МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

СОБЧУК Юлія Євгеніївна

Гібридна модель фактчекінгу на основі NLP/Hybrid Model for Fact-Checking Based on NLP

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма – Штучний інтелект

Кваліфікаційна робота

Виконав студентка групи КНШІ-41
Ю.Є. Собчук
_____

Науковий керівник:
д.т.н., доцент Х.В. Ліп'яніна-Гончаренко
_____

Кваліфікаційну роботу
допущено до захисту:
«___» _____ 20___ р.
Завідувач кафедри
_____ Н.В. Дзюбановська

ТЕРНОПІЛЬ – 2026

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «бакалавр»
спеціальність: 122 – Комп'ютерні науки
освітньо-професійна програма – Штучний інтелект

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
_____ Н.В. Дзюбановська
«____» _____20__р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
СОБЧУК Юлії Євгеніївній
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи
Гібридна модель фактчекінгу на основі NLP/Hybrid Model for Fact-Checking Based on NLP
керівник роботи д.т.н., доцент, Х.В. Ліп'яніна-Гончаренко
затверджений наказом по університету від 01 грудня 2025 року № 894.
2. Строк подання студентом закінченої кваліфікаційної роботи 25 травня 2026 р.
3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.
4. Основні питання, які потрібно розробити:
- провести аналіз предметної області та сучасних методів автоматизованого виявлення фейкових новин і маніпулятивного контенту;
- розробити архітектуру гібридної моделі автоматичного фактчекінгу на основі модулів Bi-LSTM, Agentic RAG та механізму динамічної маршрутизації;
- сформувати корпус україномовних новинних текстів, реалізувати модель попередньої обробки та векторизації текстових даних;
- реалізувати програмні модулі первинної класифікації, фактологічної перевірки та локальної векторної бази знань;
- розробити клієнтський інтерфейс у вигляді Telegram-бота для взаємодії користувача із системою;
- провести експериментальне дослідження та оцінити точність, стійкість і швидкодію розробленої гібридної системи.

5. Перелік графічного матеріалу у роботі:
- архітектура гібридної моделі автоматичного фактчекінгу;
- схема процесу обробки текстових даних у гібридній системі фактчекінгу;
- схема обробки тексту модулем Bi-LSTM;
- схема роботи модуля Agentic RAG;
- схема маршрутизатора;
- узагальнена схема обґрунтування вибору моделей, програмних засобів та критеріїв оцінювання гібридної системи фактчекінгу.

6. Консультанти розділів кваліфікаційної роботи

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

7. Дата видачі завдання 01 грудня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Огляд літературних джерел відповідно до предметної області та складання плану роботи	до 01.01.2026 р.	
2	Написання 1 розділу кваліфікаційної роботи	до 01.02.2026 р.	
3	Написання 2 розділу кваліфікаційної роботи	до 15.03.2026 р.	
4	Написання 3 розділу кваліфікаційної роботи	до 01.05.2026 р.	
5	Представлення попереднього варіанту кваліфікаційної роботи, перевірка та внесення змін керівником	до 15.05.2026 р.	
6	Опрацювання зауважень та представлення завершеного варіанту кваліфікаційної роботи. Підготовка супроводжуючих документів	до 25.05.2026 р.	
7	Перевірка кваліфікаційної роботи на оригінальність тексту	до 30.05.2026 р.	
8	Оформлення кваліфікаційної роботи та отримання допуску до захисту	до 10.06.2026 р.	
9	Подання кваліфікаційної роботи до захисту на засіданні атестаційної комісії	до 10.06.2026 р.	

Студент _____ Ю.Є. Собчук
підпис

Керівник роботи _____ д.т.н., доцент Х.В. Ліп’яніна-Гончаренко
підпис

АНОТАЦІЯ

Пояснювальна записка до кваліфікаційної роботи: 80 с., 20 рис., 1 табл., 39 джерел, 1 додаток.

Об’єктом дослідження є процеси автоматизованого фактчекінгу та виявлення недостовірної інформації у текстовому контенті з використанням методів обробки природної мови.

Метою кваліфікаційної роботи є розробка гібридної моделі фактчекінгу новинного контенту на основі технологій NLP для автоматизованого виявлення фейкових новин і перевірки достовірності інформації.

Методи дослідження: методи обробки природної мови та векторного представлення тексту (FastText), методи глибокого навчання (Bi-LSTM) для первинної класифікації, і механізми динамічної маршрутизації на базі великих мовних моделей.

Результатом роботи є програмне забезпечення для фактчекінгу, що включає Telegram-бот, модуль первинної класифікації (FastText + Bi-LSTM) та модуль фактологічної перевірки на основі Agentic RAG із використанням ChromaDB, Gemini та Google Search API. Реалізовано механізм динамічної маршрутизації запитів для оптимізації використання обчислювальних ресурсів.

