхід до побудови графів станів та потоків обробки повідомлень [21]. LangGraph дозволив чітко структурувати логіку в окремі вузли (ноди), кожен з яких відповідає за конкретну функцію: від завантаження даних із бази до генерації відповідей за допомогою LLM. Такий підхід покращує модульність та тестованість рішення, а також спрощує масштабування системи шляхом додавання нових вузлів.

Як агент для аналізу тексту у створеній системі інтелектуального аналізу сентименту до теми були обрані кілька великих мовних моделей. Умовами відбору слугували відкритість ваг, обчислювальний об'єм та рейтингова оцінка

відносно інших LLM на відкритих бенчмарках (рис. 2.3-2.4). За результатами були обрані gemma-3-4b-it-qat, gemma-3-4b-it, qwen3-4b та llama-3.2-3b-instruct. Всі вони відносяться до LLM без аргументації та мають помірний об’єм, що допоможе збільшити швидкодію обчислень. Використовуючи кожен з них як агент моделі аналізу на тестових даних можна визначити, яка з них найкраще підходить під такий вид задачі.

ArenaNEW: OverviewArena (Vision)Arena-Hard-AutoFull Leaderboard

Total #models: 241. Total #votes: 3,007,425. Last updated: 2025-06-01.

Code to recreate leaderboard tables and plots in this [notebook](#). You can contribute your vote at [lmarena.ai](#)!

CategoryOverall

Apply filter

Style ControlShow Deprecated

Overall Questions

#models: 241 (100%) #votes: 3,007,425 (100%)

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	1	Gemini-2.5-Pro-Preview-05-06	1446	+7/-6	9503	Google	Proprietary	Unknown
2	1	o3-2025-04-16	1419	+5/-6	13133	OpenAI	Proprietary	Unknown
2	3	Gemini-2.5-Flash-Preview-05-20	1419	+6/-6	8669	Google	Proprietary	Unknown
2	3	ChatGPT-4o-latest (2025-03-26)	1415	+4/-5	17656	OpenAI	Proprietary	Unknown
2	7	Grok-3-Preview-02-24	1411	+4/-4	19977	xAI	Proprietary	Unknown
5	3	GPT-4.5-Preview	1404	+5/-6	15271	OpenAI	Proprietary	Unknown
7	7	Gemini-2.5-Flash-Preview-04-17	1393	+5/-6	12720	Google	Proprietary	Unknown
8	7	GPT-4.1-2025-04-14	1375	+6/-5	11773	OpenAI	Proprietary	Unknown
8	10	DeepSeek-V3-0324	1374	+4/-5	14408	DeepSeek	MIT	Unknown
8	4	Claude Opus 4 (20250514)	1366	+7/-8	7729	Anthropic	Proprietary	Unknown
8	15	Hunyuan-Turbos-20250416	1364	+7/-7	5230	Tencent	Proprietary	Unknown
10	10	DeepSeek-R1	1365	+4/-4	19430	DeepSeek	MIT	Unknown
10	9	o4-mini-2025-04-16	1355	+5/-5	11452	OpenAI	Proprietary	Unknown
10	42	Grok-Mini-beta	1355	+8/-7	4764	xAI	Proprietary	Unknown

Рисунок 2.3 – Бенчмарк порівняння LLM від HuggingFace [22]

Best Open LM

Model	Arena Elo	MMLU	License
DeepSeek-V3-0324	1374	88.5	MIT
DeepSeek-R1	1365	90.8	MIT
Qwen3-235B-A22B	1354	88.5	Apache 2.0
Gemma-3-27B-it	1349		Gemma

Full Leaderboard

Model	Arena Elo	Coding	Vision	Arena Hard	MMLU	Votes	Organization	License
Gemini-2.5-Pro-Preview-05-06	1446	1446	1355	96.4		9503	Google	Proprietary
Gemini-2.5-Flash-Preview-05-20	1419	1434	1308			8669	Google	Proprietary
o3-2025-04-16	1419	1430	1305			13133	OpenAI	Proprietary
ChatGPT-4o-latest (2025-03-26)	1415	1417	1310			17656	OpenAI	Proprietary
Grok-3-Preview-02-24	1411	1415			92.7	19977	xAI	Proprietary
GPT-4.5-Preview	1404	1404	1254			15271	OpenAI	Proprietary
Gemini-2.0-Pro-Exp-02-05	1387	1382	1239			20120	Google	Proprietary
Gemini-2.0-Flash-Thinking-Exp-01-21	1387	1367	1276			27618	Google	Proprietary
GPT-4.1-2025-04-14	1375	1380	1279			11773	OpenAI	Proprietary
DeepSeek-V3-0324	1374	1387		85.5	88.5	14408	DeepSeek	MIT
Claude Opus 4 (20250514)	1366	1402	1233			7729	Anthropic	Proprietary
DeepSeek-R1	1365	1367		93.2	90.8	19430	DeepSeek	MIT

Рисунок 2.4 – Бенчмарк порівняння LLM від OpenLM [23]

Мовні моделі запускалися локально через LM Studio, що була обрана за її зручність та можливість запуску без потреби в хмарній інфраструктурі [24]. Це дозволило забезпечити автономність системи та уникнути обмежень API-запитів до сторонніх сервісів. Використання LM Studio як середовища для локального запуску моделей забезпечило стабільне підключення до LLM через OpenAI-сумісний API, що значно спростило її інтеграцію.

Для збору даних з відкритих джерел використовувався Selenium, що дозволив автоматизувати процес отримання динамічного контенту з новинних сайтів. Дані, які отримувались через Selenium, передавались у Pandas для первинної обробки: очищення, нормалізації текстів і підготовки до подальшої аналітики. Pandas забезпечив зручний інтерфейс для маніпулювання табличними даними, що є ключовим у підготовці корпусу новин для аналізу.

Збереження та організація даних здійснювалась у реляційній базі даних MySQL, що є стабільним та перевіреним рішенням для зберігання структурованої інформації. Для взаємодії з базою використовувався ORM-фреймворк SQLAlchemy, який значно спростив роботу з SQL-запитами, забезпечивши при цьому гнучкість у побудові моделей даних.

Розгортання серверної логіки було реалізовано за допомогою хмарної платформи Railway, що надала швидкий спосіб публікації backend-частини, забезпечила автоматичне розгортання з GitHub та постійну доступність API. Це дало змогу зосередитися на розробці логіки, а не на конфігурації серверного середовища.

Для візуалізації результатів аналізу була розроблена фронтендна частина на React Native з використанням бібліотеки Vite, яка виконувала роль віджету. Такий підхід дозволив створити кросплатформений інтерфейс, що може бути вбудований у більші системи або мобільні застосунки. React Native дав змогу реалізувати інтуїтивно зрозумілий інтерфейс, який показує результати аналізу у зручному для користувача вигляді.

В якості основного середовища розробки було обрано Visual Studio Code, яке забезпечило зручну інтеграцію з Git, підтримку форматування коду, автодоповнення та можливість одночасної роботи з кількома технологіями

завдяки розширенням. Це значно підвищило ефективність роботи над проєктом, особливо у фазі інтенсивної розробки та налагодження.

Таким чином, обраний стек технологій був спрямований на досягнення балансу між продуктивністю, модульністю, швидкістю розробки та гнучкістю налаштувань. Усі інструменти були підібрані з урахуванням специфіки задачі, а також з можливістю майбутнього масштабування та адаптації системи до нових вимог.

2.4 Особливості налаштування моделі аналізу настрою

Налаштування інформаційної системи розпочинається з побудови принципу багатокрокового конвеєра аналізу, який забезпечує динамічне формування та оновлення контексту при роботі з LLM. На початковому етапі до системи подається структурований чанк тексту статті разом із метаданими (дата, джерело, тема обговорення). Для кожного етапу обробки формується власний контекст: спочатку до промпту входить лише текстова частина, надалі – витягнені дані подій чи висловлювань, історичні дані висловлювань з бази даних, а на завершальних кроках враховуються попередні оцінки настрою та виявлені суперечності. Такий підхід дозволяє уникнути перевантаження початкового контексту та водночас гарантовано передати до LLM всі необхідні дані на кожному з етапів [19]. Також це дозволяє покращити якість аналізу, оскільки на кожному етапі відбувається цілеспрямоване визначення одного показника, за яким формується висновок, який потім використовується на наступних етапах.

Перша логічна взаємодія з LLM-агентом спрямовується на класифікацію настрою, де до промпту додаються інструкції з оцінювання тональності висловлювання до певної теми. Отриману відповідь у вигляді «сирого» значення тональності автоматично коригують за допомогою алгоритму нормалізації, що перетворює відносну оцінку в уніфікований числовий діапазон.

Далі система здійснює вбудований SQL-запит до сховища історичних даних, витягуючи минулі оцінки тієї ж особи за обраний проміжок часу, за

умови, що вони присутні. На сформовану послідовності значень застосовується формула, яка зважає кожен позицію з урахуванням давності: найсвіжіші висловлювання отримують більші коефіцієнти, а більш віддалені в часі – прогресивно зменшені.

Наступна взаємодія з LLM-агентом орієнтована на виявлення суперечностей. На цьому кроці відбувається фільтрування, для аналізу суперечностей обираються лише актуальні дані, які мають суттєву відмінність у позиції від поточної. До промπτу включається перелік відфільтрованих оцінок настрою з прив'язкою до конкретних дат і цитат. Якщо протиріччя виявлено, це відповідно фіксується, а на основі цитат формується коментар, що пояснює у чому полягає відмінність. Якщо ж історичних даних немає або протиріччя не виявлено, модель пропускає цей крок.

У завершальному логічному етапі конвеєра відбувається агрегування всіх проміжних результатів – нормалізованої оцінки настрою, маркерів суперечностей і вагових коефіцієнтів, та формування єдиного об'єкта для візуалізації. Тут формуються висновки за особою та оцінкою його висловлювань, разом із фоновими подіями. Одразу після цього інформація записується в базу даних із позначенням рівня повноти даних: якщо для запису всіх полів ще не вистачає даних (наприклад, очікується результуюча оцінка), система записує ці дані, проте до візуалізації вони не потрапляють. Завдяки такому прогресивному запису забезпечується гнучкість у конвеєрі обробки: навіть за відмови окремих кроків дані не втрачаються, а накопичуються та доповнюються за першої можливості.

Важливим етапом налаштування є підбір конфігурацій даної інтелектуальної системи аналізу. Це включає в себе визначення мінімального рівня цікавості висловлювання чи події задля включення її у візуалізацію, визначення періоду актуальності й мінімальної міри різниці суперечливих висловлювань, обрання формули для зважування актуальності подій, визначення коефіцієнтів зваженої суми оцінок для формування «цікавості» та створення шкал для суперечливості та позиції щодо теми.

Задля фіналізації налаштування моделі визначення настрою щодо теми потрібно обрати найкращу LLM для даної задачі. Це включає у себе проведення експертного оцінювання на обраній вибірці даних та порівняльний аналіз результатів обробки цих даних системою з подальшим визначенням моделі з найкращим результатом. Критерієм ефективності є максимізація точності (f1-score) відповіді системи від узагальненої експертної оцінки. Додатково визначаються якість згенерованого тексту за рішенням експертів та оцінки precision, recall та accuracy. Таким чином, модель, що досягає найбільших оцінок та має найвищу якість тексту буде задовольняти умову оптимальності – найбільшу можливу точність прогнозу системи. Варто зазначити, що підбір конкретних параметрів та визначення їх доцільності можливий лише після реалізації всього конвеєру обробки.

Таким чином, особливості налаштування системи побудовано на комбінованому використанні динамічного контексту, багатокрокової обробки з LLM-агентом, вбудованих запитів до історичних даних і прогресивного запису результатів, що разом гарантує точність, гнучкість і стійкість процесу аналізу публічного ставлення.

2.5 Висновки

У даному розділі проведено відбір оптимальних методів розв'язання поставленої задачі, використовуючи сформований концепт і конкретні вимоги до моделі аналізу настрою публічних осіб. Це включало визначення технології інтелектуального аналізу тексту з метою отримання висновків щодо позиції особи за досліджуваною темою та зазначення способу досягнення необхідних результатів з її використанням. Також визначено конкретні інформаційні технології для роботи. Основою інтелектуальної частини став фреймворк LangGraph та велика мовна модель, що запускала локально через LM Studio. У даному розділі також відібрано найкращі LLM для виконання поставленої задачі, що будуть використані як гіперпараметр моделі аналізу настрою. Візуалізаційна частина реалізована за допомогою React Native та бібліотеки Vite.

3 МОДЕЛЮВАННЯ ТА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ДЛЯ ОЦІНЮВАННЯ ЗМІН У РИТОРИЦІ ПУБЛІЧНИХ ОСІБ

3.1 Моделювання змін у риториці публічних осіб

Модель підсистеми визначення сентименту побудована на основі архітектури LangGraph – спеціального фреймворка для організації послідовного оброблення тексту за допомогою великих мовних моделей у вигляді спрямованого ациклічного графа станів [21]. LLM у створеній системі відіграє роль агента для виконання дій, тому її можна змінювати й таким чином оптимізувати оброблення.

На вході кожного виконання графа формується об'єкт `ChunkState`, який містить у собі унікальний ідентифікатор чанку, текстовий фрагмент, розмір якого налаштований під обмеження контексту LLM (зазвичай близько 2500 символів), а також накопичувальні поля для результатів попередніх кроків [19]. Також у ньому зазначаються певні метадані, що важливі для подальшого формування контексту. Варто зазначити, що під час виконання кроків графу вхідні дані (рис. 3.1) розподіляються на кілька потоків паралельної асинхронної обробки (зазвичай 4), кожен з яких одночасно обробляє один блок тексту. Сам граф описується як низка вузлів (нод), кожен із яких відповідає певному етапу аналітики, та ребер, які визначають напрямки потоків обробки залежно від наявних у стані даних (рис. 3.2).

	chunk_id	content	topic	topic_id	article_date	origin_article_id
0	145	A new blueprint from the Biden administration ...	Green energy	1	2021-09-08	45
1	146	What's in the Democrats' \$3.5 trillion spendin...	Green energy	1	2021-09-08	45
2	147	As the climate crisis escalates, the world is ...	Green energy	1	2021-08-31	46
3	148	.Operating a plane or a large ship, for instan...	Green energy	1	2021-08-31	46
4	149	. Methane, a powerful greenhouse gas and the m...	Green energy	1	2021-08-31	46
5	150	." Jackson is also the CEO and founder of Prot...	Green energy	1	2021-08-31	46
6	151	Scientists say this invisible gas could seal o...	Green energy	1	2021-08-31	46
7	152	ESG investing — funds that are conscious of co...	Green energy	1	2021-08-16	47
8	153	. Apple's manufacturing contributed more than ...	Green energy	1	2021-08-16	47
9	154	."Trying to find the perfect company is impos...	Green energy	1	2021-08-16	47

Рисунок 3.1 – Вигляд вхідних даних, кожен рядок формує новий `ChunkState`

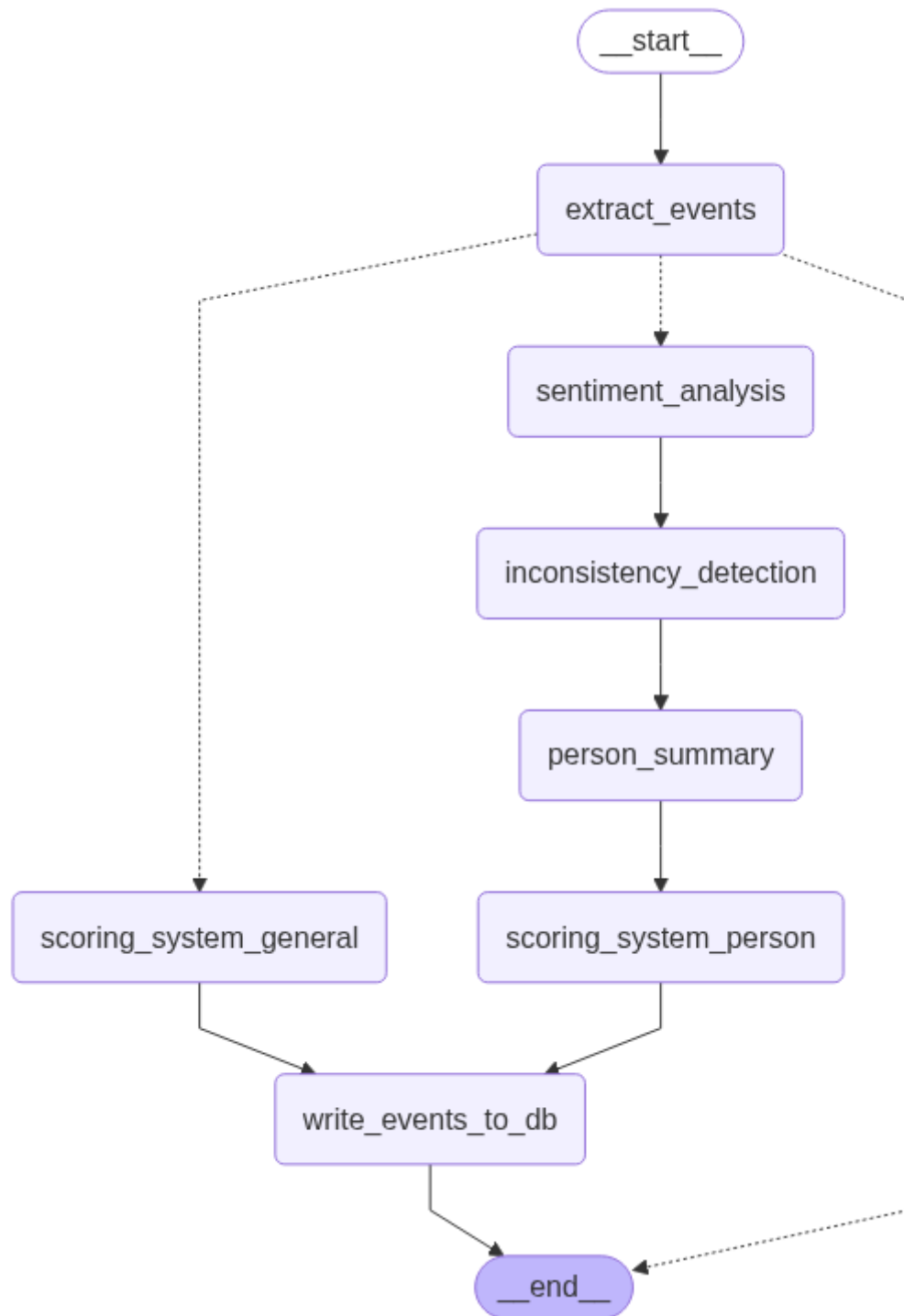


Рисунок 3.2 – Модель графу аналізу

Першим кроком у графі є нода `extract_events`, яка відповідає за базове розпізнавання наявності в чанку подій загального характеру (`general_event`) та заяв публічних осіб (`person_event`). Використовуючи потужності мовної моделі, ця нода аналізує текстовий фрагмент і намагається виокремити дві різні структури даних: для загальних подій – фонову подію із заголовком та коротким

описом, а для висловлювань – відповідно ім'я особи та цитату. У разі успішного виявлення подій кожного типу відбувається відповідна маршрутизація на дві асинхронні гілки графу – обробку подій та обробку висловлювань. Приклад подібного стану наведено на рисунку 3.3.

```
{
  "chunk_id": 152,
  "topic": "Green energy",
  "topic_id": 1,
  "origin_article_id": 47,
  "article_date": "2021-08-16 00:00:00",
  "content": "ESG investing \u2014 funds that are conscious of companies\u2019 actions on the environment, society and governance \u2014 is an increasingly popular trend for people concerned about how their money is being used. It sounds great, but it\u2019s complicated. The ESG label is meant to make investing green easier by giving investors a simple way to allocate their money to good causes. And, indeed, there are plenty of sustainability- and climate change-focused exchange traded funds available. But it\u2019s still not the catch-all stamp of approval you\u2019d expect. The label started with the idea that ESG issues should be included when valuing a business. More recently, however, the emphasis has moved to the impact company\u2019s products and services have, said Jon Hale, director of sustainability research at Sustainalytics. \u201cA lot of ESG funds are oriented more toward the ESG valuations rather than the impact,\u201d he told CNN Business. That\u2019s why the blanket ESG designation might not mean all companies in an ETF are up-to-snuff on all fronts. \u201cYou can\u2019t tell just from the label. You actually have to figure out what they\u2019re doing,\u201d said Hale. That essentially forces investors into tough choices between the E, the S and the G. Invest in one, and you often have to sacrifice another. This puts climate and socially conscious investors in a tough spot. For example, ESG ETFs stay away from weapons manufacturers, tobacco companies or firms in the coal and oil sands businesses but still give investors access to the controversial golden geese of America\u2019s stock market: tech stocks. Big Tech companies like Apple (AAPL) and Amazon (AMZN) make up a big chunk of these funds. They\u2019re included because both of these companies have made commitments to run net-zero carbon supply chains and operations in the next decades. That doesn\u2019t mean they\u2019re carbon-neutral now. Amazon\u2019s ultra-fast delivery options come at an environmental cost, for example. Last year, when online shopping went through the roof as people tried to avoid exposure to Covid-19, Amazon\u2019s carbon emissions grew by 19% \u2014 even though the business reduced its overall carbon intensity. As for Apple, though the company is working on getting the carbon intensity out of its mostly Asia-based supply chain, it\u2019s not there yet. Apple\u2019s manufacturing contributed more than two-thirds of its carbon footprint last year, according to the company\u2019s environmental progress report",
  "general_event": [
    {
      "title": "ESG Funds' Valuation Focus",
      "description": "Shift in ESG funds prioritizing business valuations over impact on products and services."
    }
  ],
  "person_event": [
    {
      "person_name": "Jon Hale, Sustainalytics",
      "citation": "Hale stated that many ESG funds focus on ESG valuations rather than the actual impact of companies\u2019 products and services. Investors need to investigate beyond the label."
    }
  ],
  "is_processed": 0
}
```