Розроблена система може використовуватися в інформаційно-аналітичних центрах, системах журналістського фактчекінгу, інформаційної безпеки та як інструмент перевірки достовірності новин для кінцевих користувачів.

Ключові слова: ВИЯВЛЕННЯ ДЕЗІНФОРМАЦІЇ, ФЕЙКОВІ НОВИНИ, МАШИННЕ НАВЧАННЯ, РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ, AGENTIC RAG, ВЕЛИКІ МОВНІ МОДЕЛІ, ФАКТЧЕКІНГ.

ANNOTATION

Explanatory note to the qualification thesis: 80 pages, 20 figures, 1 table, 36 references, 1 appendice.

The object of research is the processes of automated fact-checking and detection of unreliable information in textual content using natural language processing methods.

The aim of the thesis is to develop a hybrid fact-checking system for news content based on Bi-LSTM and Agentic RAG technologies for the automated detection of fake news and verification of information credibility.

Research methods: natural language processing and text vector representation methods (FastText), deep learning methods (Bi-LSTM) for primary classification, and dynamic routing mechanisms based on large language models.

The result of the work is a software fact-checking system that includes a Telegram bot, a primary classification module (FastText + Bi-LSTM), and a factual verification module based on Agentic RAG using ChromaDB, Gemini, and Google Search API. A dynamic query routing mechanism was implemented to optimize computational resource usage.

The developed system can be used in information and analytical centers, journalistic fact-checking platforms, information security systems, and as a tool for verifying the credibility of news content for end users.

Keywords: DISINFORMATION DETECTION, FAKE NEWS, MACHINE LEARNING, RECURRENT NEURAL NETWORKS, AGENTIC RAG, LARGE LANGUAGE MODELS, FACT-CHECKING.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

Agentic RAG — гібридна архітектура, що інтегрує технологію генерації з доповненим пошуком та концепцію великих мовних моделей як автономних агентів.

API (Application Programming Interface) — інтерфейс програмування застосунків.

Bi-LSTM — двонаправлена довга короткострокова пам'ять.

CNN — згорткові нейронні мережі.

IMI — Інститут масової інформації.

ІПСО — Інформаційно-психологічні спеціальні операції.

LLM (Large Language Models) — Великі мовні моделі.

LSTM — архітектура довгої короткострокової пам'яті.

NLP (Natural Language Processing) — Обробка природної мови (або автоматизована обробка природної мови).

OOV (Out-Of-Vocabulary) — слів, відсутніх у словнику.

RAG (Retrieval-Augmented Generation) — технологія генерації з доповненим пошуком.

RNN — Рекурентні нейронні мережі.

ВСТУП

Інформаційний простір постійно піддається інформаційно-психологічним атакам та масовому поширенню дезінформації. Зростання обсягів текстових даних у соціальних мережах та медіа робить ручний фактчекінг неефективним та ресурсозатратним. Існуючі системи автоматизованого аналізу на базі класичного машинного навчання здатні виявляти маніпулятивний стиль чи токсичність текстів, проте вони не мають механізмів перевірки фактичної достовірності тверджень. Використання ж виключно великих мовних моделей LLM супроводжується проблемою генерації хибних фактів та відсутністю прив'язки до реального часу. Тому розроблення гібридної системи, яка поєднує нейромережевий аналіз тексту з технологією генерації з доповненим пошуком RAG для верифікації фактів на основі надійних україномовних джерел, є надзвичайно актуальним завданням.

Метою кваліфікаційної роботи є розробка гібридної моделі фактчекінгу новинного контенту на основі технологій Bi-LSTM та Agentic RAG для автоматизованого виявлення фейкових новин і перевірки достовірності інформації.

Для досягнення поставленої мети сформульовано та вирішено такі завдання:

– провести аналіз предметної області та сучасних методів автоматизованого виявлення фейкових новин і маніпулятивного контенту;

– розробити архітектуру гібридної моделі автоматичного фактчекінгу на основі модулів Bi-LSTM, Agentic RAG та механізму динамічної маршрутизації;

– сформувати корпус україномовних новинних текстів, реалізувати модель попередньої обробки та векторизації текстових даних;

– реалізувати програмні модулі первинної класифікації, фактологічної перевірки та локальної векторної бази знань;

– розробити клієнтський інтерфейс у вигляді Telegram-бота для взаємодії користувача із системою;

– провести експериментальне дослідження та оцінити точність, стійкість і швидкодію розробленої гібридної системи.