Рисунок 3.3 – Приклад визначених загальних подій та висловлювань

У разі наявності висловлювання (`person_event`) система переходить до специфічного аналізу сентименту (`sentiment_analysis`). Вона компонує промпт до LLM із переданим контекстом: ім'я особи, текст цитати та сама цитата. Модель повертає числовий `sentiment_score` в діапазоні від -1 до $+1$, який означає негативне або позитивне ставлення. Результат миттєво зберігається в полях стану, а при виникненні помилки або невизначеності система може повторити запит або пропустити фрагмент як невалідний. Приклад стану на даному етапі наведено на рисунку 3.4.

```

"person_event": [
  {
    "person_name": "Jon Hale, Sustainalytics",
    "citation": "Hale stated that many ESG funds focus on ESG valuations
rather than the actual impact of companies\u2019 products and
services. Investors need to investigate beyond the label.",
    "sentiment": -0.6
  }
],

```

Рисунок 3.4 – Стан після проходження sentiment_analysis (частина з висловлюваннями)

Далі у процесі обробки висловлювання викликається нода виявлення суперечностей (inconsistency_detection). На цьому етапі нова зафіксована цитата порівнюється з історією висловлювань тієї ж особи, отриманою через захищений інтерфейс зв'язку із базою даних на основі SQLAlchemy. Проаналізувавши зважені за актуальністю минулі сентименти певної особи до теми визначається, чи присутня у них значуща нещодавня суперечливість. Якщо виявлено розбіжність, у стан додаються поля inconsistency_flag, список ідентифікаторів попередніх заяв, що протирічать, а також короткий коментар-інтерпретація (inconsistency_comment). Приклад стану після виявлення у висловлюванні суперечливості наведено на рисунку 3.5.

```

"person_event": [
  {
    "person_name": "Fatih Birol (IEA)",
    "citation": "The IEA\u2019s executive director warned that accelerating clean energy innovation is \u2018critical\u2019 for achieving net-zero targets by 2050, stating the current pace is insufficient to meet global climate goals.",
    "sentiment": 0.6,
    "person_id": 118,
    "inconsistency_with_id": [
      121,
      193,
      203
    ],
    "inconsistency_flag": true,
    "inconsistency_comment": "The new citation emphasizes urgency for innovation; previous statements focused on investment needs and technology inclusion, lacking a similar sense of immediate critical action."
  }
]

```

Рисунок 3.5 – Стан після виявлення протиріччя (частина із висловлюваннями)

Наступною в ланцюжку обробки йде нода `person_summary`. Вона агрегує отримані дані про всі цитати особи в межах одного чанку та вичитує їхній загальний контекст. На вхід вона отримує всю попередньо аналізовану інформацію. Таким чином модель має змогу чітко розуміти конкретний випадок. За допомогою LLM генерується коротка аналітична виписка (`person_summary`), яка узагальнює позицію особи: основний наратив, можливі суперечності та напрямок зміни ставлення. Цей текст стає зручним для людини-читача, оскільки подає висновки у стислому, але змістовному вигляді, у який входять короткі відомості про особу, його загальна позиція, схильність змінювати свої думки та наявність нещодавніх суперечливих висловлювань. Приклад стану після проходження цього етапу наведено на рисунку 3.6.

```

"person_event": [
  {
    "person_name": "Jon Hale, Sustainalytics",
    "citation": "Hale stated that many ESG funds focus on ESG valuations
rather than the actual impact of companies\u2019 products and
services. Investors need to investigate beyond the label.",
    "sentiment": -0.6,
    "person_id": 708,
    "inconsistency_flag": false,
    "sentiment_deviation": 0.11579900522888784,
    "stance": "against",
    "person_summary": "Jon Hale, Sustainalytics, now argues against ESG
funds prioritizing valuations over actual impact. He emphasizes
investor due diligence beyond labels, reflecting a medium openness
to change and aligning with recent concerns about greenwashing."
  }
],

```

Рисунок 3.6 – Стан після проходження person_summary (частина із висловлюваннями)

Нарешті, гілка обробки висловлювань завершується виконанням ноди `scoring_system_person`, яка об'єднує весь отриманий раніше контекст та передає до мовної моделі. LLM, у свою чергу осмисливши отримані дані, дає свою оцінку конкретному висловлюванню за такими категоріями: релевантність до теми, внесок до теми обговорення та суперечливість цитати. Оскільки оцінювання від мовної моделі не є ідеальним, на основі нього формується інтегральна оцінка «цікавості» висловлювання, що буде використана для відокремлення найцікавіших висловлювань.

Обробка загальних подій, якщо такі виявлено, відбувається простіше. Після кроку `extract_events` отриманий стан із присутніми загальними подіями передається до вузла `scoring_system_general`, що подібним чином оцінює цікавість фонові події. Оскільки фонові події є радше довідниковими, вони оцінюються лише за їх вмістом та відношенням до теми у конкретному блоці тексту. За цим контекстом мовна модель аналогічно формує оцінки – релевантність, вплив та новизну, з яких потім формується показник цікавості.

Після того, як обидві гілки графу завершують роботу, вони зливають свої накопичені стани. Приклад фінального стану наведено на рисунку 3.7.

Отриманий повний стан далі передається до ноди `write_events_to_db`, яка зберігає його у відповідних таблицях бази даних, забезпечуючи етапність відповідно до повноти даних. Таким чином досягається безпечність запису у базу даних та контроль над помилками оброблення. Внесена інформація є достатньою для використання у необхідній візуалізації.

```
{
  "chunk_id": 152,
  "topic": "Green energy",
  "topic_id": 1,
  "origin_article_id": 47,
  "article_date": "2021-08-16 00:00:00",
  "content": "...",
  "general_event": [
    {
      "title": "ESG Funds' Valuation Focus",
      "description": "Shift in ESG funds prioritizing business valuations over impact on products and services.",
      "relevance_score": 0.85,
      "influence_score": 0.7,
      "novelty_score": 0.6,
      "event_hotness": 0.6875
    }
  ],
  "person_event": [
    {
      "person_name": "Jon Hale, Sustainalytics",
      "citation": "Hale stated that many ESG funds focus on ESG valuations rather than the actual impact of companies\u2019 products and services. Investors need to investigate beyond the label.",
      "sentiment": -0.6,
      "person_id": 708,
      "inconsistency_flag": false,
      "sentiment_deviation": 0.11579900522888784,
      "stance": "against",
      "person_summary": "Jon Hale, Sustainalytics, now argues against ESG funds prioritizing valuations over actual impact. He emphasizes investor due diligence beyond labels, reflecting a medium openness to change and aligning with recent concerns about greenwashing.",
      "relevancy_score": 0.9,
      "contribution_score": 0.8,
      "controversy_score": 0.7,
      "opinion_hotness": 0.81
    }
  ],
  "is_processed": 0
}
```

Рисунок 3.7 – Приклад фінального повного стану (зміст блоку тексту опущено для зручності)

Завдяки такому поетапному підходу з чітким розподілом відповідальності між нодами забезпечується висока гнучкість та обчислювальна ефективність

системи: можна легко додавати нові етапи обробки, змінювати алгоритми класифікації та розширювати контекст, не порушуючи глобальну структуру графа станів, допоки на наступний крок обробки передається коректний стан. Це дозволяє згодом масштабувати підсистему та адаптувати її під інші тематики або джерела новин.

3.2 Підбір оптимальних налаштувань моделі аналізу настрою

Як було зазначено вище, модель аналізу настрою інформаційної системи інтелектуального аналізу настрою на даному етапі є лише конвеєром із кроків аналізу вхідних даних заданого формату. Для досягнення оптимального результату роботи готової моделі потрібно підібрати такі налаштування гіперпараметрів, які максимізують критерій точності отриманих результатів. Для цього потрібно передусім досягти точності прогнозування та покращити якість текстових результатів аналізу.

Налаштування гіперпараметрів моделі інтелектуальної системи аналізу умовно можна поділити на дві частини. Першою частиною є підбір формул для обрахунку вже отриманих показників. Насамперед потрібно обрати спосіб зважування подій за актуальністю та максимальний вік висловлювання, що ще вважається актуальним для аналізу суперечностей. Експериментально було визначено, що лінійне спадання актуальності висловлювань за часом та максимальний період незмінності думки у 1 рік є найближчими до істини та надають контекст потрібного розміру у більшій частині випадків. Ці показники були обрані через підбір й перевірку кількості та якості отриманих відповідей із бази даних. Аналогічно проводячи тестові вибірки із бази даних, було також встановлено значення мінімальної різниці у настрої для аналізу суперечності та пороги цікавості події й висловлювання в подальшому.

Для визначення наступних параметрів було створено тестову вибірку 50 актуальних статей з однієї теми, висловлювання та події яких оцінено за настроєм, суперечливістю, релевантністю, внеском та «цікавістю». Було запроваджено певні оцінки, так настрій до теми визначався як: -2 –

«негативний», -1 – «здебільшого негативний», 0 – «нейтральний чи змішаний», 1 – «здебільшого позитивний» та 2 – «позитивний». Інші ж оцінки призначались за 5-бальною шкалою, де 0 – найменше можливе значення, а 5 – найбільше. Середню експертну оцінку кожної статті за цими критеріями було записано у таблицю та використано як правильний результат аналізу (рис. 3.8).

Article_id	Expert				
	sentiment (-2/+2)	controversy (0-5)	relevancy (0-5)	contribution (0-5)	hotness (0-5)
1	2	0	2	2	4
2	1	0	5	5	3
3	0	2	3	3	3
4	-2	1	5	5	4
5	1	0	1	2	2
6	-1	4	5	5	5
7	2	0	4	2	4
8	2	0	5	5	4
9	0	2	3	5	5
11	0	4	0	0	1
12	0	1	0	2	1
13	0	3	3	3	4
14	1	4	3	4	4
15	2	1	4	5	5
16	0	4	5	4	5

Рисунок 3.8 – Результати експертної оцінювання тестової вибірки (частина)

Коефіцієнти зваженої суми для визначення цікавості висловлювання було обрано шляхом визначення важливості параметрів лінійної регресії прогнозування «цікавості» у експертних оцінках. Звідси, на «цікавість» новини найбільше впливає релевантність (68,19%), суперечливість (13,928%) та внесок (10,861%). Ці показники у подальшому адаптовані до результатів оцінювання від обраної мовної моделі з метою досягнення максимально подібного значення «цікавості». Як додаткова міра подальшого аналізу було створено шкалу настрою для призначення до класу за позицією та шкалу мору суперечливості людини. Основа цих мір – це базові статистичні поняття діапазону та

середньоквадратичного відхилення, а джерелом даних для них при цьому є зважений показник сентименту.

За проміжними результатами роботи підсистеми визначення сентименту було проаналізовано розподіл результируючих показників цікавості подій та висловлювань. Таким чином були зазначені межові значення для відображення елемента на візуалізації. Для висловлювань (рис. 3.9) ним стало значення 0.665, все що отримало оцінку більше мало право на відображення, при чому візуально розмежовувалось за цікавістю у відрізках між 0.698, 0.798, 0.832 та 0.96. Загальні ж події (рис. 3.10) не потребують такої деталізації на категорії, тому було визначене лише межеве значення, яким є 0.6963.

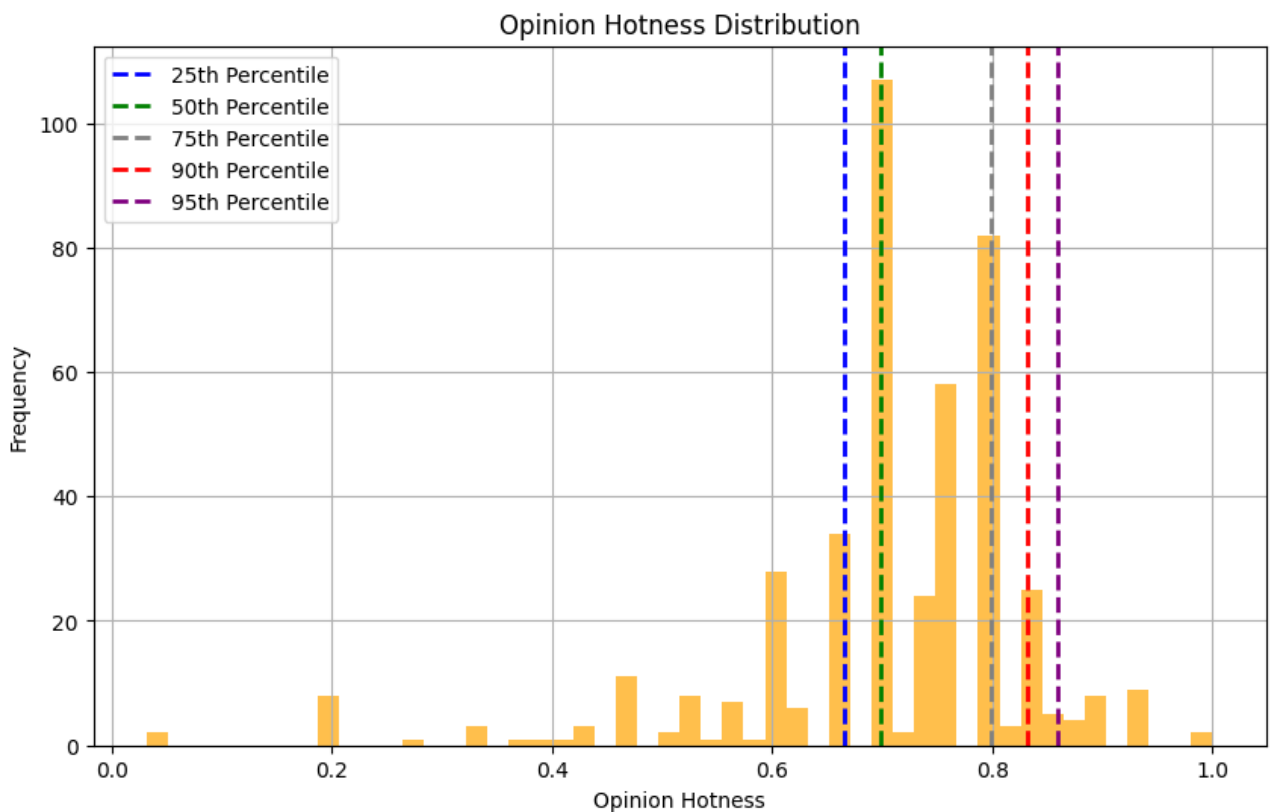


Рисунок 3.9 – Розподіл оцінок цікавості висловлювань

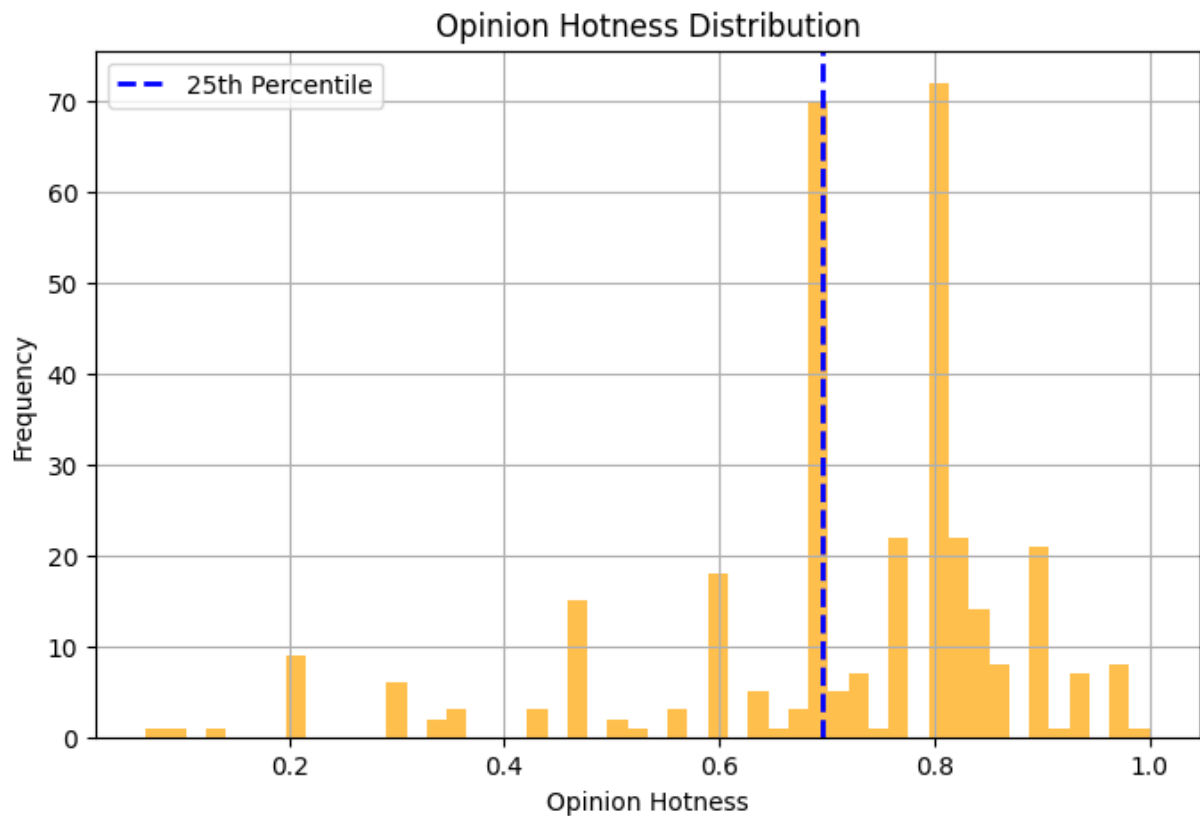


Рисунок 3.10 – Розподіл оцінок цікавості подій

Таким чином визначено, що використання описаних налаштувань надає найбільший контекст подальшому використанню великих мовних моделей. Параметри витягнення та фільтрування минулих записів забезпечують достатній та досить актуальний контекст аналізу, сформовані оцінки обґрунтовано формують цікавість подій та висловлювань, обрані межові значення цікавості залишають помірну кількість найкращих тверджень. Проведене експертне оцінювання вибірки даних, що була здійснене для підтвердження формування оцінок, буде використане у подальших кроках для порівняння великих мовних моделей.

3.3 Вибір оптимальної великої мовної моделі для вирішення поставленої задачі

Продовжуючи налаштування моделі системного інтелектуального аналізу ставлення осіб до теми із відстежуванням протиріч було проведено порівняння

ефективності використання різних гіперпараметрів, а саме – підбір LLM-агентів, що найбільше підходять для виконання подібних задач. Із великих мовних моделей з відкритими вагами було обрано ті, що зайняли найвищі показники на тестувальних бенчмарках [22,23]. Додатковими вимогами були: орієнтація на інтелектуальну обробку тексту, а не генерацію коду, принцип роботи без блоку аргументації, через часові затрати на виконання, та наявність розміру, що забезпечує потрібну інтелектуальність при відносній швидкодії. За результатами були обрані gemma-3-4b-it-qat, gemma-3-4b-it, qwen3-4b та llama-3.2-3b-instruct. Використовуючи кожен з цих великих мовних моделей у якості гіперпараметру моделі аналізу настрою було опрацьовано перед цим оцінені дані вибірки із 50 статей (рис. 3.8).

За принципом здійсненого перед цим експертного оцінювання висловлювань результати було групувано за новиною статтею, після цього оцінено загальний настрій, суперечливість, релевантність, внесок та «цікавість». Додатково, зважаючи на необхідність оцінити не лише точність аналізу, але і його якість, було додано показники кількості думок на статтю та 10-бальну експертну оцінку якості тексту.

Перша модель, gemma-3-4b-it-qat отримала наступні результати (рис. 3.11). За якістю тексту, зважаючи на кількість отриманих висловлювань, ця модель має оцінку 9.04 за 10-бальною шкалою. При цьому на кожну статтю в середньому приходить 3.09 висловлювань. За оцінками влучності (precision) ця LLM має показник 0.69, повноти (recall) 0.67, зважений f1-score 0.68. Також, якщо перевести настрій за визначеною перед цим шкалою, можна обрахувати коефіцієнт детермінації (r^2 score), що у даному випадку становить 0.69. На даному етапі видно, що модель гарно генерує текстову частину та достатньо точно визначає клас настрою.

Article_id	gemma-3-4b-it-qat						
	n_opins	text_quality (1-10)	sentiment (-2/+2)	controversy (0-1)	relevancy (0-1)	contribution (0-1)	hotness (0-1)
1	2	5	1	0,3	1	0,8	0,8
2	3	9	0	0,6	0,9	0,8	0,75
3	3	10	1	0,4	0,75	0,7	0,6
4	5	10	-2	0,7	0,88	0,78	0,77
5	3	7	0	0,3	0,9	0,767	0,613
6	9	8	-1	0,744	0,922	0,833	0,82
7	2	7	1	0,5	0,9	0,8	0,705
8	2	9	1	0,65	0,925	0,825	0,78
9	4	9	-1	0,638	0,862	0,762	0,738
11	2	8	0	0,3	0,1	0,15	0,198
12	4	10	0	0,175	0,4	0,325	0,284
13	5	10	-1	0,6	0,88	0,78	0,733
14	6	9	1	0,633	0,833	0,717	0,712
15	3	10	2	0,6	0,917	0,85	0,767
16	6	10	1	0,567	0,875	0,758	0,711
17	2	10	0	0,65	0,925	0,825	0,78
18	3	9	2	0,467	0,967	0,9	0,743
19	4	8	-2	0,663	0,862	0,737	0,739
20	4	8	2	0,45	0,862	0,788	0,671
23	2	10	0	0,3	0,75	0,675	0,544
24	3	10	1	0,3	0,833	0,717	0,579

Рисунок 3.11 – Результати роботи моделі gemma-3-4b-it-qat (частина)

Для додаткового визначення поведінки системи за таких налаштувань складемо матрицю невідповідностей (рис. 3.12) та графік ROC AUC (рис. 3.13).