Об’єктом дослідження є процеси автоматизованого фактчекінгу та виявлення недостовірної інформації у текстовому контенті з використанням методів обробки природної мови.

Предметом дослідження є методи класифікації текстових даних на основі Bi-LSTM, технології Agentic RAG та механізми динамічної маршрутизації запитів у системах автоматизованого фактчекінгу.

Методи дослідження. Для вирішення поставлених завдань у роботі використано: методи обробки природної мови та векторного представлення тексту (FastText), методи глибокого навчання (Bi-LSTM) для первинної класифікації, і механізми динамічної маршрутизації на базі великих мовних моделей.

Практичне значення отриманих результатів полягає у створенні функціонуючого програмного модуля та користувацького інтерфейсу у вигляді бота, що дозволяє у реальному часі аналізувати україномовні новини, класифікувати маніпулятивний стиль та верифікувати факти. Результати роботи можуть бути інтегровані у платформи новинних агрегаторів або використовуватися як незалежний інструмент для фактчекінгу.

Результати кваліфікаційної роботи апробовані та опубліковані у матеріалах студентської науково-практичної конференції “XI Міжнародна наукова конференція «Науковий простір: актуальні питання, досягнення та інновації»”, м. Дніпро, Україна, 8 травня 2026 року.

Кваліфікаційна робота складається із вступу, трьох розділів, висновків, списку використаних джерел та додатків.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Опис предметної області

Сучасний етап розвитку інформаційного суспільства характеризується експоненційним зростанням обсягів цифрових даних, безперервним генеруванням контенту користувачами соціальних мереж та миттєвим поширенням новинних повідомлень. В умовах глобальної цифровізації комунікаційних процесів однією з найбільш критичних загроз для інформаційної безпеки постає масове поширення дезінформації, маніпулятивних матеріалів та фейкових новин. Особливої гостроти ця проблема набуває в контексті гібридних інформаційних війн, де цілеспрямовані та добре фінансовані кампанії з дезінформації використовуються як інструмент дестабілізації суспільства, маніпулювання громадською думкою та підриву довіри до офіційних інституцій [6, 23].

Предметна область даного дослідження охоплює процеси виявлення, аналізу та спростування неправдивої інформації в україномовному сегменті мережі Інтернет. Специфіка предметної області визначається тим, що швидкість поширення сфабрикованих новин у соціальних медіа часто значно перевищує швидкість розповсюдження об'єктивних фактів. Алгоритми соціальних платформ, орієнтовані на максимізацію залученості користувачів, схильні просувати емоційно забарвлений та сенсаційний контент, яким найчастіше і є фейкові новини. Це призводить до формування так званих інформаційних «ехо-камер», де користувачі споживають лише ту інформацію, яка підтверджує їхні існуючі упередження, повністю ігноруючи альтернативні або об'єктивні факти [35].

Крім того, багаторазове повторення згенерованих фейків або глибоких підробок (deepfakes) у новинних стрічках спричиняє психологічний ефект «ілюзії правди». Згідно з останніми дослідженнями, часте споживання новин через соціальні мережі посилює цей ефект, змушуючи користувачів сприймати багаторазово побачену маніпуляцію як достовірний факт, навіть якщо первинно вони ставилися до неї скептично [5].

Для українського інформаційного простору ситуація ускладнюється безпрецедентним рівнем інформаційно-психологічних спеціальних операцій ІПСО,

пов'язаних із повномасштабною збройною агресією. Основним каналом швидкого розповсюдження ворожої пропаганди та неперевірених чуток стали платформи миттєвого обміну повідомленнями, зокрема месенджер Telegram [7]. Наявність великої кількості анонімних каналів, відсутність редакційної політики та інструментів модерації контенту створюють ідеальне середовище для посіву дезінформації [20].

Окремим викликом у даній предметній області є специфічні характеристики даних, що підлягають обробці. Текстові дані, які циркулюють у соціальних мережах та месенджерах, є неструктурованими та мають низку лінгвістичних особливостей. Україномовний контент характеризується багатою морфологією, вільним порядком слів у реченні, частим використанням сленгу, неологізмів, емодзі, а також наявністю суржику та змішуванням кириличних і латинських символів [8]. Короткий формат повідомлень, до прикладу, пости у Telegram або заголовки новин, часто не містить достатнього контексту для класичного семантичного аналізу, що суттєво ускладнює завдання автоматизованої обробки природної мови NLP.