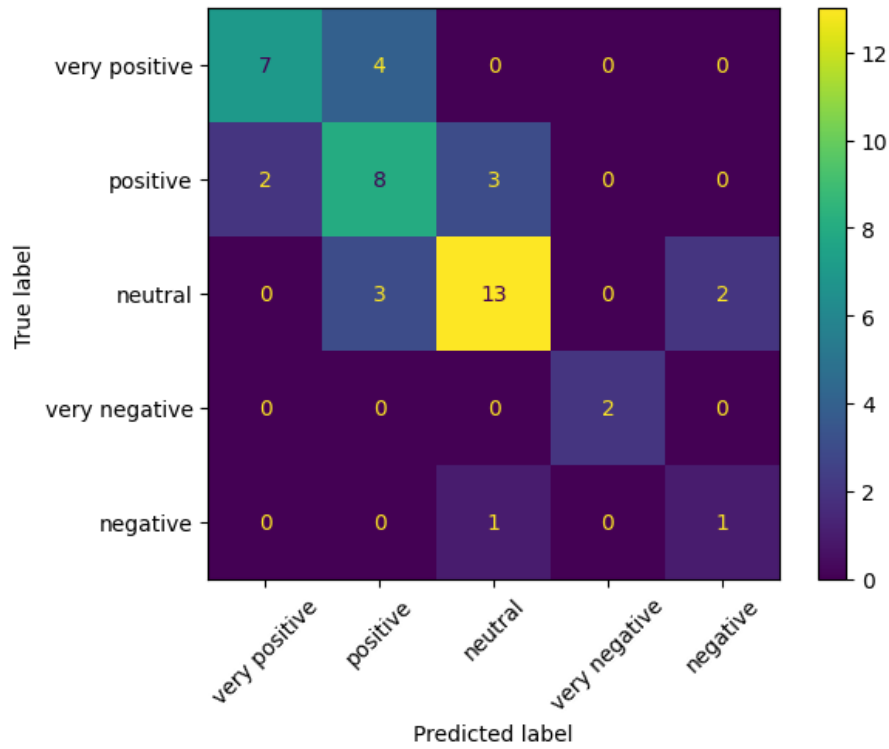


Рисунок 3.12 – Матриця невідповідностей за результатами роботи gemma-3-4b-it-qat

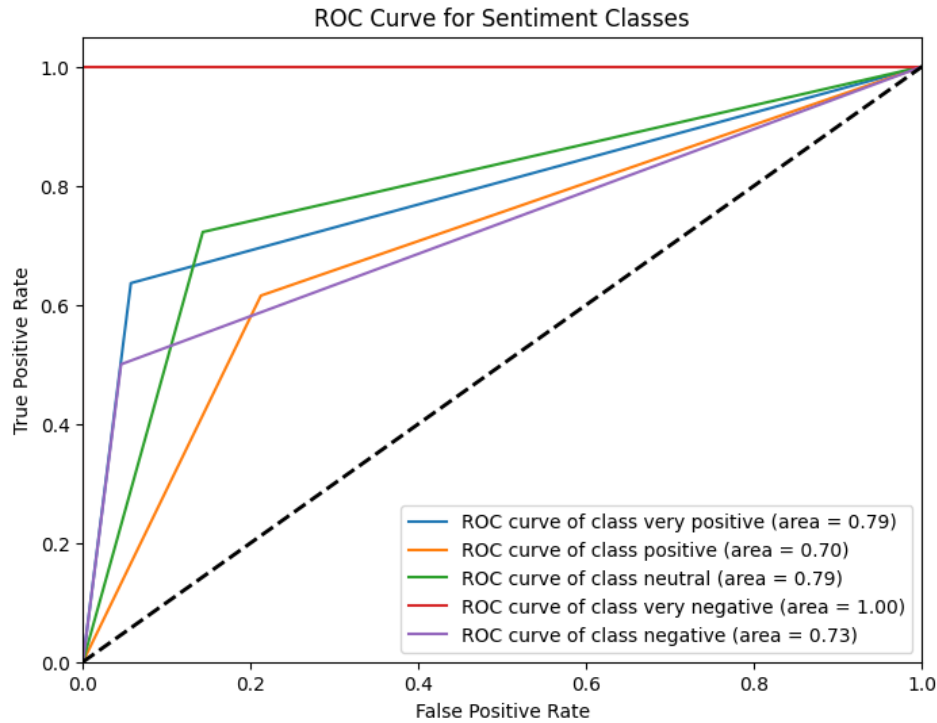


Рисунок 3.13 – ROC AUC за результатами роботи gemma-3-4b-it-qat

Використовуючи ці візуалізації стає зрозуміло, що за даних налаштувань модель найкраще визначає строго негативні позиції, але непогано справляється і

з визначенням інших. Також видно, якщо і відбувається помилка судження, прогнозоване значення, як правило, займає сусідній клас, що у випадку шкали не є настільки критичним як просте невгадування.

Друга модель, gemma-3-4b-it була оцінена аналогічно (рис. 3.14). За якістю тексту, вона має оцінку 8.82 за 10-бальною шкалою. При цьому на кожну статтю в середньому також приходиться 3.09 висловлювань. За оцінками влучності (precision) ця LLM має показник 0.67, повноти (recall) 0.62, f1-score 0.63. За описаним раніше принципом, обрахований коефіцієнт детермінації (r2 score) становить 0.34. Видно, що якість тексту помітно впала, також гіршою є і точність порівняно із версією qat. При цьому низький коефіцієнт детермінації свідчить про те, що попри непогані результати точності, модель аналізу не розуміє розподілу сентименту.

Article_id	gemma-3-4b-it						
	n_opins	text_quality (1-10)	sentiment (-2/+2)	controversy (0-1)	relevancy (0-1)	contribution (0-1)	hotness (0-1)
1	2	10	0	0,3	0,75	0,65	0,535
2	3	6	0	0,633	0,883	0,883	0,783
3	3	9	0	0,483	0,85	0,75	0,668
4	5	9	-2	0,79	0,92	0,86	0,847
5	3	9	1	0,367	0,9	0,8	0,652
6	9	9	-1	0,644	0,883	0,767	0,747
7	2	9	0	0,75	0,9	0,8	0,805
8	2	9	1	0,5	0,95	0,85	0,735
9	4	9	0	0,6	0,912	0,812	0,752
11	2	7	-1	0,4	0,45	0,35	0,395
12	4	9	-1	0,35	0,45	0,375	0,384
13	5	6	0	0,6	0,96	0,85	0,745
14	6	10	1	0,592	0,9	0,792	0,739
15	3	10	2	0,533	0,917	0,817	0,728
16	6	10	0	0,542	0,9	0,783	0,716
17	2	10	0	0,65	0,875	0,725	0,723

Рисунок 3.14 – Результати роботи моделі gemma-3-4b-it (частина)

Аналогічно побудуємо матрицю невідповідностей (рис. 3.15) та графік ROC AUC (рис. 3.16).

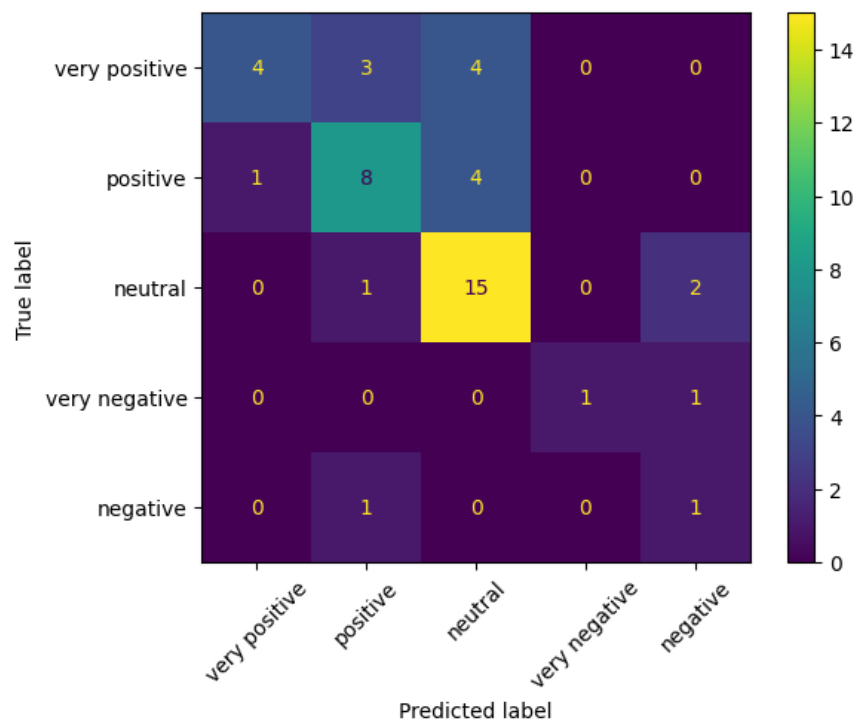


Рисунок 3.15 – Матриця невідповідностей за результатами роботи gemma-3-4b-it

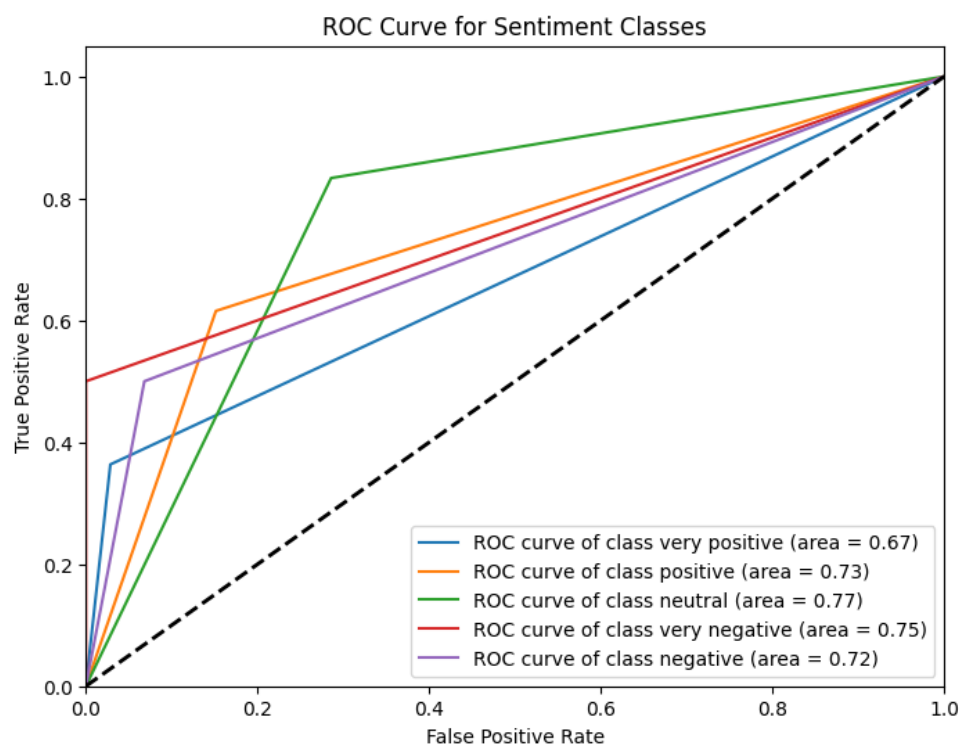


Рисунок 3.16 – ROC AUC за результатами роботи gemma-3-4b-it

За графіками аналізу роботи gemma-3-4b-it особливих залежностей не видно, середня якість інтелектуального аналізу цієї моделі загалом менша.

Далі перевіримо та оцінимо llama-3.2-3b-instruct (рис. 3.17). За якістю тексту, вона має оцінку 5.15 за 10-бальною шкалою. При цьому на кожну статтю в середньому також припадає 3.07 висловлювань. За оцінками влучності (precision) ця LLM має показник 0.45, повноти (recall) 0.38, f1-score 0.37. За описаним раніше принципом, обрахований коефіцієнт детермінації (r2 score) становить -0.56. Дана модель продемонструвала жахливі результати, як у якості тексту, так і точності оцінок. Коренем проблеми є генерація неправильного тексту, який потім не міг бути правильно оціненим. На даному етапі вже можна зробити висновок, що ця LLM не підходить для виконання даної задачі.

Article_id	llama-3.2-3b-instruct						
	n_opins	text_quality (1-10)	sentiment (-2/+2)	controversy (0-1)	relevancy (0-1)	contribution (0-1)	hotness (0-1)
1							
2	1	10	1	0,8	0,9	1	0,895
3	3	9	0	0,45	0,85	0,9	0,708
4	5	8	1	0,22	0,9	0,9	0,628
5	3	6	1	0,133	0,867	0,533	0,457
6	9	5	2	0,211	0,933	0,933	0,644
7	2	6	2	0,45	0,9	0,85	0,702
8	2	6	2	0,25	1	1	0,7
9	4	5	0	0,325	0,925	0,825	0,65
11	2	5	1	0,65	0,9	0,6	0,695

Рисунок 3.17 – Результати роботи моделі llama-3.2-3b-instruct (частина)

Наприкінці перевіримо qwen3-4b (рис. 3.18). Варто зауважити, що ця модель за замовчуванням є аргументаційною та спроектованою для використання як інструмент, проте її також можна використовувати для звичайної генерації. За якістю тексту, вона має оцінку 9,73 за 10-бальною шкалою. При цьому на кожну статтю в середньому також припадає 3.11 висловлювань. Попри гарні результати екстракції подій і висловлювань й

надзвичайну високу якість отриманого тексту, дана модель не змогла провести числове оцінювання потрібних показників, тому можна зробити висновок, що вона підходить лише для безпосередньої роботи із текстом.

Article_id	qwen-4b						
	n_opins	text_quality (1-10)	sentiment (-2/+2)	controversy (0-1)	relevancy (0-1)	contribution (0-1)	hotness (0-1)
1							
2							
3	3	10					
4	5	9					
5	3	9					
6	9	10					
7	2	10					
8	2	10					
9	4	9					
11	2	8					
12	4	10					
13	5	10					
14	6	10					

Рисунок 3.18 – Результати оцінювання роботи моделі qwen3-4b (частина)

Для формування конкретних висновків потрібно систематизувати отримані оцінки у таблицю (рис. 3.19).

Модель	Якість тексту	Висловлювань на статтю	Влучність (precision)	Повнота (recall)	F1-score	Accuracy	R2 score	Час обробки (сек/чанк)	Загальний час обробки (50 ст.)
gemma-3-4b-it-qat	9,04	3,09	0,69	0,67	0,68	0,67	0,69	43,88	105 хв.
gemma-3-4b-it	8,82	3,09	0,67	0,63	0,62	0,63	0,34	40,9	97 хв.
llama-3.2-3b-instruct	5,15	3,07	0,45	0,38	0,37	0,38	-0,56	34,78	82 хв.
qwen3-4b	9,73	3,11						33,64	79 хв.

Рисунок 3.19 – Таблиця порівняльного аналізу обраних LLM

Підсумовуючи проведений аналіз результатів роботи моделі інформаційної системи інтелектуального аналізу настрою за різних задіяних LLM-агентів, можна сформулювати декілька висновків. Насамперед варто зазначити, що показник покриття висловлювань на статтю виявився малоінформативним, кожна з моделей виокремлювала в середньому 3.09 висловлювання на статтю. За якістю тексту найкращою є qwen3-4b, проте gemma-3-4b-it-qat також має непоганий результат. Моделі llama-3.2-3b-instruct та qwen3-4b не змогли виконати всі поставлені завдання коректно, тому вони не можуть бути використані як єдиний гіперпараметр моделі аналізу. Модель gemma-3-4b-it при аналогічних часових й ресурсних витратах показала себе гірше за -qat версію. Час обробки одного блоку тексту є здебільшого однаковим, відхилення у меншу сторону спричинені лише фрагментарністю необхідних кроків аналізу. Звідси, модель gemma-3-4b-it-qat виявилась найкращою як агент системи, проте вона програє за якістю тексту. Якщо можливим є використання лише однієї LLM, потрібно обирати gemma-3-4b-it-qat, але за умови можливості паралельного використання qwen3-4b для генерування текстових відповідей. Тоді загальна якість суттєво зросте.

Закріпивши вищезазначені гіперпараметри моделі інформаційної системи аналізу настрою можна досягти оптимального вирішення поставленої задачі – зменшення часу та спрощення проведення системного аналізу позиції осіб стосовно теми із виявленням суперечностей. Умовою оптимальності у даному випадку є досягнення достатнього рівня точності отриманого аналізу від інформаційної системи, що було досягнуто за результатами підбору відповідних налаштувань.

3.4 Висновки

У даному розділі було змодельовано та реалізовано конвеєр аналітичної обробки настрою публічних осіб, що формує модель інтелектуального аналізу. Також було оптимізовано загальні гіперпараметри системи, задля максимізації точності автоматизованого інтелектуального аналізу. Це включало

в себе підбір формул та способів обрахунку базових показників, коригування переданого контексту та вибір оптимальної великої мовної моделі. Було встановлено, що період актуальності висловлювання становить максимум 1 рік, при чому вплив старіших думок на поточне ставлення спадає лінійно від дати публікації. При цьому на фінальну цікавість висловлювання найбільше впливають релевантність (68,19%), суперечливість (13,928%) та внесок (10,861%), а його порогове значення для відображення становить 66.5%. Також проведене експертне оцінювання на тестовій вибірці із 50 статей із порівняльним аналізом великих мовних моделей дозволило обрати оптимальну конфігурацію. LLM-агенти були оцінені за кількома критеріями, після чого було обрано найкращу із можливих моделей для вирішення поставленої задачі. Так, найбільш оптимальним рішенням є використання gemma-3-4b-it-qat, яка показала найкращі результати за f1-score (0.68) та точністю (0.69), при цьому вона найкраще спрогнозувала класи за ROC AUC, а відхилення в оцінці відносились переважно до сусідніх класів шкали, з чого можна зробити висновок про найбільшу загальну точність аналізу. Коефіцієнт детермінації становив 0.69, а загальна якість тексту за 10-бальною шкалою дорівнює 9.04, що свідчить про достатній рівень розуміння даних моделлю та загальну високу якість відповіді. Це у свою чергу підтверджує точність автоматизованого аналізу, що задовольняє умову оптимальності. Було зазначено, що за можливості використання двох моделей у паралелі для роботи із генерацією тексту, варто використовувати qwen3-4b, оскільки він показав вищий результат якості тексту (9.73), але в цілому є гіршим. За результатами досліджено було чітко сформульовані параметри, використання яких дозволяє досягти оптимального результату під час використання моделі інформаційної системи інтелектуального аналізу настроїв.

4 МОДЕЛЮВАННЯ Й РЕАЛІЗАЦІЯ СИСТЕМИ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ СИСТЕМИ

4.1 Підбір оптимального способу представлення результатів аналізу

Усі результати аналізу розробленої інформаційної системи зберігаються в чотирьох таблицях бази даних. Таблиця dimEvent (рис. 4.1) містить заголовки подій, короткі описи й числові оцінювання релевантності, впливу та новизни, які можуть слугувати індикатором «цікавості» кожної події. У dimOpinion (рис. 4.2) фіксуються думки публічних осіб: сама цитата, числовий бал сентименту від -1 до 1, позначка суперечливості з коментарем (за наявності) та внутрішні оцінки релевантності, внеску й суперечливості, що формують інтегральну міру «цікавості» висловлювання. Таблиця dimPerson (рис. 4.3) зберігає базову інформацію про особу, а fctAttitude (рис. 4.4) – узагальнену позицію кожної людини стосовно певної теми: короткий коментар, статичний показник загальної позиції та індикатор ймовірності зміни думки.

	123 event_id	A2 event_title	123 fk_origin_article	A1 description	123 relevance_score	123 influence_score	123 novelty_score	123 event_hotness
1	1	Emissions Cut Potential	86	Budget bill measures could reduce emissions by nearly 1 billion tons annually by 2030.	0.95	0.85	0.6	0.7775
2	2	Greenhouse Energy Use	85	High-tech greenhouses historically required significant electricity due to cost and operati	0.9	0.7	0.6	0.695
3	3	Biden's Climate Agenda Cuts Costs	86	White House economists argue Biden's agenda reduces emissions and lowers consumer c	0.9	0.8	0.6	0.745
4	4	AI-Powered Greenhouses	85	AppHarvest's high-tech greenhouse uses robotics, AI and data to grow tomatoes efficient	0.9	0.8	0.7	0.78
5	5	Vertical Farms Expanding	85	Several companies are developing indoor vertical farms in locations like South Korea, Abu	0.95	0.85	0.7	0.8125
6	6	Transition to Clean Energy	84	Growing concern over extreme weather and environmental disasters is driving a shift tow	1	0.85	0.6	0.785
7	7	Emissions Cut Potential	84	Budget measures could reduce emissions by nearly 1 billion tons annually by 2030.	0.9	0.8	0.6	0.745
8	8	Google Maps Route Finder	84	Google Maps now offers eco-friendly route options for travelers.	0.6	0.4	0.3	0.395
9	9	Nuclear Energy Statistics	84	US nuclear power generated ~20% of national electricity since 1990; 94 reactors operating	0.85	0.75	0.5	0.6775
10	10	Coal Plants Increase Output	83	Chinese officials are ordering coal plants to increase production.	0.8	0.7	0.6	0.68
11	11	Energy Crunch Concerns	83	Global energy demand is rising amid economic recovery, straining renewable energy supp	0.9	0.8	0.7	0.78
12	12	Energy Crisis Intensifies	83	China faces power shortages due to coal imports and consumption limits, leading to incr	0.9	0.8	0.6	0.745
13	13	Minerals Shortage Risk	82	A minerals shortage poses a significant risk to accelerating decarbonization efforts, simila	0.9	0.8	0.7	0.78
14	14	Gas Crisis & Renewables	83	Europe's gas crisis is linked to renewables; wind and solar are cheapest electricity.	0.95	0.85	0.7	0.8125
15	15	Mineral Demand for Renewables	82	Transition to clean energy requires significantly more raw materials than fossil fuels, possi	1	0.9	0.7	0.845
16	16	Global Net-Zero Goal	83	World needs to stop expanding fossil fuels to reach net-zero by 2050, according to the IEA	1	0.8	0.7	0.795
17	17	Cobalt Mining Concerns	82	Majority of world's cobalt mined in DRC with poor protections & child labor.	0.9	0.8	0.7	0.78
18	18	Minerals Shortage Concerns	82	The article highlights a looming minerals shortage due to the increasing demand for elect	0.9	0.8	0.6	0.745
19	19	Supply Gap Concerns	81	Gas & renewables face a supply gap due to coal phase-out; governments must balance cl	0.9	0.8	0.6	0.745
20	20	Gas Prices Impact Low-Income	81	High natural gas prices will disproportionately affect low-income Americans this winter.	0.7	0.6	0.3	0.51
21	21	Energy Crisis Intensifies	81	Global energy shortages and price spikes are worsening, particularly in Europe and China.	0.9	0.8	0.6	0.745
22	22	Tax Bill Emissions Reduction	80	Finance Committee bill will achieve 73% of emissions reductions in the next 10 years via t	0.95	0.85	0.6	0.7775
23	23	Biden's Domestic Agenda	80	President Biden is focused on pushing Democrats to finalize his domestic agenda, includi	0.9	0.8	0.6	0.745
24	24	CEP Program Faces Hurdles	80	Democrats are working on a replacement for the Clean Electricity Performance Program (	0.9	0.8	0.6	0.745
25	25	Grassland Loss in UK	79	Over 2 million acres of grasslands have been lost in the UK due to agriculture, developme	0.7	0.6	0.4	0.545
26	26	Grasslands' Role Missed	79	Report and policy overlooked grasslands' importance in carbon capture, biodiversity, and	0.9	0.8	0.7	0.78
27	27	Grasslands Store Carbon	79	UK grasslands store 1.8 billion metric tons of carbon, equivalent to 400 million cars' emiss	0.95	0.85	0.7	0.8125
28	28	Call to Plant Trillion Trees	79	Environmental groups are urging global tree planting of one trillion trees over the next de	0.9	0.7	0.6	0.695
29	29	Tidal Energy Lag	78	Tidal energy lags behind wind and solar in adoption.	0.85	0.75	0.6	0.7125
30	30	Tidal Energy Pilot Project	78	Verdant Power's underwater turbines are generating electricity for Roosevelt Island, suppl	0.9	0.7	0.6	0.695
31	31	Tidal Energy Development	78	Two firms are developing tidal energy technologies – Orbital Marine Power in Scotland an	1	0.8	0.6	0.76
32	32	Ghana Green Bond Pledge	77	Government pledges \$2 billion in green bonds for environmental priorities and sustainabi	0.95	0.85	0.6	0.7775
33	33	ESG Costs Drive Inflation	76	ESG integration and the transition to eco-friendly business models are contributing to risi	0.9	0.8	0.7	0.78
34	34	ESG Investing Impacts Inflation	76	Increased ESG investing may be contributing to inflation due to 'en-flation' – the impact c	0.85	0.75	0.6	0.7125
35	35	Arctic Ocean Warming	75	New research indicates the Arctic Ocean began warming decades earlier than previously e	0.8	0.6	0.5	0.595
36	36	Wind Farm Approval	75	Federal officials approved plans for a wind farm off Rhode Island coast.	0.9	0.8	0.6	0.745
37	37	COP26 Pact Outcomes	74	Pledges fell short of needed emissions cuts; fossil fuel language deferred.	0.9	0.8	0.6	0.745
38	38	Renewable Energy Growth	74	Global renewable energy capacity is projected to increase significantly by 2026, reachin	1	0.9	0.7	0.845
39	39	Diversity in Outdoors	73	Groups are working to broaden access and representation in outdoor recreation, challeng	0.8	0.7	0.6	0.68

Рисунок 4.1 – Вид даних таблиці dimEvents

	123	opinion_id	123	fk_opinion_article_id	123	fk_person_id	A2: citation	123	sentiment_score	123	inconsistency_flag	A2:	inconsistency_with_id	A2:	inconsistency_comment	123	c
1	1	86	1				Economists highlight the costs of climate change and advocate for clean energy policies	0.6		0	[NULL]		[NULL]				
2	2	85	2				Lam stated that big greenhouses have historically been expensive to operate and required	0.25		0	[NULL]		[NULL]				
3	3	86	3				White House Council of Economic Advisers member Heather Boushey, along with Noah	0.65		0	[NULL]		[NULL]				
4	4	85	4				We have to figure out how to grow a lot more food with a lot less resources, all the while	0.25		0	[NULL]		[NULL]				
5	5	85	5				He believes greenhouse farming can help ensure food security and reduce costs/emission	0.25		0	[NULL]		[NULL]				
6	6	84	6				Professor Vijay Modi believes that widespread understanding of renewable possibilities wi	0.65		0	[NULL]		[NULL]				
7	7	84	7				Zhang stated the US has sufficient land for solar deployment, creating jobs and requiring	0.65		0	[NULL]		[NULL]				
8	8	84	8				Modi highlighted the challenge of locating geothermal hotspots in the US and exploring	-0.25		0	[NULL]		[NULL]				
9	9	84	9				Senators Bernie Sanders and Elizabeth Warren have called to phase out nuclear energy.	-0.6		0	[NULL]		[NULL]				
10	10	83	10				US President Joe Biden's power crunch is that it appears to be strengthening the argument	-0.8		0	[NULL]		[NULL]				
11	11	83	11				The worry with China's power crunch is that it appears to be strengthening the argument	-0.6		0	[NULL]		[NULL]				
12	12	83	12				Chinese President Xi Jinping announced a halt to financing overseas coal projects, aiming	0.85		0	[NULL]		[NULL]				
13	13	82	13				Materials are the first-order problem in fighting climate change, and we need to ramp up	0.6		0	[NULL]		[NULL]				
14	14	83	14				The Green Deal would provide the 'only lasting solution' to Europe's energy challenge.	0.85		0	[NULL]		[NULL]				
15	15	82	15				Associate producer for 'Fareed Zakaria GPS' offered commentary on the resource deman	0.25		0	[NULL]		[NULL]				
16	16	83	16				Here are a few options, all under 300 characters: ""Option 1 (Concise):"" "Extreme climate	0.65		0	[NULL]		[NULL]				
17	17	82	17				Rare earth elements (neodymium) extraction primarily occurs in China, used for wind tur	-0.6		1	[13]		The new citation shifts from a brc				
18	18	82	18				The bipartisan infrastructure bill would provide billions of dollars to support the product	0.65		0	[NULL]		[NULL]				
19	19	81	19				Governments need to address the challenge of acting on climate change while ensuring a	0.6		0	[NULL]		[NULL]				
20	20	81	20				Governments must ensure they don't neglect the remaining need for fossil fuels, especial	-0.8		1	[13]		The new citation shifts from prior				
21	21	81	21				Director of Energy, Climate and Resources at Eurasia Group, believes the crisis will accel	0.65		0	[NULL]		[NULL]				
22	22	80	22				Clean electricity can catalyze a clean economy, enabling decarbonization across transport	0.85		0	[NULL]		[NULL]				
23	23	80	23				Manchin's spokeswoman Sam Runyon stated that Senator Manchin supports combating	-0.25		0	[NULL]		[NULL]				
24	24	80	24				Sen. Tina Smith of Minnesota confirmed that Democrats are developing a replacement fo	-0.6		0	[NULL]		[NULL]				
25	25	79	25				Ranscombe Farm nurtures rare plants in small sections, hoping they will grow into bigger	0		0	[NULL]		[NULL]				
26	26	79	26				Bennett criticized the Net Zero Strategy for lacking recognition of nature's role and recycl	-0.6		0	[NULL]		[NULL]				
27	27	79	27				But they are pretty much ignored or have been ignored in many sustainability policies,"	-0.6		0	[NULL]		[NULL]				
28	28	79	28				Ian Dunn, chief executive of Plantlife, stated that iconic grassland habitats like those in th	-0.25		0	[NULL]		[NULL]				
29	29	78	29				We're underwater, and it's probably easier to get a rocket to the moon than to get these tr	-0.8		0	[NULL]		[NULL]				
30	30	78	30				The best way to envision what Verdant Power's technology is, is to think of wind turbines	0.65		0	[NULL]		[NULL]				
31	31	78	31				The bits of technology that generate power look not too different to a wind turbine," He	0.65		0	[NULL]		[NULL]				
32	32	77	32				We are creating employment, and also saving our environment," - Trashy Bags country'r	0.65		0	[NULL]		[NULL]				
33	33	76	33				Orlando predicts the Fed will need to hike rates twice next year and four more times in 20	-0.6		0	[NULL]		[NULL]				
34	34	76	34				I can't see a reason why the Fed will not acknowledge the 'E' in its inflation outlook for m	-0.25		0	[NULL]		[NULL]				
35	35	75	35				RODA Executive Director Annie Hawkins stated that the South Fork project will generate	-0.85		0	[NULL]		[NULL]				
36	36	75	36				The Biden administration aims to grow renewable energy capacity. Officials said the deve	0.85		1	[20, 17]		The new citation shifts focus to B				
37	37	74	37				India requested a change to the text from 'phasing out' coal to 'phasing down' in the Glo	-0.25		0	[NULL]		[NULL]				
38	38	74	38				The International Energy Agency (IEA) estimates that renewables need a far bigger push to	-0.6		1	[19, 13, 18]		The new citation shifts from bro				
39	39	73	39				Here are a few options, all under 300 characters: ""Option 1 (Concise):"" "Taylor wants 202	0		0	[NULL]		[NULL]				
40	40	73	40				Nicole Boyd and Naphara Cade co-founded Black Girls Hike RVA to address disparities in	0		0	[NULL]		[NULL]				
41	41	73	41				Historical institutional racism has banished people of color from the great outdoors," He	-0.25		0	[NULL]		[NULL]				
42	42	73	42				Researchers argue that racial disparities in outdoor participation stem from cultural differ	-0.25		0	[NULL]		[NULL]				

Рисунок 4.2 – Вид даних таблиці dimOpinion (частково)

1	1	Boushey, Kaufman and Bressler	https://upload.wikimedia.org/wikipedia/commons/thumb/5/50/Heather_M._Boushey%2C_CEA_Member.jpg/1200px-Heather_M._Boushey%2C_CEA_Member.jpg
2	2	Lam	https://skla.cuhk.edu.hk/wp-content/uploads/member/Liming-Xiong.jpg
3	3	Heather Boushey	https://upload.wikimedia.org/wikipedia/commons/thumb/5/50/Heather_M._Boushey%2C_CEA_Member.jpg/1200px-Heather_M._Boushey%2C_CEA_Member.jpg
4	4	Jonathan Webb	https://upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Jonathan_Webb_August_2011.jpg/1200px-Jonathan_Webb_August_2011.jpg
5	5	Professor Hon-ming Lam	https://skla.cuhk.edu.hk/wp-content/uploads/member/Liming-Xiong.jpg
6	6	Vijay Modi	https://www.earth.columbia.edu/sites/default/files/image/education/profiles/Modi_250.jpg
7	7	K. Max Zhang	https://www.engineering.cornell.edu/wp-content/uploads/2024/06/k-max-zhang.jpg
8	8	Narendra Modi	https://upload.wikimedia.org/wikipedia/commons/c/c4/Official_Photograph_of_Prime_Minister_Narendra_Modi_Portrait.png
9	9	Bernie Sanders	https://upload.wikimedia.org/wikipedia/commons/thumb/2/2e/Bernie_Sanders_2023.jpg/960px-Bernie_Sanders_2023.jpg
10	10	Joe Biden	https://upload.wikimedia.org/wikipedia/commons/thumb/6/68/Joe_Biden_presidential_portrait.jpg/960px-Joe_Biden_presidential_portrait.jpg
11	11	Christine Shearer	https://sustainable.solutions.stanford.edu/sites/g/files/tbjybj31186/files/styles/large_square/public/media/image/shearer_0_0.jpeg?h=eb250db1&tok=BUY56pRJ
12	12	Xi Jinping	https://imageio.forbes.com/specials-images/imageserve/5a6ccdd31358e612fb80afa/0x0.jpg?format=jpg&crop=2320,2319,x1234,y348,safe&height=416&width=416&fit=bounds
13	13	Unspecified	[NULL]
14	14	Kadi Simson	https://upload.wikimedia.org/wikipedia/commons/9/95/Kadi_Simson_2017-05-25_%28cropped%29.jpg
15	15	Jonah Bader	https://media.licdn.com/dms/image/v2/D5E3QAQh4_6RtG035w/profile-displayphoto-shrink_200_200/0/1665893551412?e=2147483647&v=beta&h=3V7yGjAL7v38OcjTg2n2AvZOIbw0&PZL
16	16	Charles Moore	https://upload.wikimedia.org/wikipedia/commons/9/9b/Charles_Moore%2C_former_editor_of_the_Daily_Telegraph%2C_at_Edmund_Burke_Philosophy%2C_Politician%2C_Prophet_%28cropped%29.jpg
17	17	Henning Gloystein	https://cdn.streamlyvideo.org/_9_Henning_Gloystein_portrait_2_He_c83ac9d9f6d4d48a22768991024b61_6347cd34c05eecd1e4025fd6876e08_0cf1d0731.png
18	18	Leah Stokes	https://ewp.jonmimix.com/wp-content/uploads/2022/07/Leah-Stokes-Headshot-2023-Sq.jpg
19	19	Sam Runyon	https://img-legisform.com/people/324/72234.jpg?rand=535800
20	20	Tina Smith	https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRLBaiQFQOCdljilngJfNPB3xmsKYLE-n5AA8s
21	21	Sweeney	https://upload.wikimedia.org/wikipedia/commons/thumb/2/2f/Sydney_Sweeney_at_the_2024_Toronto_International_Film_Festival_%28cropped%2C_rotated%29.jpg/1200px-Sydney_Sweeney_at_the_2024_Toronto_International_Film_Festival_%28cropped%2C_rotated%29.jpg
22	22	Craig Bennett	https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcOPr-YWC7qDDBelQZu483wv98cc3XQyCg8s
23	23	Richard Bardgett	https://www.research.lancs.ac.uk/portal/files/421659327/1920_professorrichardbardgett_1_.jpg
24	24	Ian Dunn	https://lovetlogan.co.uk/wp-content/uploads/2021/12/Ian-Dunn-HIGH-RES.jpg
25	25	Taylor	https://media.pff.com/player-photos/nfl/125278.png
26	26	Trey Taylor	https://media.pff.com/player-photos/nfl/125278.png
27	27	Andrew Scott	https://upload.wikimedia.org/wikipedia/commons/1/17/Andrew_Scott_at_Radiata_2025-05-28_cropped%29.jpg

Рисунок 4.3 – Вид даних таблиці dimPerson (частина)

	121	fk_topic_id	122	fk_person_id	created_at	modified_at	123	sentiment_deviation	A2	stance	A2	person_summary
1	1	1	1	1	2025-05-28 19:43:53	2025-05-28 19:43:53	0	pro	Economists Boushey, Kaufman and Bressler advocate for pro-green energy policies citing potential cost savings through clean electricity programs and			
2	1	2	1	2	2025-05-28 19:44:19	2025-05-28 19:44:19	0	mixed	Lam advocates for high-tech greenhouses using LED tech & insulation, citing reduced energy needs despite historical costs; however, he deems cereal			
3	1	3	1	3	2025-05-28 19:45:02	2025-05-28 19:45:02	0	pro	Here are a few options, all under 300 characters: "Option 1 (Most Concise):" Heather Boushey supports Biden's climate agenda, citing potential for low			
4	1	4	1	4	2025-05-28 19:45:25	2025-05-28 19:45:25	0	mixed	Here are a few options within 300 characters, refining Jonathan Webb's statements: "Option 1 (Concise):" "Jonathan Webb champions AI-driven green			
5	1	5	1	5	2025-05-28 19:46:51	2025-05-28 19:46:51	0	mixed	Professor Lam advocates for greenhouse farming near consumption areas to bolster food security and reduce emissions, citing examples like NextOn's			
6	1	6	1	6	2025-05-28 19:47:08	2025-05-28 19:47:08	0	pro	Vijay Modi advocates for rapid renewable adoption due to affordable solar and wind, emphasizing public understanding as a key driver. He believes Gov			
7	1	7	1	7	2025-05-28 19:47:38	2025-05-28 19:47:38	0	pro	K. Max Zhang advocates for expanding US green energy through solar and wind deployment, citing sufficient land and job creation potential. He emp			
8	1	8	1	8	2025-05-28 19:47:56	2025-05-28 20:46:34	0.230588	mixed	Narendra Modi increasingly links Reliance's green energy push to his vision, despite ongoing reliance on coal. Recent reports show a shift towards tying			
9	1	9	1	9	2025-05-28 19:48:31	2025-05-28 19:48:31	0	against	Bernie Sanders now opposes nuclear energy, citing its non-renewable nature despite being considered "clean" by some. He and Warren advocate phasi			
10	1	10	1	10	2025-05-28 19:48:45	2025-05-28 21:50:58	0.488185	mixed	President Biden is shifting toward supporting green energy despite Republican resistance, exemplified by lawsuits challenging transition proposals and			
11	1	11	1	11	2025-05-28 19:50:14	2025-05-28 19:50:14	0	against	Christine Shearer opposes accelerating renewable energy transition due to China's power crunch, citing concerns about pro-coal interests and infrastru			
12	1	12	1	12	2025-05-28 19:50:41	2025-05-28 19:50:41	0	pro	Xi Jinping now supports peaking emissions by 2030 and carbon neutrality by 2060 after halting overseas coal financing, despite recent increases in dam			
13	1	13	1	13	2025-05-28 20:07:47	2025-05-28 20:01:58	0.576456	mixed	Shifting stance on green energy, now emphasizing Republican control over high-carbon states and aligning with GOP geographical strength. Recent de			
14	1	14	1	14	2025-05-28 20:08:54	2025-05-28 20:08:54	0	pro	Here are a few options, all under 300 characters: "Option 1 (Concise):" Kadi Simson champions the Green Deal as Europe's energy future, citing renew			
15	1	15	1	15	2025-05-28 20:08:59	2025-05-28 20:08:59	0	mixed	Here are a few options within 300 characters, refining the original text: "Option 1 (Concise):" "Jonah Bader (CNN) now sees renewable energy's huge m			
16	1	16	1	16	2025-05-28 20:09:07	2025-05-28 20:09:07	0	pro	Pro-green energy, citing UK's 2035 decarbonization plan as an example. Moore suggests fossil fuel impacts could push nations toward renewables, dees			
17	1	17	1	17	2025-05-28 20:13:31	2025-05-28 20:37:03	0.680604	mixed	Henning Gloystein now expresses concern about nuclear's future following a potential accident, contrasting with his previous optimistic view due to fu			
18	1	18	1	18	2025-05-28 20:13:36	2025-05-28 20:46:29	0.0267711	pro	Leah Stokes remains a strong proponent of clean energy policies like electrification and EVs to combat volatile fossil fuel prices, acknowledging particu			
19	1	19	1	19	2025-05-28 20:13:41	2025-05-28 20:13:41	0	mixed	Sam Runyon supports combating climate change and American energy independence, expressing concerns about taxpayer funding for private compai			
20	1	20	1	20	2025-05-28 20:14:41	2025-05-28 20:14:41	0	against	Tina Smith opposes the Clean Electricity Performance Program (CEPP), a key part of Biden's climate plan, due to opposition from Sen. Manchin. She's s			
21	1	21	1	21	2025-05-28 20:16:05	2025-05-28 20:16:05	0	mixed	Sweeney advocates for careful nuclear expansion at Ranscombe Farm, acknowledging decades of environmental damage and a long-term (20+ year) re			
22	1	22	1	22	2025-05-28 20:16:11	2025-05-28 20:16:11	0	against	Craig Bennett opposes the Net Zero Strategy's neglect of nature's role, citing limited funding and reliance on recycled policies. He argues for greater rev			
23	1	23	1	23	2025-05-28 20:16:16	2025-05-28 20:16:16	0	against	Richard Bardgett opposes prioritizing giant fans over grasslands for carbon capture, citing their neglect in sustainability policies and the significant rev			
24	1	24	1	24	2025-05-28 20:18:08	2025-05-28 20:18:08	0	mixed	Here are a few options within 300 characters, refining the original text: "Option 1 (Concise):" "Ian Dunn champions global grassland protection – like Si			
25	1	25	1	25	2025-05-28 20:18:55	2025-05-28 20:18:55	0	against	Taylor remains firmly against green energy, particularly underwater sources like tidal power, citing significant technical challenges. He acknowledges th			
26	1	26	1	26	2025-05-28 20:20:16	2025-05-28 20:20:16	0	pro	Here are a few options within 300 characters, refining the original text: "Option 1 (Concise):" "Trey Taylor promotes underwater river turbines for contin			
27	1	27	1	27	2025-05-28 20:20:22	2025-05-28 20:20:22	0	pro	Here are a few options, all under 300 characters: "Option 1 (Concise):" "Andrew Scott champions tidal energy – like wind – as a reliable, clean source v			
28	1	28	1	28	2025-05-28 20:21:16	2025-05-28 20:21:16	0	pro	Bright Biney supports green energy and sustainable practices, citing Trashy Bags' employment creation and environmental benefits. He emphasizes tha			
29	1	29	1	29	2025-05-28 20:22:01	2025-05-28 20:22:01	0	against	Phil Orlando opposes green energy's impact on inflation, predicting Fed rate hikes to combat rising prices due to ESG spending and supply constraints;			
30	1	30	1	30	2025-05-28 20:22:44	2025-05-28 20:22:44	0	mixed	Seema Shah believes rising ESG investing and "en-flation" driven by green policies will keep inflation elevated, urging the Fed to acknowledge this facto			
31	1	31	1	31	2025-05-28 20:22:50	2025-05-28 20:22:50	0	against	Annie Hawkins opposes green energy, citing minimal climate benefit for the South Fork project. Her stance aligns with RODA's concerns about potent			
32	1	32	1	32	2025-05-28 20:24:32	2025-05-28 20:24:32	0	mixed	India now favors "phasing down" coal in the Glasgow Pact, shifting from a previous stance against strong fossil fuel language. This reflects India's com			
33	1	33	1	33	2025-05-28 20:27:18	2025-05-28 20:27:18	0	mixed	Here are a few options, all under 300 characters: "Option 1 (Concise):" "Eddie Taylor champions diversifying outdoor recreation by amplifying BIPOC,			
34	1	34	1	34	2025-05-28 20:27:23	2025-05-28 20:27:23	0	mixed	Nicole Boyd & Narshara Cade advocate for equitable access to outdoor spaces, citing historical exclusion rooted in eugenics and White supremacy with			
35	1	35	1	35	2025-05-28 20:28:22	2025-05-28 20:28:22	0	mixed	Jerry Lee argues historical racism has excluded people of color from outdoor spaces, linking this to elite romanticization of nature and its roots in eug			
36	1	36	1	36	2025-05-28 20:29:18	2025-05-28 20:29:18	0	mixed	Lee argues institutional racism, not cultural differences, drives outdoor participation disparities due to higher poverty rates, contradicting research sugg			
37	1	37	1	37	2025-05-28 20:31:17	2025-05-28 20:31:17	0	against	Greig Aitken opposes new gas infrastructure, citing stranded asset risks and potential undermining of climate goals. He suggests a shift towards green f			
38	1	38	1	38	2025-05-28 20:31:23	2025-05-28 20:31:23	0	mixed	Here are a few options within 300 characters, refining the original text: "Option 1 (Concise):" "TotalEnergies promotes natural gas & carbon capture fo			
39	1	39	1	39	2025-05-28 20:31:28	2025-05-28 20:31:28	0	against	Equinor supports carbon limits in electricity/heat, opposes them in new gas projects, despite green hydrogen investments. Remains largely resistant to			
40	1	40	1	40	2025-05-28 20:32:28	2025-05-28 20:32:28	0	mixed	Pascal Canfin advocates a mixed approach to green energy, proposing limited gas use with CO2 limits and a 2030 sunset clause, despite criticism as "a			
41	1	41	1	41	2025-05-28 20:33:58	2025-05-28 20:33:58	0	mixed	EU officials' stance shifted to include natural gas and nuclear in the taxonomy, potentially boosting green investment despite previous reluctance. This			
42	1	42	1	42	2025-05-28 20:34:12	2025-05-28 20:34:12	0	against	Jeff Green opposes the Mormon church due to its wealth and hindering progress on civil rights, including LGBTQ+ rights, citing unhealthy gender par			
43	1	43	1	43	2025-05-28 20:34:43	2025-05-28 20:34:43	0	against	Fossil fuel companies exploit hydrogen hype to prolong natural gas use and influence EU energy policy. InfluenceMap analysis shows this lobbying inn			
44	1	44	1	44	2025-05-28 20:35:03	2025-05-28 20:35:03	0	pro	Andrew Forrest firmly supports green hydrogen as a practical solution to global warming, dismissing blue/grey options as misleading distractions. He's			

Рисунок 4.4 – Вид даних таблиці fctAttitude

Оскільки кожна з цих таблиць містить різні типи даних, доцільно попередньо підібрати варіанти візуалізації. По-перше, для fctAttitude можна використовувати динамічні картки з ім'ям особи, узагальненим коментарем і базовою шкалою позиції – це дозволить відразу бачити загальний тон кожного учасника. По-друге, події з dimEvent доцільно розміщувати як позначки на часовій шкалі з підказками, що міститимуть назву та короткий опис. Такий підхід допоможе синхронізувати хронологію важливих етапів із даними про поведінку осіб. По-третє, погляди з dimOpinion можуть бути відображені комбінованим графіком: вісь X – дата публікації, вісь Y – числовий бал настрою, а розмір «бульбашки» (або товщина маркера) може відповідати інтегральній мірі «цікавості» висловлювання. Колір чи умовні позначки штрихування використовуватимуться для сигналізації про наявність суперечностей, а лінії, що з'єднують цитати однієї людини, дозволять простежити еволюцію її позиції.

Вибір саме таких типів візуалізації виглядає доцільним з огляду на завдання аналітика: картки надають швидкий загальний огляд сутностей, шкала з позначками подій – доповнює інформацію в часі, а комбінований графік зі настроєм-балами й показником «цікавості» дає змогу оцінити динаміку тональності та важливість кожного висловлювання. Застосування інтерактивних

елементів може суттєво спростити подальший етап проектування інтерфейсу та зменшити час на аналіз сирих даних.

4.2 Прототипування інтерфейсу віджету

Враховуючи зазначені доцільні типи візуалізації варто сформувати комбіновану візуалізацію із додатковою інформацією у вигляді карток осіб та карток подій. Підсумкова інформація має відображати позицію особи, дозволяти фільтрування для прослідковування думки та відображати підсумок за необхідності. Фонові події загального характеру також мають формувати картки, які будуть прив'язуватись до дати події на часовій шкалі та створювати хронологію суспільних змін. Головна ж інформація – дані висловлювань. Вона має бути представлена у вигляді комбінованого графіку, що поєднує елементи bubble chart та beeswarm chart, представлених на рисунках 4.5 та 4.6 відповідно. Головними відстежуваними показниками є: час висловлювання, який сформує вісь абсцис та ставлення до теми, яке сформує вісь ординат. При цьому факт висловлювання буде відзначений маркером, розмір якого залежатиме від загальної цікавості події. Маркери будуть розміщуватись без накладання, притягуючись до дати висловлювання (подібно до swarm chart). Для забезпечення інтегральності думки певної особи проводилась лінія між кожним висловлюванням кожної особи, особливо виділяючи кольором наявність протиріччя.

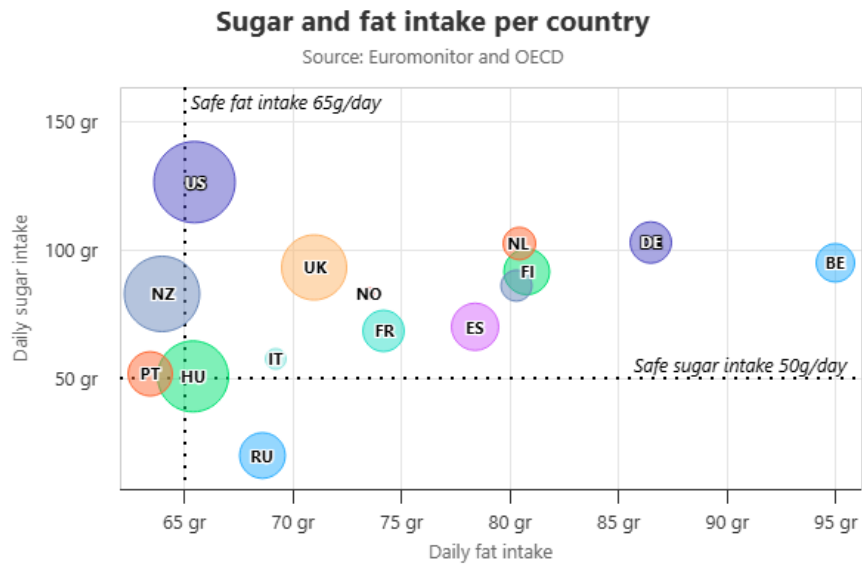


Рисунок 4.5 – Бульбашковий графік (має змінний розмір маркера та дві вісі)

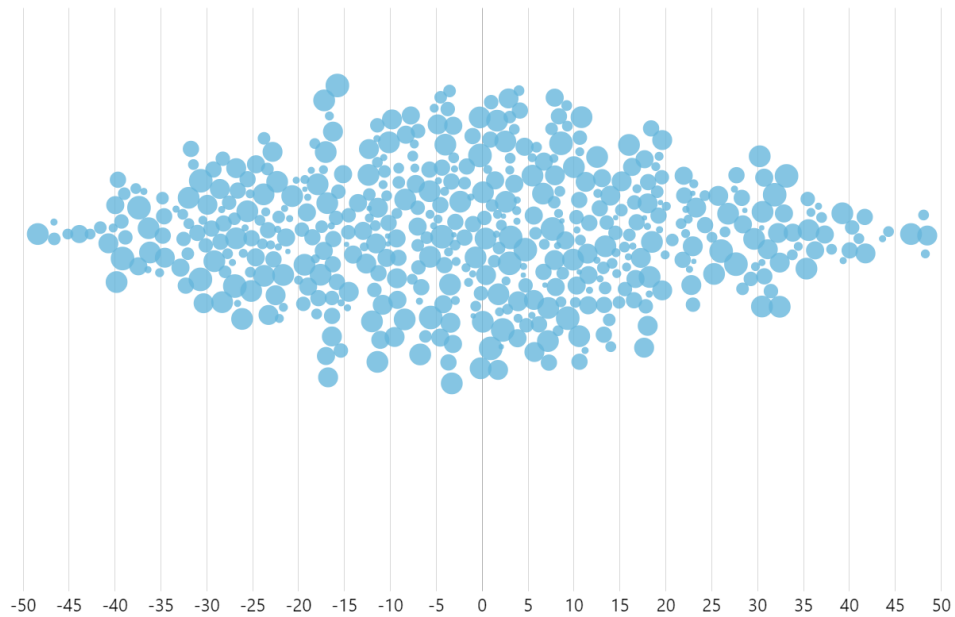


Рисунок 4.6 – Beeswarm chart (маркери групуються навколо деякого значення, для даного проекту – до дати висловлювання)

Реалізуючи вищезазначені зауваження був розроблений прототип інтерактивного віджету у Figma (рис. 4.7).

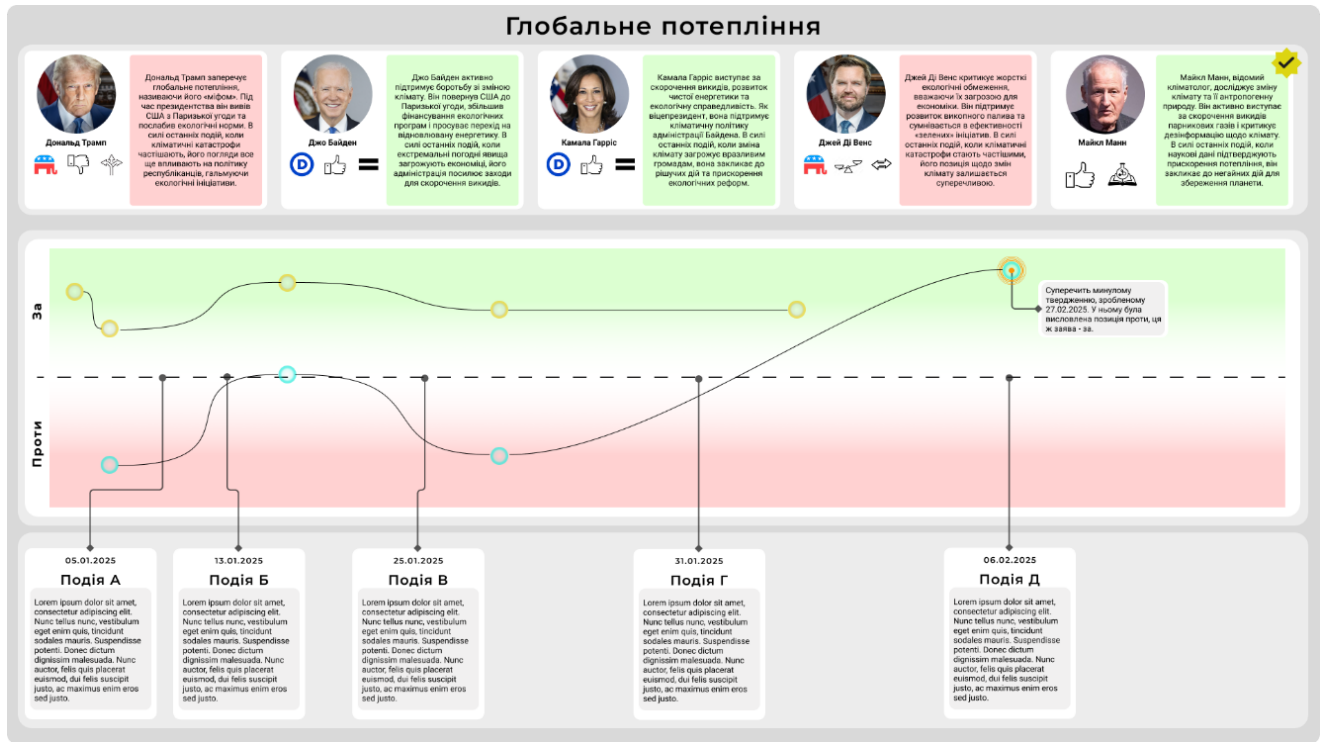


Рисунок 4.7 – Прототип візуалізації аналізованої інформації

Як можна побачити, віджет має три секції: картки персон, головну візуалізацію та картки фонових подій. При наповненні даної візуалізації контекстом має відбуватись формування картини висловлювань, що, при умові достатньої точності аналізу, дозволить швидко та у зручній формі отримувати висновок про загальне ставлення до теми, хронологію позиції конкретної публічної особи, події у світі, що могли змінити її думку, та можливі непослідовності.

4.3 Реалізація системи візуалізації сентименту

Задля реалізації прототипу візуалізації як основний інструмент було обрано React Native та бібліотеку Vite. Розроблений інтерактивний віджет, представлений на рисунку 4.8, відтворив весь необхідний функціонал. Його структура включає три ключові елементи: блок фільтрів і вибору осіб, основну хронологічну шкалу з маркерами тверджень і подій, а також панель із фоновими новинами або подіями, що створюють контекст.

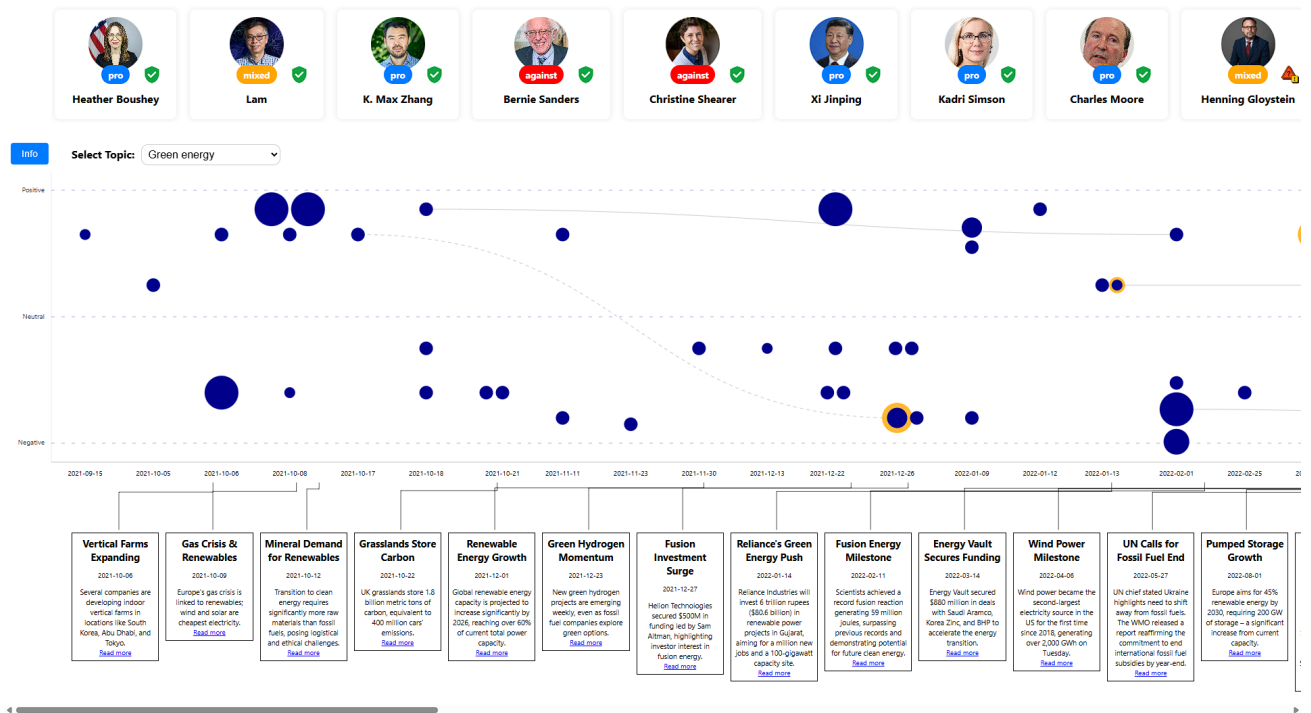


Рисунок 4.8 – Віджет візуалізації отриманої інформації

Підсумки позицій за особою розміщено у верхній частині віджету. На них відображається портрет особи, ставлення до теми та позначка суперечливості. Це дозволяє швидко дізнатись позицію особи та отримати висновок щодо послідовності думок. Якщо присутній значок верифікації – людина притримується відносно постійної позиції, інакше буде присвоєно мітку протиріч. Ці картки відображають всіх дійових осіб за аналізований період, чия думка виявилась мінімально цікавою, а при натисненні на картку можна отримати підсумок позиції та відфільтрувати всі висловлювання на графіку (рис.4.9).

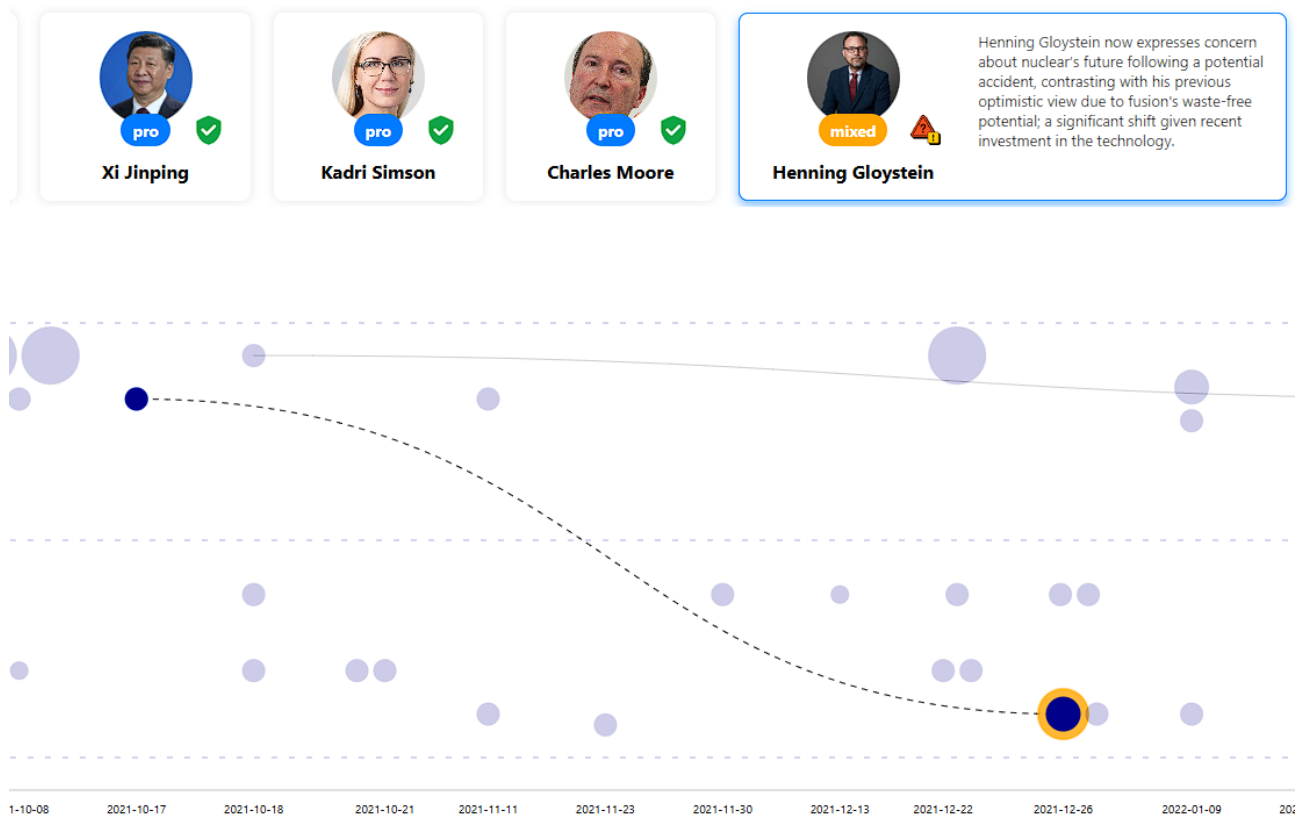


Рисунок 4.9 – Підсумок позиції особи із фільтруванням записів

Із рисунку 4.9 зрозуміло, що послідовні висловлювання однієї і тієї ж особи пов'язуються лінією, при чому наявне протиріччя виділяється окремо. Від висловлювання, у якому з'являється відмінна від попередніх позиція, проводиться пунктирна лінія до усіх актуальних висловлювань, у яких містилась інформація про інше ставлення. При цьому подія, у якій виникло протиріччя блимає помаранчевим, а усі джерела відмінної думки перед цим – блимають блакитним.

При наведенні на бульбашку відображається підказка, що зазначає особу, дату, цікавість та цитату персони із зазначенням суперечностей, якщо вони є (рис. 4.10). Даний елемент стає ключовим у дослідженні позиції особи, оскільки він у стисло подає цитату на часовій шкалі. Це в свою чергу дозволяє у зручній формі переглянути всю хронологію думок. Також запроваджене автоматичне відслідковування настрою щодо теми та суперечностей із минулими висловлюваннями дозволяє ще більше спростити аналіз ставлення, дотримуючись достатнього рівня точності. Сформована підказка також має міру

протидії помилкам у аналізі. За кожним висловлюванням прикріплене першоджерело, тому зацікавлена особа завжди може зробити власні висновки.

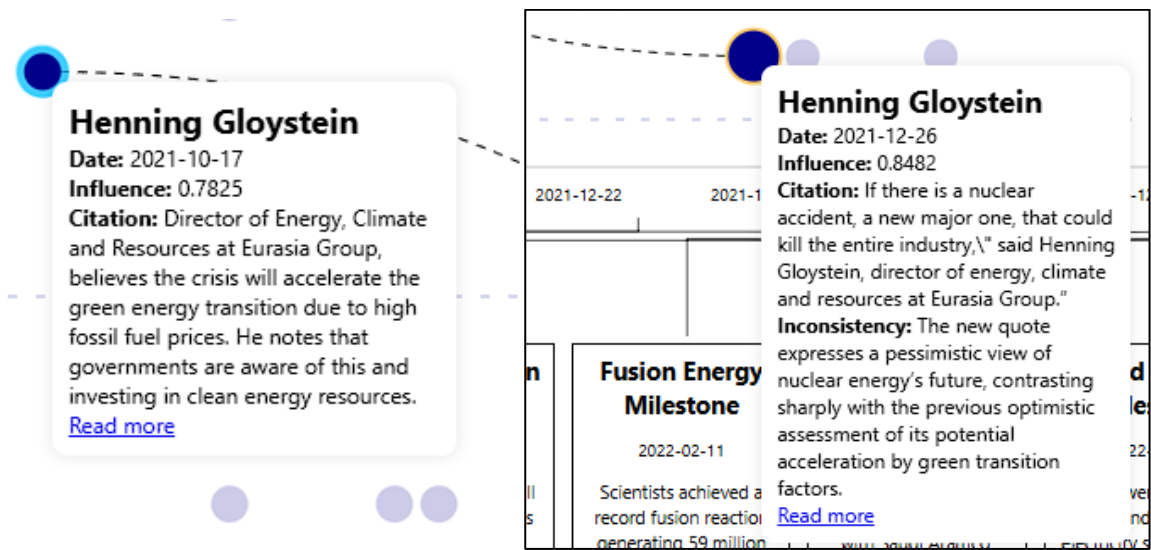


Рисунок 4.10 – Підказки із контекстом висловлювань

Такий підхід до зберігання та виведення даних дозволяє поєднати гнучкість аналітики з наочністю подання, що є важливим у контексті дослідження інформаційних ландшафтів та динаміки змін у публічному дискурсі.

4.4 Інтерпретація результатів аналізу настрою із визначенням суперечностей у риториці публічних осіб

Задля оцінювання ефективності роботи моделі інформаційної системи інтелектуального аналізу ставлення публічної особи до теми потрібно дослідити результати опрацювання широкого об'єму статей на різну тему. Заради цього було відібрано три найпопулярніші наразі теми обговорень із явною відмінністю у позиціях ключових осіб із сайту CNN [25]. Ними стали – зелена енергетика, цензура у соціальних мережах та торговельна війна США. За кожною з них було зібрано 3 пошукові сторінки актуальних статей, сумарно досягнувши кількості у 258. Після попереднього оброблення було отримано 884 блока тексту орієнтованою довжиною 2500 символів кожен. Повне оброблення цього корпусу

статей системою аналізу зайняло 11 годин 40 хвилин, усі результати в процесі вивантажувались у базу даних.

За результатами аналізу системою теми зеленої енергетики (рис. 4.11) видно, що її обговорення є тривалим, не дуже популярним та досить контрверсійним. Висловлювання публічних осіб розтягнені у часі, тому можна прослідкувати позицію особи на тривалому проміжку і достатньо точно перевірити підсистему визначення протиріч. Через досить невелику новизну теми зеленої енергетики послідовних висловлювань однієї особи мало, проте все ж можна прослідкувати їх позицію. Переважна кількість таких висловлювань дотримується попередньої позиції, або ж залишається у її межах. Все ж деяка зміна позицій присутня. Так, директор з питань енергетики, клімату та ресурсів Eurasia Group спершу яскраво підтримав ініціативу зеленої енергії, але потім розкритикував використання атомних електростанцій. Європейська комісія притримувалась нейтралітету, проте після трьох місяців очікування яскраво підтримала зміну, назвавши конкретні цільові показники впровадження відновлюваних джерел енергетики. Джо Байден, який спершу підтримував відбудову нафтової промисловості, змінив риторику у бік підтримки зеленої енергетики, а ще через рік заявив про проблеми із впровадженням цих ініціатив. Дональд Трамп, який відзначився своєю непостійністю у даному питанні, спочатку підтримував зелену енергетику, але наразі активно виступає проти будь-яких програм часів Джо Байдена, заохочуючи використання викопного палива.

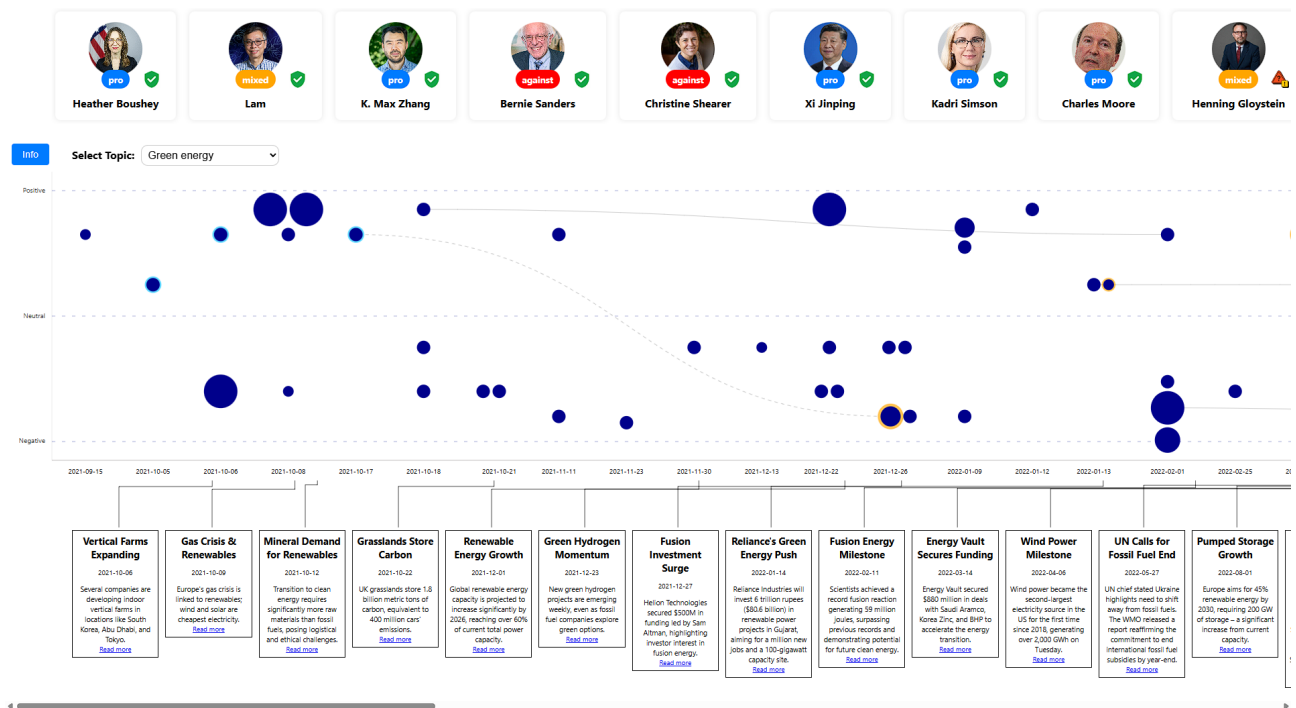


Рисунок 4.11 – Результати роботи системи стосовно теми зеленої енергетики (частина)

Тема цензури у соціальних мережах є досить новітньою, їй характерна досить висока суперечливість (рис. 4.12). Переважна кількість висловлювань є негативними, проте останнім часом ключові дієві особи виявляють деяку симпатію щодо цензури. Оскільки тема має активне обговорення, легко прослідкувати висловлювання однієї особи. Так, Ілон Маск, всупереч публічній підтримці свободи слова, де-факто виступає проти неї. Він знизив контроль над перевіркою на істинність новин у власній мережі X та активно поширює дезінформацію (в тому числі власноруч), яка підтримує його політичну позицію. Дональд Трамп у період до президентських виборів висловлювався проти цензури, звинувачуючи своїх політичних опонентів у односторонності висловлювань деяких тем. Проте після вступу на посаду президента США він прийняв деякі закони, що впроваджують публічну цензуру, з метою «відновлення платформ свободи слова». Марк Цукерберг, який завжди активно виступав за свободу слова, у свої останніх рішеннях скасував перевірки на істинність у соціальних мережах компанії Meta, підтверджуючи публічно, що це рішення виявилось коректним через зменшення політичної заангажованості

медіапростору. Наразі він співпрацює із Дональдом Трампом щодо зменшення кількості модерації та перевірки фактів у медіасфері, що потенційно може допомогти поширенню фейкової інформації.

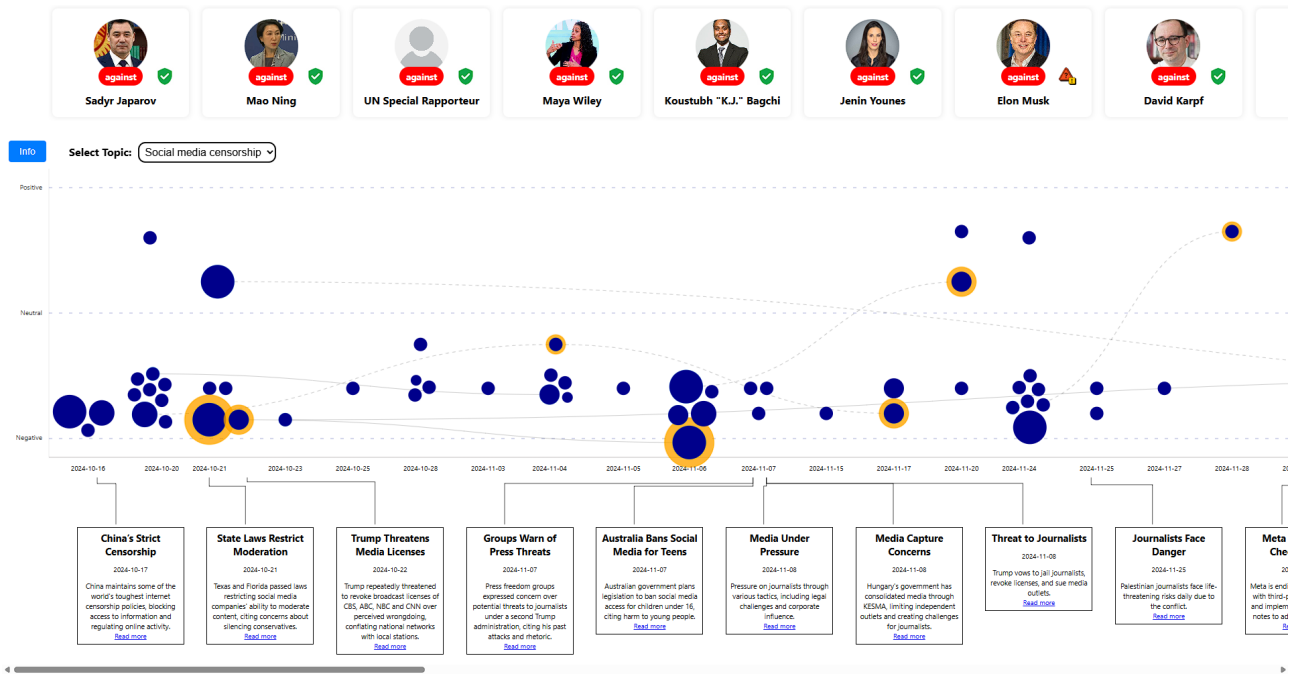


Рисунок 4.12 – Результати роботи системи стосовно теми цензури у соціальних мережах (частина)

Тема торговельної війни США виявилась найновішою, тому має найбільше скупчення висловлювань (рис. 4.13). Переважне ставлення до неї є різко негативним. Темі притаманна дуже висока суперечливість, переважну частину якої складають висловлювання Президента США Дональда Трампа. Йому характерне множинне висловлювання на підтримку тарифів та обіцянки найшвидшого процвітання країни. Він одночасно пропагує незалежність для США та покращення стосунків із іншими країнами, попри різкі дії у їх сторону. Поміж підтримки своєї політики тарифів, Трамп активно заспокоює населення, та переконує у тимчасовості погіршення добробуту громадян.

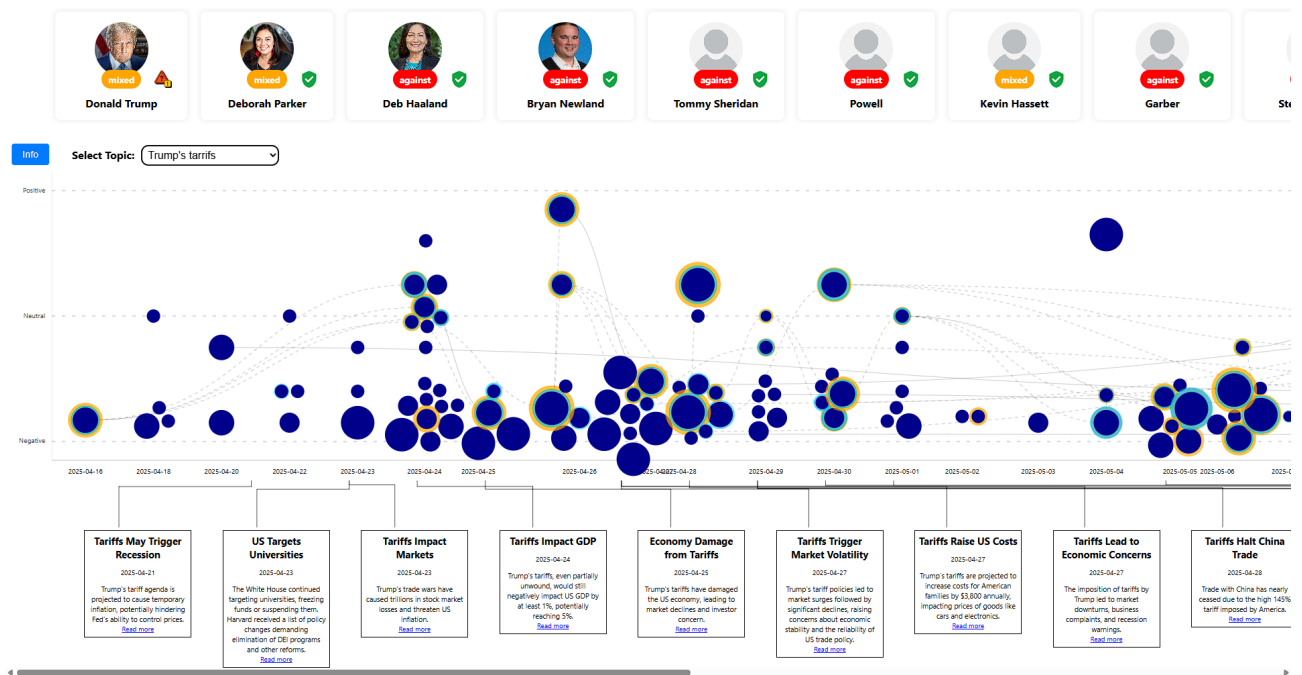


Рисунок 4.13 – Результати роботи системи стосовно теми торговельної війни США (частина)

Після аналізу ключових висловлювань за різними темами стає зрозуміло, що система правильно виокремлює цитати й події, коректно оцінюючи їх загальну цікавість. Визначаючи ставлення персони до теми, система досягла достатнього рівня точності, як було оцінено раніше. Вона точно присвоює позитивність чи негативність ставлення до теми. Варто зазначити, що часто через розмите визначення теми дискурсу система не здатна виявити істинну позицію особи. У таких випадках визначається лише емоційне забарвлення, що спричиняє переважну більшість помилок.

Беручи до уваги системність збору даних висловлювань протягом великих проміжків часу та коректність їх екстракції можна стверджувати, що модель аналізу ставлення публічних осіб до теми є корисною для практичного використання. Таким чином створюється необхідний контекст аналізу із зазначенням часу, що дозволяє користувачу у зручній формі отримати підготовлену інформацію стосовно потрібної теми, а також суттєво спрощує процес прийняття рішень. Через дотримання високого рівня точності прогнозів

та можливості звернутись до першоджерела цитати система досягає умови точності, що підтверджує її оптимальність.

4.5 Висновки

У даному розділі реалізовану модель інформаційної системи інтелектуального аналізу настрою було використано для оброблення корпусу новинних статей з 3 ключових тем – зеленої енергетики, цензури у соціальних мережах та торговельної війни США. У результаті дослідження підібрані оптимальні способи їх представлення кінцевому користувачу. Було здійснено прототипування інтерфейсу віджету із зазначенням обов’язкових функцій. Використовуючи відібрані технології розроблено віджет, що задовольнив усі поставлені вимоги. За його допомогою оцінено результати роботи системи з точки зору кінцевого користувача. Для цього було опрацьовано великий обсяг даних та отримано відгук про якість формування висновків про позиції публічних осіб з теми. Позитивний користувацький досвід та дотримання високих показників точності оцінювання настрою моделлю аналізу дають змогу стверджувати про оптимізацію процесу прийняття рішень з умовою оптимальності у вигляді досягнення достатнього рівня точності.

ВИСНОВКИ

Дослідження бакалаврської кваліфікаційної роботи стосувалося процесу аналізу настрою публічних осіб з новинних публікацій для виявлення суперечностей у позиціях. Для цього потрібно підвищити точність оцінювання настрою публічних осіб до тематики з аналізу новинних публікацій за рахунок використання штучного інтелекту у складних системах підтримки прийняття рішень та забезпечити зручне представлення отриманих даних.

Аналіз предметної області моделювання ставлення публічних осіб до теми із відстежуванням протиріч досліджувався з оцінкою аналогів подібних інформаційних систем та конкретизацією області можливого використання програмного рішення. За отриманими вимогами визначилися задачі із чітким зазначенням необхідних функціональних та нефункціональних вимог до розроблюваної моделі.

Сформований концепт і конкретні вимоги до моделі зумовили відбір оптимальних методів розв’язання задачі аналізу настрою риторичних публічних осіб, а саме технологію інтелектуального дослідження тексту з метою отримання висновків щодо позиції особи за обраною темою та зазначено спосіб досягнення необхідних результатів з її використанням. Це дозволило розробити модель аналітичної інформаційної системи, що відслідковує ставлення осіб до теми та виявляє протиріччя у позиціях. Основою інтелектуальної частини є фреймворк LangGraph та велика мовна модель, що запускається локально через LM Studio. У ході проведеного аналізу обрано найкращі LLM для виконання поставленої задачі, що використовуються як гіперпараметр моделі аналізу настрою. Візуалізаційна частина реалізована за допомогою React Native та бібліотеки Vite.

Розроблена підсистема аналізу настрою до теми є конвеєром інтелектуальної обробки настрою публічних осіб, що формує спроектовану модель аналізу. Оптимізація загальних гіперпараметрів системи сприяє максимізації точності автоматизованого інтелектуального аналізу. Це включає в себе підбір формул та способів обрахунку базових показників, коригування переданого контексту та вибір оптимальної великої мовної моделі.

Дослідженням встановлено, що період актуальності висловлювання становить максимум 1 рік, при чому вплив старіших думок на поточне ставлення спадає лінійно від дати публікації. При цьому на фінальну цікавість висловлювання найбільше впливають релевантність (68,19%), суперечливість (13,928%) та внесок (10,861%), а його порогове значення для відображення становить 66.5%. Також проведене експертне оцінювання на тестовій вибірці із 50 статей із порівняльним аналізом великих мовних моделей дозволило обрати оптимальну конфігурацію. LLM-агенти були оцінені за кількома критеріями, після чого обрано найкращу із можливих моделей для вирішення поставленої задачі. Так, найбільш оптимальним рішенням є використання gemma-3-4b-it-qat, яка показала найкращі результати за f1-score (0.68) та точністю (0.69), при цьому вона найкраще спрогнозувала класи за ROC AUC, а відхилення в оцінці відносились переважно до сусідніх класів шкали. З цього можна зробити висновок про найбільшу загальну точність аналізу. Коефіцієнт детермінації становить 0.69, а загальна якість тексту за 10-бальною шкалою дорівнює 9.04, що свідчить про достатній рівень розуміння даних моделлю та загальну високу якість відповіді. Це у свою чергу підтверджує точність автоматизованого аналізу, що задовольняє умову оптимальності. За можливості використання двох моделей у паралелі для роботи із генерацією тексту варто використовувати qwen3-4b, оскільки він показав вищий результат якості тексту (9.73), але в цілому є гіршим. За результатами досліджено чітко сформульовані параметри, використання яких дозволяє досягти оптимального результату під час використання моделі інформаційної системи інтелектуального аналізу настроїв.

Реалізована модель інформаційної системи інтелектуального аналізу настроїв є основою для оброблення корпусу новинних статей з 3 ключових тем – зеленої енергетики, цензури у соціальних мережах та торговельної війни США. У результаті дослідження підбираються оптимальні способи їх представлення кінцевому користувачу та здійснюється прототипування інтерфейсу віджету із зазначенням обов'язкових функцій. Розроблена на основі обраних технологій підсистема візуалізації настроїв публічних осіб

задовольняє усі поставлені вимоги. За його допомогою можна оцінити результати роботи системи з точки зору кінцевого користувача. Для цього опрацьовується великий обсяг даних і формується відгук про якість генерування висновків щодо позицій публічних осіб з досліджуваного елементу суспільного дискурсу. Позитивний користувацький досвід та дотримання високих показників точності оцінювання настрою моделлю аналізу дають змогу стверджувати про оптимізацію процесу прийняття рішень з умовою оптимальності у вигляді досягнення високого рівня точності.

Результати виконаної роботи були впроваджені на підприємстві ТОВ «Кваліа Консалтинг Едженси», зокрема, реалізовану інформаційну підсистему настрою-аналізу інформаційних потоків та аналітичний віджет інтегровано в існуючу інфраструктуру.

За результатами даної роботи зроблено доповідь на тему «Аналіз змін у позиціях публічних осіб на основі новинних публікацій» на LIV Всеукраїнській науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету (2025) з публікацією тез [1].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Демков І.В., Козачко О.М. Аналіз змін у позиціях публічних осіб на основі новинних публікацій. *Всеукраїнська науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету (Вінниця, 2025 р.)*». URL: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2025/paper/view/24134> (дата звернення: 03.06.2025).
2. Alieinyk V. Factors of the growing role of mass communication as a tool for shaping political consciousness in the modern world. *Society and Security*. 2024. № 4(4). С. 3–13. URL: [https://doi.org/10.26642/sas-2024-4\(4\)-3-13](https://doi.org/10.26642/sas-2024-4(4)-3-13) (дата звернення: 03.06.2025).
3. Goyal T., Jessy Li J., Durrett G.. News Summarization and Evaluation in the Era of GPT-3. *arXiv*. 2023. No. 2209.12356. URL: <https://doi.org/10.48550/arXiv.2209.12356> (дата звернення 03.06.2025)
4. A. Jawad Z., J. Obaid A. Combination Of Convolution Neural Networks And Deep Neural Networks For Fake News Detection. *arXiv*. 2022. №2210.08331. URL: <https://doi.org/10.48550/arXiv.2210.08331> (дата звернення: 03.06.2025).
5. Wu J., Guo J., Hooi B. Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks. *KDD '24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, м. Barcelona Spain. New York, NY, USA, 2024. С. 3367–3378. URL: <https://doi.org/10.1145/3637528.3671977> (дата звернення: 03.06.2025).
6. Кулаженко В., Мазур О. Дослідження алгоритму побудови моделі сентимент аналізу повідомлень у соціальних мережах. *Herald of Khmelnytskyi National University. Technical sciences*. 2024. Т. 333, № 2. С. 421–427. URL: <https://doi.org/10.31891/2307-5732-2024-333-2-66> (дата звернення: 02.06.2025).
7. Batiuk T., Dosyn D. Implementation of the intellectual system of sentiment analysis and clusterization of publications in the twitter social network. *Innovative Technologies and Scientific Solutions for Industries*. 2023. № 1 (23). С. 25–44. URL: <https://doi.org/10.30837/itssi.2023.23.025> (дата звернення: 03.06.2025).

8. Event registry / G. Leban та ін. *the 23rd International Conference*, м. Seoul, Korea, 7–11 квіт. 2014 р. New York, New York, USA, 2014. URL: <https://doi.org/10.1145/2567948.2577024> (дата звернення: 03.06.2025).
9. Leetaru K., Schrod P. GDELT: Global data on events, location, and tone. *ISA Annual Convention*. 2013. URL: <http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf> (дата звернення: 03.06.2025).
10. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities / D. Antons та ін. *R&D Management*. 2020. Т. 50, №3. С. 329–351. URL: <https://doi.org/10.1111/radm.12408> (дата звернення: 03.06.2025).
11. Text Mining: Techniques, Applications and Issues / R. Talib та ін. *International Journal of Advanced Computer Science and Applications*. 2016. Т. 7, №11. URL: <https://doi.org/10.14569/ijacsa.2016.071153> (дата звернення: 03.06.2025).
12. ALDayel A., Magdy W. Stance detection on social media: State of the art and trends. *Information Processing & Management*. 2021. Т. 58, № 4. С. 102597. URL: <https://doi.org/10.1016/j.ipm.2021.102597> (дата звернення: 03.06.2025).
13. Nadkarni P.M., Ohno-Machado L., Chapman W.W. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*. 2011. Т. 18, №5. С. 544–551. URL: <https://doi.org/10.1136/amiajnl-2011-000464> (дата звернення: 03.06.2025).
14. Haque Khan Tusar T., Islam T. A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data. *arXiv*. 2021. №2110.00859. URL: <https://doi.org/10.48550/arXiv.2110.00859> (дата звернення: 03.06.2025).
15. Singh A.K., Shashi M. Vectorization of Text Documents for Identifying Unifiable News Articles. *International Journal of Advanced Computer Science and Applications*. 2019. Т. 10, № 7. URL: <https://doi.org/10.14569/ijacsa.2019.0100742> (дата звернення: 03.06.2025).
16. Smith B. A Complete Guide to BERT with Code. *towardsdatascience.com*.

URL: <https://towardsdatascience.com/a-complete-guide-to-bert-with-code-9f87602e4a11/> (дата звернення: 03.06.2025).

17. Молчанова М. О., Залуцька О. О., Бармак О. В. Метод інтелектуального аналізу тональності текстів. Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання». Київ, 2023. С. 113-116.

18. TnT-LLM: Text Mining at Scale with Large Language Models / M. Wantha і ін. *KDD '24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, м. Barcelona Spain. New York, NY, USA, 2024. С. 5836–5847. URL: <https://doi.org/10.1145/3637528.3671647> (дата звернення: 03.06.2025).

19. Determination of Context Window Size / K.Y. HUNG та ін. *International Journal of Computer Processing of Languages*. 2001. Т. 14, №01. С. 71–80. URL: <https://doi.org/10.1142/s0219427901000291> (дата звернення: 03.06.2025).

20. Guido van Rossum, Fred L. Drake, Jr. The Python Tutorial URL: <https://docs.python.org/3/tutorial/> (дата звернення: 03.06.2025).

21. Wang J., Duan Z. Agent AI with LangGraph: A Modular Framework for Enhancing Machine Translation Using Large Language Models. *arXiv*. 2024. №2412.03801. URL: <https://doi.org/10.48550/arXiv.2412.03801> (дата звернення: 03.06.2025).

22. LLM Arena Leaderboard. URL: <https://huggingface.co/spaces/lmarena-ai/chatbot-arena-leaderboard> (дата звернення: 03.06.2025).

23. Chatbot Arena. URL: <https://openlm.ai/chatbot-arena/> (дата звернення: 03.06.2025).

24. LM Studio. URL: <https://lmstudio.ai/> (дата звернення: 03.06.2025)

25. CNN. URL: <https://edition.cnn.com/> (дата звернення: 03.06.2025)

Додаток А
(обов'язковий)

Міністерство освіти і науки України
Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

_____ д.т.н., проф. Віталій МОКІН

“ ____ ” _____ 2025 р.

ТЕХНІЧНЕ ЗАВДАННЯ
на бакалаврську кваліфікаційну роботу
СЕНТИМЕНТ-АНАЛІЗ НОВИН ДЛЯ ОЦІНЮВАННЯ ЗМІН У РИТОРИЦІ
ПУБЛІЧНИХ ОСІБ ТА ВИЯВЛЕННЯ СУПЕРЕЧНОСТЕЙ ЗА ДОПОМОГОЮ
ШТУЧНОГО ІНТЕЛЕКТУ. ПІДСИСТЕМА ВИЗНАЧЕННЯ ТА ВІЗУАЛІЗАЦІЇ
СЕНТИМЕНТУ

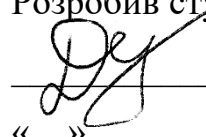
08-34.БКР.003.00.000 ТЗ

Керівник: к.т.н., доц.

_____ Олексій КОЗАЧКО

« ____ » _____ 2025 р.

Розробив студент гр. СА-216

 _____ Іван ДЕМКОВ

« ____ » _____ 2025 р.

1. Підстава для проведення робіт.

Підставою для виконання роботи є наказ №__ по ВНТУ від «__» _____2025р., та індивідуальне завдання на БКР, затверджене протоколом №__ засідання кафедри САІТ від «__» _____ 2025р.

2. Джерела розробки:

- 1) CNN. URL: <https://edition.cnn.com/> (дата звернення: 03.06.2025)
- 2) Goyal T., Jessy Li J., Durrett G.. News Summarization and Evaluation in the Era of GPT-3. arXiv. 2023. No. 2209.12356. URL: <https://doi.org/10.48550/arXiv.2209.12356> (дата звернення 03.06.2025)
- 3) ALDayel A., Magdy W. Stance detection on social media: State of the art and trends. Information Processing & Management. 2021. T. 58, № 4. С. 102597. URL: <https://doi.org/10.1016/j.ipm.2021.102597> (дата звернення: 03.06.2025).

3. Мета і призначення роботи.

Метою дослідження є підвищення точності оцінювання сентименту публічних осіб до тематики з аналізу новинних публікацій за рахунок використання штучного інтелекту, що дозволяє виявити суперечності у позиціях для використання у складних системах підтримки прийняття рішень.

4. Вихідні дані для проведення робіт:

Дані висловлювань публічних осіб із новинного веб-сайту CNN.

5. Методи дослідження:

Огляд наукових публікацій за темою, дослідження існуючих інформаційних систем, моделювання процесу аналізу ставлення до теми із відстежуванням протиріч, оптимізація параметрів моделі, формалізація визначень базових понять, реалізація аналітичного конвеєру, дослідження способу візуалізації даних аналізу, прототипування інтерфейсу віджету, реалізація візуалізаційної частини, оцінювання результатів роботи системи аналізу.

6. Етапи роботи і терміни їх виконання:

- | | | | |
|---|-------|---|-------|
| a) Аналіз предметної області | _____ | — | _____ |
| b) Порівняльний аналіз проблематики | _____ | — | _____ |
| c) Вибір оптимальних інформаційних технологій | _____ | — | _____ |
| d) Розроблення моделі аналізу риторики публічних осіб | _____ | — | _____ |
| e) Оформлення матеріалів до захисту БКР | _____ | — | _____ |

7. Очікувані результати та порядок реалізації

Отримання інформаційної системи автоматизованого аналізу новин із визначенням ставлення публічної особи до теми та відслідковуванням протиріч з минулими твердженнями та модуль інтерактивного представлення результатів.

8. Вимоги до розробленої документації

Текстова та ілюстративна частини роботи оформлені у відповідності до вимог «Методичних вказівок до виконання бакалаврських кваліфікаційних робіт для студентів спеціальностей: 124 «Системний аналіз», 126 «Інформаційні системи та технології» (освітня програма «Прикладні інформаційні технології»)).

9. Порядок приймання роботи

Публічний захист

Початок розробки

Граничні терміни виконання БКР

«__» _____ 2025 р.

«__» _____ 2025 р.

«__» _____ 2025 р.

Розробив студент групи СА-216

 Іван ДЕМКОВ

Додаток Б
(обов'язковий)

ПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Назва роботи: «Сентимент-аналіз новин для оцінювання змін у риторичності публічних осіб та виявлення суперечностей за допомогою штучного інтелекту. Підсистема визначення та візуалізації сентименту»

Тип роботи: бакалаврська кваліфікаційна робота

Підрозділ: кафедра САІТ, ФІТА, гр. СА-216

Коефіцієнт подібності текстових запозичень, виявлених у роботі
системою StrikePlagiarism 0,79 %

Висновок щодо перевірки кваліфікаційної роботи (відмітити потрібне):

- ☒ Запозичення, виявлені у роботі, є законними і не містять ознак плагіату, фабрикації, фальсифікації. Роботу прийняти до захисту
- ☐ У роботі не виявлено ознак плагіату, фабрикації, фальсифікації, але надмірна кількість текстових запозичень та/або наявність типових розрахунків не дозволяють прийняти рішення про оригінальність та самостійність її виконання. Роботу направити на доопрацювання.
- ☐ У роботі виявлено ознаки плагіату та/або текстових маніпуляцій як спроб укриття плагіату, фабрикації, фальсифікації, що суперечить вимогам законодавства та нормам академічної доброчесності. Робота до захисту не приймається.

Експертна комісія:

Віталій МОКІН, зав. каф. САІТ

(підпис)

Сергій ЖУКОВ, доц. каф. САІТ

(підпис)

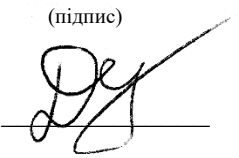
Особа, відповідальна за перевірку _____
(підпис)

Сергій ЖУКОВ

З висновком експертної комісії ознайомлений(-на)

Керівник _____
(підпис)

Олексій КОЗАЧКО, к.т.н., доц. каф. САІТ

Здобувач  _____

Іван ДЕМКОВ

(підпис)

Додаток В
(довідниковий)
Фрагмент лістингу програми

Код файлу аналітичної системи graph_config.py

```
from langgraph.graph import StateGraph, START, END

from nodes import write_event, event_classifier, event_scoring, sentiment_analysis, inconsistency_detection,
person_summary, opinion_scoring

from .states_setup import ChunkState

from typing import List

def create_app():

    graph = StateGraph(state_schema=ChunkState)

    graph.add_node("extract_events", event_classifier.main)

    # Opinion processing
    graph.add_node("sentiment_analysis", sentiment_analysis.main)
    graph.add_node("inconsistency_detection", inconsistency_detection.main)
    graph.add_node("person_summary", person_summary.main)
    graph.add_node("scoring_system_person", opinion_scoring.main)

    # General event processing
    graph.add_node("scoring_system_general", event_scoring.main)

    graph.add_node("write_events_to_db", write_event.main)

    graph.add_edge(START, "extract_events")

    def router_extract_events(state: ChunkState) -> List[str]:

        routes = []

        if state.general_event:

            routes.append("scoring_system_general")

        if state.person_event:
```

```

        routes.append("sentiment_analysis")

    if not routes:
        routes.append(END)

    return routes

graph.add_conditional_edges(
    "extract_events",
    router_extract_events,
    {
        "scoring_system_general": "scoring_system_general",
        "sentiment_analysis": "sentiment_analysis",
        END: END
    }
)

# Opinion processing
graph.add_edge('sentiment_analysis', 'inconsistency_detection')
graph.add_edge('inconsistency_detection', 'person_summary')
graph.add_edge('person_summary', 'scoring_system_person')

graph.add_edge(['scoring_system_person', 'scoring_system_general'], 'write_events_to_db')

graph.add_edge("write_events_to_db", END)

return graph.compile()

```

Код файлу аналітичної системи model.py

```

import logging, os, sys

from langchain_openai.chat_models import ChatOpenAI

sys.path.append(os.path.dirname(os.path.dirname(os.path.abspath(__file__))))

from config import USE_TWO_LLMS

llm_main = ChatOpenAI(openai_api_base="http://127.0.0.1:1234/v1",
                       model='gemma-3-4b-it-qat',

```

```

        temperature=0.1)

if USE_TWO_LLMS:
    llm_text = ChatOpenAI(openai_api_base="http://127.0.0.1:1234/v1",
                           model='qwen3-4b',
                           temperature=0.1)

# Configure logging
log_dir = "./logs"
os.makedirs(log_dir, exist_ok=True) # Ensure the logs directory exists
log_file = os.path.join(log_dir, "model_usage.log")

logging.basicConfig(
    filename=log_file,
    level=logging.INFO,
    format="%(asctime)s - %(message)s",
    datefmt="%Y-%m-%d %H:%M:%S"
)

def llm_invoke(*args, usage_type="unspecified", model='main', **kwargs):
    """
    Invoke the language model with logging of token usage and purpose.

    Args:
        *args: Arguments to pass to the model's invoke method
        usage_type: String describing what the LLM is being used for (e.g. "event_scoring", "sentiment_analysis")
        model: The model to use ('main' for llm_main, 'text' for llm_text)
        **kwargs: Keyword arguments to pass to the model's invoke method

    Returns:
        The model's response
    """
    if model == 'main':
        response = llm_main.invoke(*args, **kwargs)

```

```

elif model == 'text':

    # Check if llm_text is declared, else use llm_main

    if 'llm_text' in globals() and llm_text is not None:

        response = llm_text.invoke(*args, **kwargs)

    else:

        response = llm_main.invoke(*args, **kwargs)


# Extract token usage from response metadata

token_usage = response.response_metadata.get('token_usage', {})

total_tokens = token_usage.get('total_tokens', 0)

prompt_tokens = token_usage.get('prompt_tokens', 0)

completion_tokens = token_usage.get('completion_tokens', 0)


# Log the token usage with usage type

logging.info(f"{usage_type} - total_tokens: {total_tokens}, prompt_tokens: {prompt_tokens}, completion_tokens: {completion_tokens}")


return response

```

Код файлу аналітичної системи prompts.py

```
REFINING_PROMPT = "Refine the following text to fit within {max_length} characters:\n\n{field_value} /no_think"
```

```
EVENT_CLASSIFICATION_PROMPT = """
```

You are an expert journalist assistant. Your task is to extract:

1. Person events related to statements or actions regarding "{topic}" from the article.
2. General events related to the topic "{topic}" from the article.

Ensure the following constraints:

- The "title" of each general event must not exceed 50 characters.
- The "description" of each general event must not exceed 200 characters.
- The "person_name" must not exceed 150 characters.
- The "person_name" should be the name of a single person; if multiple people are present, create separate person events.

- If "person_name" is unknown, use "Unspecified".
- The "citation" must not exceed 300 characters.

Without superfluous information, extract only the most important details. Provide only the answer, do not include any additional text or explanations.

Return a JSON object in the following format:

```
{
  "person_events": [
    {
      "person_name": "...",
      "citation": "..."
    },
    {
      "person_name": "...",
      "citation": "..."
    }
  ],
  "general_events": [
    {
      "title": "...",
      "description": "..."
    },
    {
      "title": "...",
      "description": "..."
    }
  ]
}
```

If no person events are found, use an empty array for "person_events".

If no general events are found, use an empty array for "general_events".

If no events of either type are found, return:

```
{
  "person_events": [],
  "general_events": []
}
```

```
}}
```

Article:

```
{chunk}
```

```
/no_think
```

```
"""
```

```
SENTIMENT_ANALYSIS_PROMPT = """
```

You are a debate and sentiment analysis expert. Analyse positive or negative commitment to the topic "{topic}" in the following text.

Given the following topic and text, return **only** a single floating-point number between -1.0 and +1.0, where:

- -1.0 indicates strongly negative sentiment
- 0.0 indicates neutral sentiment
- +1.0 indicates strongly positive sentiment

Topic: {topic}

Text:

```
{content}
```

Respond with the number alone (e.g. "-0.75").

```
"""
```

```
INCONSISTENCY_COMMENT_PROMPT = """
```

You are an expert journalist skilled in deduction and speech analysis.

Given the following inputs, compare the new citation to previous ones and comment on any inconsistency.

Person: {person}

Topic: {topic}

New citation:

- Text: {new_citation}
- Date: {new_date}
- Score: {new_score}

Previous citations:

```
{previous_formatted}
```

Write **only one paragraph**, max **200 characters**, pointing out the sentiment inconsistency. Do not include quotes or metadata—just the concise comment.

/no_think

"""

PERSON_SUMMARY_PROMPT = """

You are an expert journalist proficient in analyzing political stances. Given the following inputs, produce **only** the final person_summary text (max 300 characters) that:

1. Summarizes the person's current position on the topic.
2. Notes their expert status and openness to change. If this is unknown, do not mention it.
3. Highlights any recent contradiction, if present.
4. Integrates key details from the new citation and, if available, the previous summary.

If a previous summary exists, should make an updated version of it.

Respond with no bullet points, no options, no additional commentary—just the **single person_summary string** of given length. Do **not** start with typical phrases like "Here's a summary under 300 characters:", start already with option text. Output **only** the most full option's text.

Inputs:

Person: {person}

Topic: {topic}

Citation: {citation}

Stance: {stance}

Tendency to change stance: {sentiment_deviation}

{{#if previous_summary}}

Previous summary: {previous_summary}

{{/if}}

{{#if inconsistency}}

Recently found inconsistency: {inconsistency}

{{/if}}

Context:

{content}

/no_think

"""

OPINION_SCORING_PROMPT = """

You are an expert journalist proficient in analyzing political stances. Given the following inputs, score the opinion (citation) on the topic.:

Make 3 scores:

1. Relevancy: 0-1 floating-point number (how relevant the opinion is to the topic)
2. Contribution: 0-1 floating-point number (how much the opinion contributes to the topic)
3. Controversy: 0-1 floating-point number (how controversial the opinion is)

Use the following scale, be as precise as possible:

- 0.0 = completely irrelevant / no contribution / not controversial at all
- 0.5 = moderately relevant / contributes somewhat / somewhat controversial
- 1.0 = absolutely relevant / major contribution / extremely controversial

Be conservative in assigning high values; only use scores ≥ 0.7 when you are highly confident. If in doubt, err on the side of a lower score.

Provide ****only**** the scores in the following format:

```
{{
  "relevancy": <relevancy>,
  "contribution": <contribution>,
  "controversy": <controversy>
}}
```

Inputs:

Topic: {topic}

Person: {person}

Person summary: {person_summary}

Tendency to change stance: {sentiment_deviation}

Citation: {citation}

Recent inconsistencies found: {inconsistency_flag}

{{#if inconsistency}}

Found inconsistency: {inconsistency}

{{/if}}

"""

EVENT_SCORING_PROMPT = """

You are an expert journalist proficient in analyzing events. Given the following inputs, score the general event in the topic of discussion:

Make 3 scores:

1. Relevance: 0-1 floating-point number (how relevant the event is to the topic)
2. Influence: 0-1 floating-point number (how much the event influences the topic)
3. Novelty: 0-1 floating-point number (how novel the event is)

Use the following scale for each score, be as precise as possible:

- 0.0 = completely irrelevant / no influence / not novel at all
- 0.5 = moderately relevant / has some influence / somewhat novel
- 1.0 = absolutely relevant / major influence / highly novel

Be conservative when assigning high values; only use scores ≥ 0.7 when you are highly confident. If in doubt, default to a lower score.

Provide **only** the scores in the following format:

```
{{
  "relevance": <relevance>,
  "influence": <influence>,
  "novelty": <novelty>
}}
```

Inputs:

Topic: {topic}

Event title: {title}

Event description: {description}

Article chunk content (for context): {content}

""""

Код файлу аналітичної системи states_setup.py

```
from pydantic import BaseModel, Field
from typing import Optional, List, Annotated
from operator import add
import json
from ast import literal_eval
from pandas import Timestamp

# Define custom reducers for different merge strategies
def keep_latest(old_val, new_val):
```

```

"""Replace old value with new value"""
return new_val

def merge_unique_events(old_events, new_events):
    """Merge events lists, keeping the most complete version of each event"""
    if not old_events:
        return new_events
    if not new_events:
        return old_events

    # Create a dictionary to store the most complete version of each event
    # Use a tuple of key identifying fields as the dictionary key
    event_dict = {}

    # Helper function to create a key for each event
    def get_event_key(event):
        if 'person_name' in event and 'citation' in event:
            # For person events
            return ('person', event['person_name'], event['citation'])
        elif 'title' in event and 'description' in event:
            # For general events
            return ('general', event['title'], event['description'])
        else:
            # Fallback for other event types
            return ('other', str(event))

    # Process old events first
    for event in old_events:
        key = get_event_key(event)
        event_dict[key] = event

    # Process new events, keeping the version with more fields when there's a conflict
    for event in new_events:
        key = get_event_key(event)
        if key in event_dict:

```

```

        # If this event already exists, keep the one with more fields
        old_event = event_dict[key]
        if len(event) > len(old_event):
            event_dict[key] = event
        else:
            # New unique event
            event_dict[key] = event

    # Return the list of most complete events
    return list(event_dict.values())

def save_state_as_json(state, filename):
    """
    Save the given state as a JSON file.

    Args:
        state (dict): The state to save.
        filename (str): The name of the file to save the state to.
    """
    with open(filename, 'w') as f:
        # Option 1: Use model_dump() to get a Python dict directly
        json.dump(state.model_dump(), f, indent=4, default=str)

def save_output_as_json(state, filename):
    """
    Save the given state as a JSON file.

    Args:
        state (dict): The state to save.
        filename (str): The name of the file to save the state to.
    """
    with open(filename, 'w') as f:
        json.dump(dict(state.items()), f, indent=4, default=str)

def load_state_from_json(filename):

```

"""

Load a state from a JSON file and convert it to a ChunkState object.

Args:

filename (str): The path to the JSON file to load.

Returns:

ChunkState: The loaded state as a ChunkState object.

"""

with open(filename, 'r') as f:

data = json.load(f)

Handle timestamp conversion if needed

if 'article_date' in data and data['article_date'] is not None:

data['article_date'] = Timestamp(data['article_date'])

Create a ChunkState object from the loaded data

return ChunkState(**data)

Full chunk state with all fields

class ChunkState(BaseModel):

model_config = {"arbitrary_types_allowed": True}

chunk_id: Annotated[int, keep_latest] = Field(description="Unique identifier for the chunk")

topic: Annotated[str, keep_latest]

topic_id: Annotated[Optional[int], keep_latest]

origin_article_id: Annotated[Optional[int], keep_latest] = None

article_date: Annotated[Optional[Timestamp], keep_latest] = None

content: Annotated[str, keep_latest]

general_event: Annotated[Optional[List[dict]], merge_unique_events] = Field(default_factory=list)

person_event: Annotated[Optional[List[dict]], merge_unique_events] = Field(default_factory=list)

is_processed: Annotated[int, add] = 0

Parser setup

```
class GeneralEvent(BaseModel):
```

```
    title: Optional[str] = Field(description="Title of the general event.", max_length=50)

    description: Optional[str] = Field(description="Brief summary of the event, limited to 200 characters.",
max_length=200)
```

```
class PersonEvent(BaseModel):
```

```
    person_name: Optional[str] = Field(description="Name of the person involved in the event.", max_length=150)

    citation: Optional[str] = Field(description="Key statement or citation related to the person and the topic.",
max_length=300)
```

```
class EventClassification(BaseModel):
```

```
    general_event: Optional[GeneralEvent]

    person_event: Optional[PersonEvent]
```

```
class SentimentScore(BaseModel):
```

```
    sentiment: float = Field(description="Sentiment score of the opinion.", ge=-1, le=1)
```

```
class InconsistencyComment(BaseModel):
```

```
    inconsistency_comment: str = Field(description="Comment explaining the inconsistency.", max_length=200)
```

```
class PersonSummary(BaseModel):
```

```
    person_summary: str = Field(description="Summary of the person's stance on topic.", max_length=300)
```

```
class OpinionScores(BaseModel):
```

```
    relevancy: float = Field(description="Relevancy score from 0 to 1", ge=0, le=1)

    contribution: float = Field(description="Contribution score from 0 to 1", ge=0, le=1)

    controversy: float = Field(description="Controversy score from 0 to 1", ge=0, le=1)
```

```
class EventScores(BaseModel):
```

```
    relevance: float = Field(description="Relevance score from 0 to 1")

    influence: float = Field(description="Influence score from 0 to 1")

    novelty: float = Field(description="Novelty score from 0 to 1")
```

Додаток Г
(обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

СЕНТИМЕНТ-АНАЛІЗ НОВИН ДЛЯ ОЦІНЮВАННЯ ЗМІН У РИТОРИЦІ
ПУБЛІЧНИХ ОСІБ ТА ВИЯВЛЕННЯ СУПЕРЕЧНОСТЕЙ ЗА ДОПОМОГОЮ
ШТУЧНОГО ІНТЕЛЕКТУ. ПІДСИСТЕМА ВИЗНАЧЕННЯ ТА ВІЗУАЛІЗАЦІЇ
СЕНТИМЕНТУ

Нормоконтроль: к.т.н., доцент

_____ Сергій ЖУКОВ

«___» _____ 2025 р.

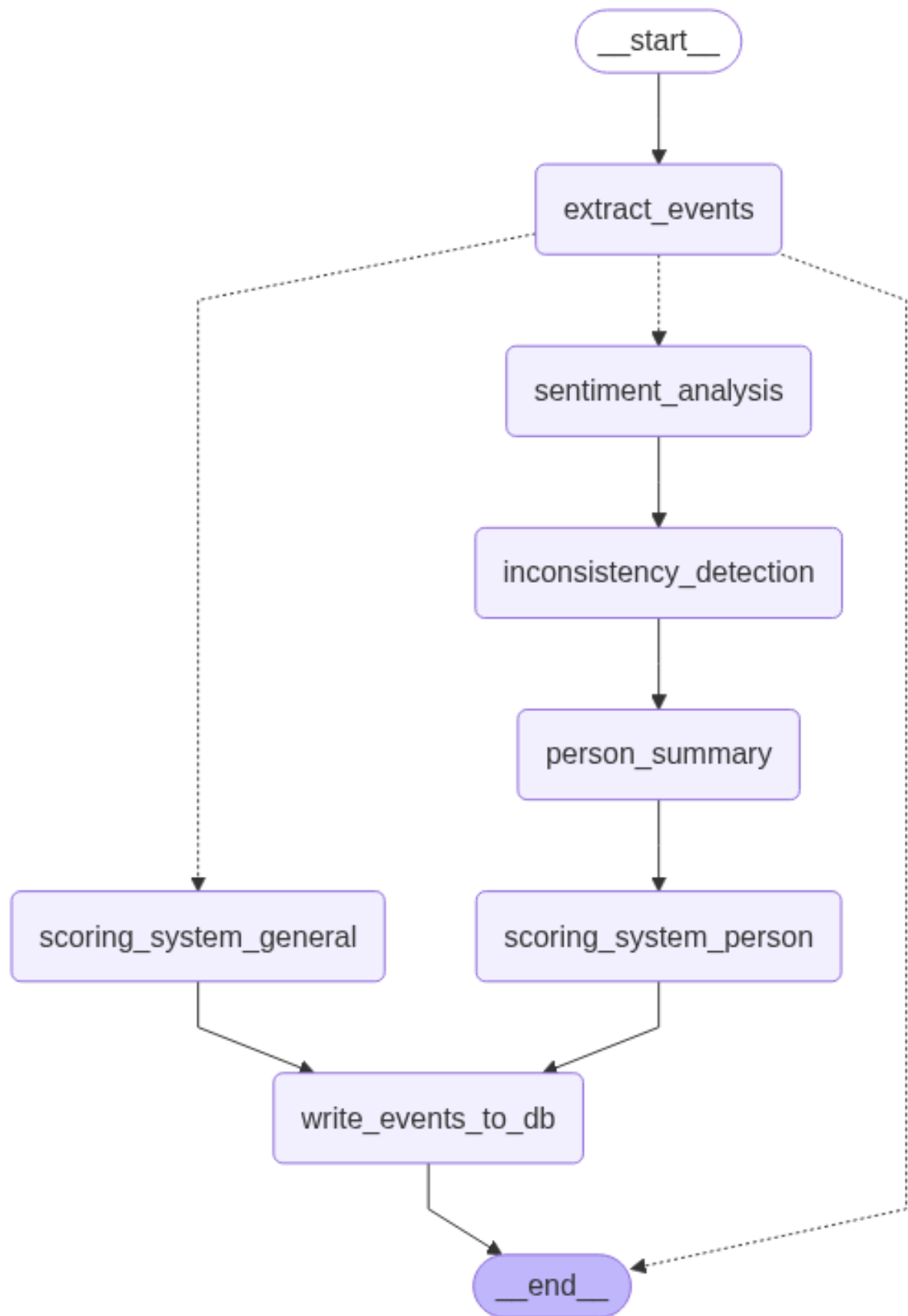


Рисунок Г.1 – Діаграма потоку даних конвеєру аналізу настрою публічних осіб

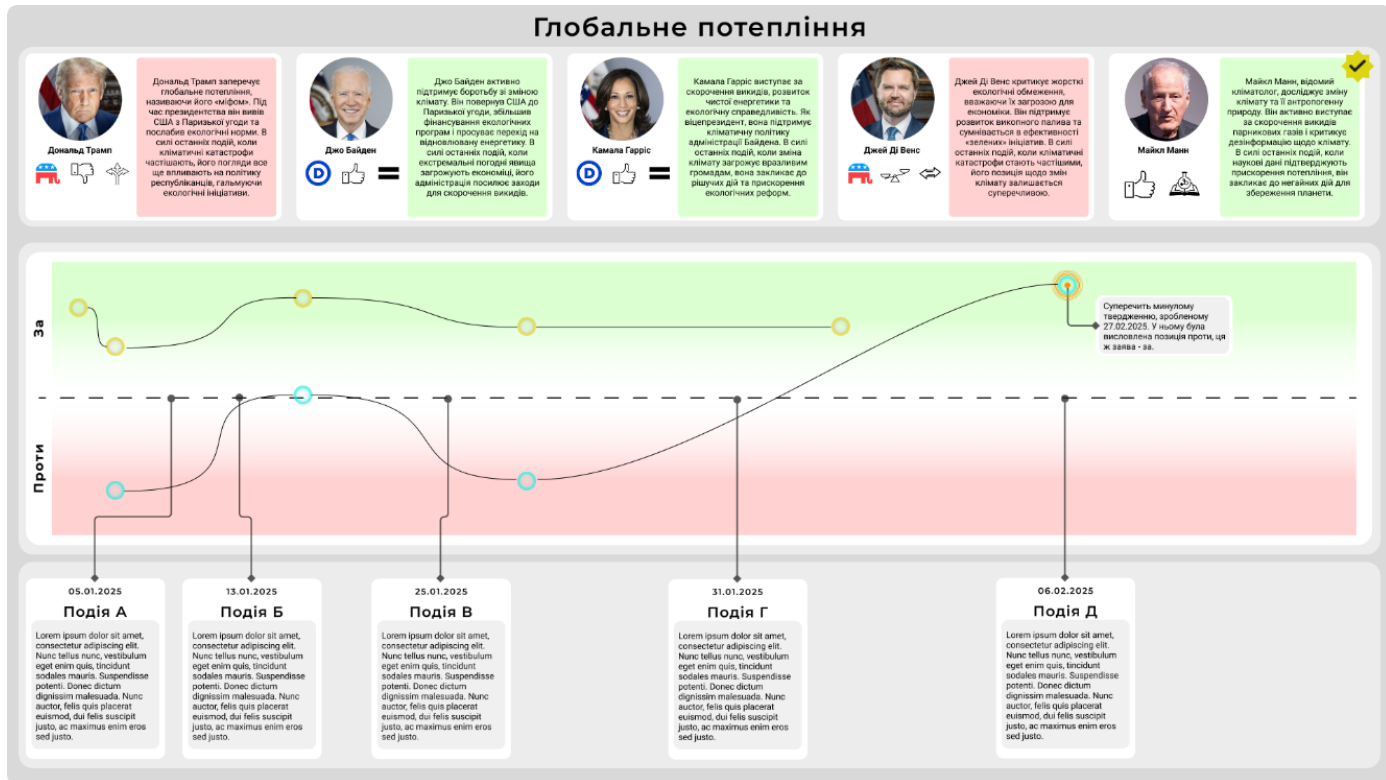


Рисунок Г.2 – Прототип візуалізації аналізованої інформації

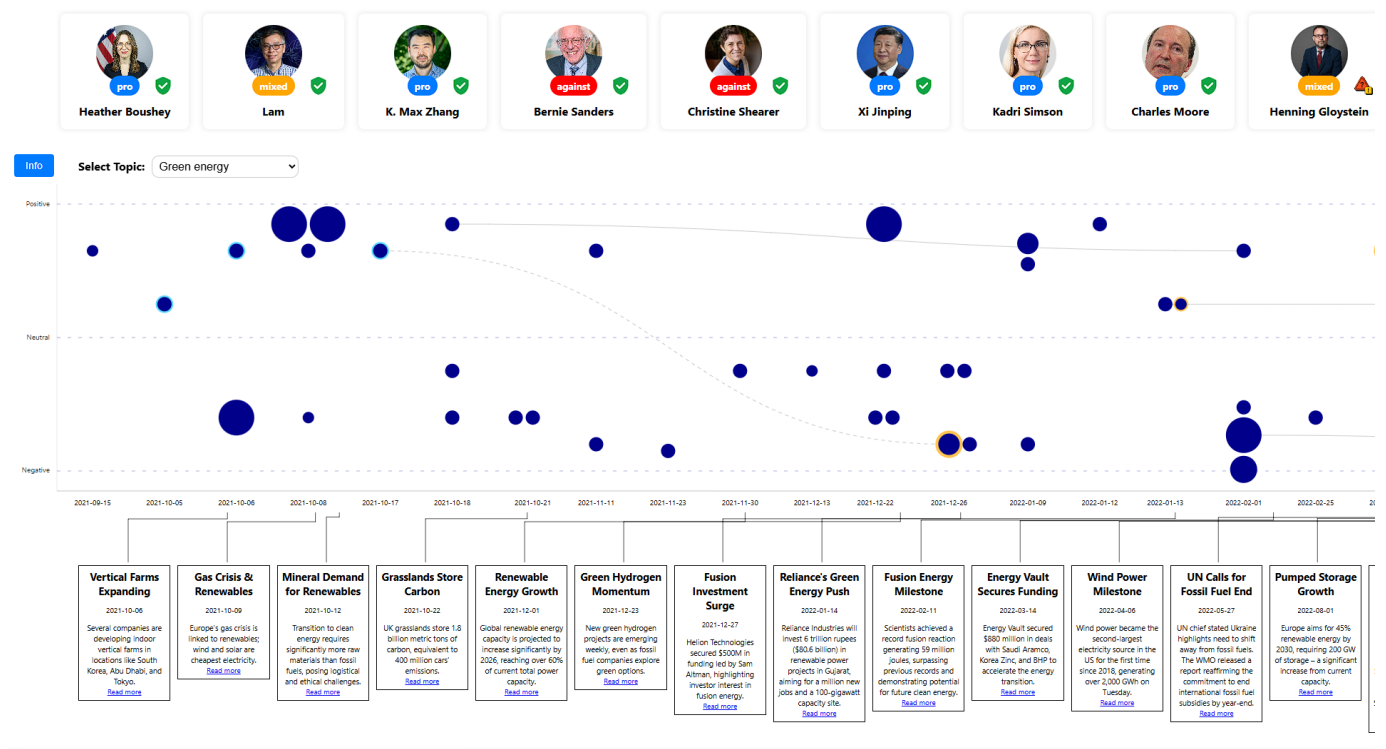


Рисунок Г.3 – Результати роботи системи аналізу настроїв публічних осіб стосовно теми зеленої енергетики (частина)

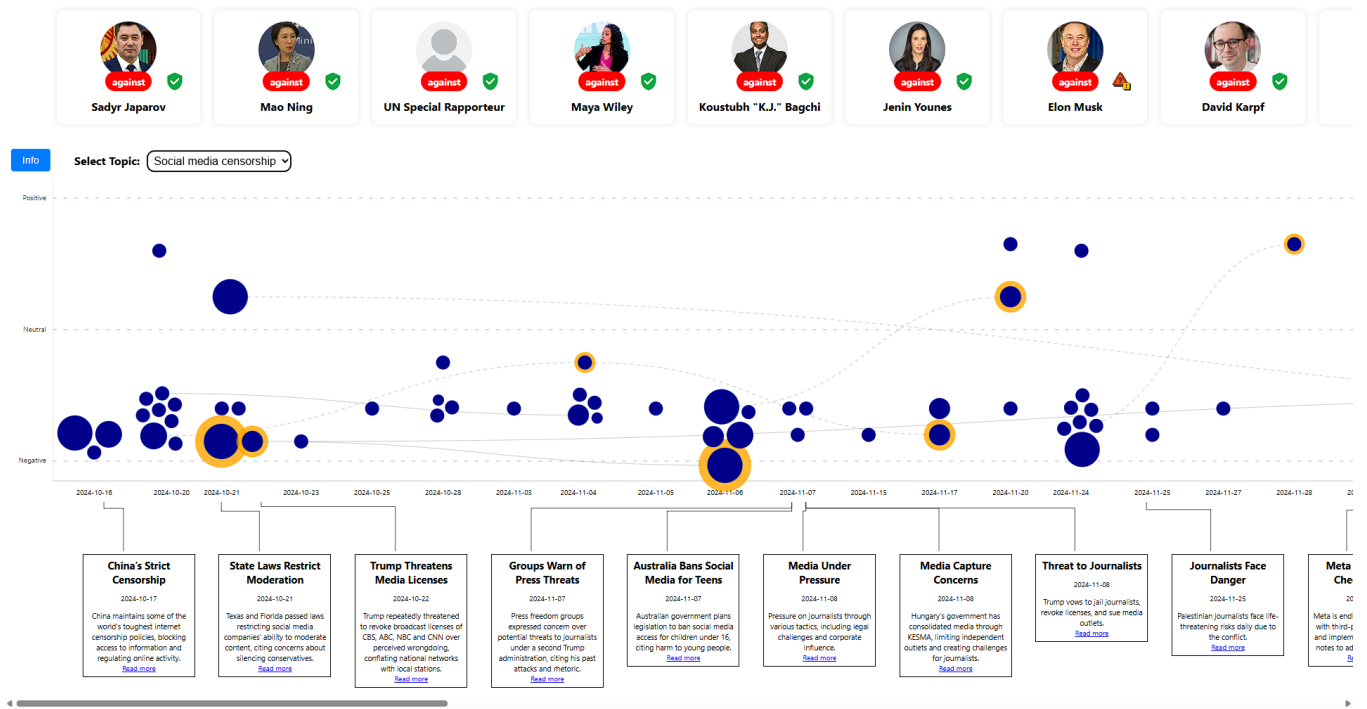


Рисунок Г.4 – Результати роботи системи аналізу настроїв стосовно теми цензури у соціальних мережах (частина)

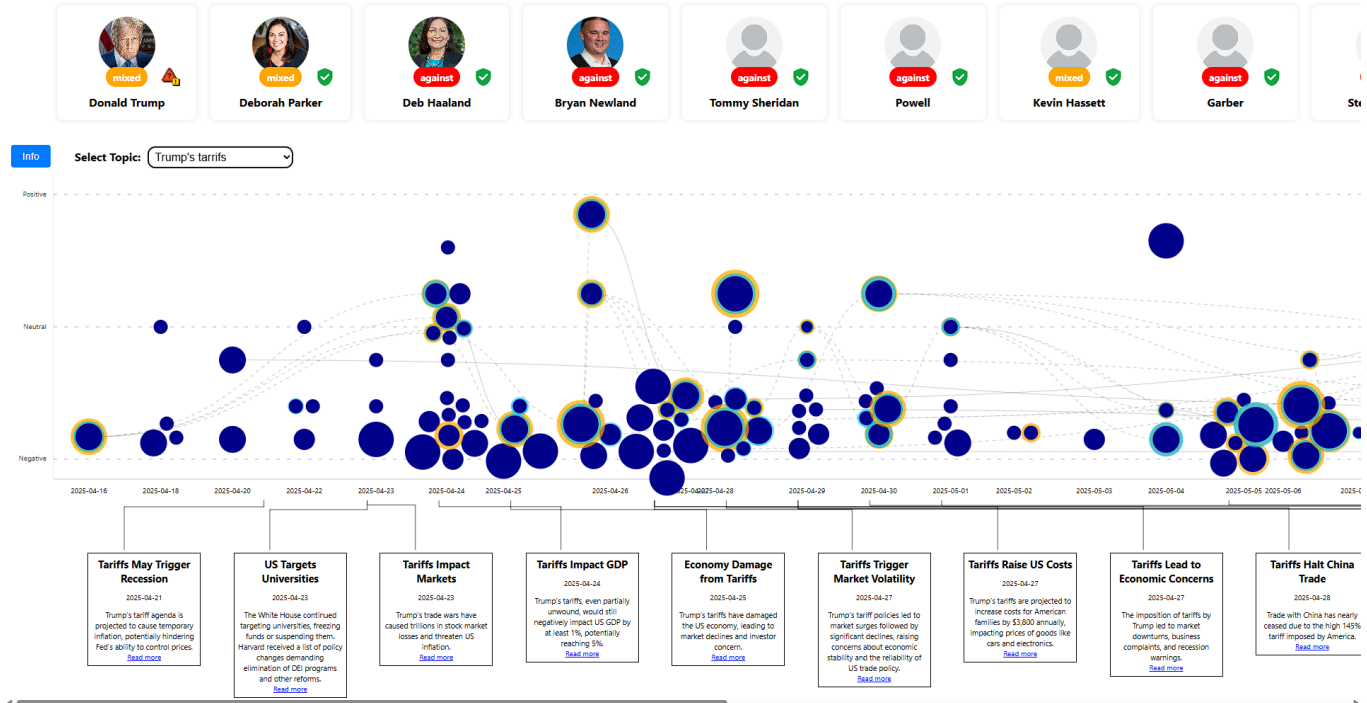


Рисунок Г.5 – Результати роботи системи аналізу настроїв стосовно теми торговельної війни США (частина)

Додаток Д

Акт впровадження



Товариство з обмеженою відповідальністю «КВАЛІА
КОНСАЛТІНГ ЕДЖЕНСІ»

вул. Полідарівська, с.Полідарівка, Київська обл., 07239, Україна | Т: +38 097 296
6579, e-mail: team@thequalia.agency | web: thequalia.agency

ЄДРПОУ 44876766

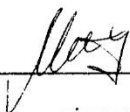
ДОВІДКА

Дана Демкову Івану Владиславовичу в тому, що результати, одержані ним в процесі виконання бакалаврської кваліфікаційної роботи, а саме методика сентимент-аналізу інформаційних потоків, що дозволяє оцінювати зміни в риториці публічних діячів та ідентифікувати потенційні суперечності в їхніх заявах на основі алгоритмів штучного інтелекту, використано в розробках ТОВ «КВАЛІА КОНСАЛТІНГ ЕДЖЕНСІ».



Керівник

ТОВ «ККЕ»


підпис

Тетяна ГУМЕНЮК