З розвитком технологій генеративного штучного інтелекту процес створення фейкових новин став економічно доступним та масовим. Великі мовні моделі здатні генерувати граматично правильні, логічно побудовані та стилістично переконливі тексти, які візуально не відрізняються від професійної журналістики. У зв'язку з цим традиційний ручний фактчекінг, який виконують спеціалізовані організації та журналісти-розслідувачі, стає неефективним. Процес ручної перевірки одного твердження може займати від кількох годин до кількох днів, тоді як за цей час фейкова новина встигає охопити мільйонну аудиторію.

Отже, виникає критична потреба у переході від ручних до автоматизованих методів перевірки інформації. Завдання полягає у створенні інтелектуальних систем, які здатні не лише виявляти підозрілі лінгвістичні та стилістичні патерни у текстах (токсичність, емоційність, клікбейтність), але й здійснювати глибинний фактологічний аналіз, зіставляючи зміст повідомлення з надійними базами знань та офіційними джерелами у режимі реального часу [22]. Це вимагає залучення

передових методів штучного інтелекту, що зумовлює напрямок подальшого дослідження існуючих алгоритмічних та архітектурних рішень у цій сфері.

1.2 Огляд існуючих підходів та методів

Дослідження у сфері автоматизованого виявлення маніпуляцій та фейкових новин інтенсивно розвиваються, і на сьогодні їх можна умовно розділити на три основні напрямки: застосування класичних нейромережевих підходів, використання чистих великих мовних моделей та впровадження систем генерації з доповненим пошуком [24].

Перший напрямок базується на використанні методів глибокого навчанн, які здатні виявляти приховані закономірності у текстах без необхідності ручного конструювання ознак. Як зазначається у огляді [14], традиційні підходи глибокого навчання, зокрема рекурентні нейронні мережі та їхня модифікація LSTM, довели свою високу ефективність у виявленні стилістичних, лінгвістичних та емоційних патернів, що є характерними для неправдивих повідомлень. Архітектура LSTM вирішує проблему зникаючого градієнта, яка притаманна звичайним RNN, що дозволяє моделі запам'ятовувати довгострокові залежності у тексті. Дослідження [10] демонструє, що моделі на основі Bi-LSTM з механізмами уваги здатні досягати точності до 97,66 % у класифікації фейкових новин. Це досягається завдяки тому, що модель аналізує контекст як у прямому, так і у зворотному напрямках, зосереджуючи увагу на найбільш вагомих словах у реченні.

Подальший розвиток цього напрямку супроводжується створенням складних гібридних архітектур. Зокрема, у роботі [3] запропоновано структуру LSTM-CGPNN із застосуванням методів метаевристичної оптимізації, що дозволило підвищити точність розпізнавання до 98 % на стандартних еталонних наборах даних. Водночас у дослідженні [21] було успішно поєднано здатність до семантичної абстракції, яка притаманна архітектурі Transformer, із модулем Self-Attention LSTM, що забезпечило ще глибший аналіз контекстуальних зв'язків між елементами тексту.

Важливим кроком для розв'язання проблеми дезінформації саме у вітчизняному інформаційному просторі є дослідження процесів створення збалансованих україномовних корпусів новин. У роботі [29] доведено, що використання базових моделей Bi-LSTM у поєднанні з алгоритмами просторової векторизації слів FastText демонструє високу здатність до виявлення згенерованих маніпулятивних текстів українською мовою. Модель FastText дозволяє враховувати морфологічну структуру слів шляхом розбиття їх на символьні n-грами, що є критично важливим для мов із багатою словозміною, таких як українська.

Незважаючи на високу точність у виявленні стилістичних аномалій, спільним недоліком суто нейромережевих класифікаторів є їхня обмеженість навчальною вибіркою. Вони спроможні виявити маніпулятивний стиль тексту або токсичність мови, проте не здатні перевіряти нові факти у реальному часі. Якщо у світі відбувається нова подія, якої не було в наборі даних під час навчання нейромережі, такий класифікатор не зможе підтвердити або спростувати цю інформацію на основі фактичних доказів.

Зі стрімким розвитком генеративного штучного інтелекту фокус наукових досліджень змістився на використання великих мовних моделей для завдань автоматизованого фактчекінгу. Завдяки навчанню на гігантських масивах текстових даних, LLM володіють глибоким розумінням семантики та здатністю до логічного висновування. Однак, як детально проаналізовано у дослідженні [1], застосування чистих мовних моделей супроводжується значними ризиками. Головною проблемою є явище так званих "галюцинацій" – коли модель генерує граматично правильні, логічно побудовані та візуально переконливі, але фактично хибні твердження [15]. Крім того, мовні моделі мають обмеження щодо актуальності знань: вони не володіють інформацією про події, що відбулися після завершення процесу їхнього тренування. Відповідно, використання генеративних моделей без доступу до актуальних зовнішніх джерел даних є нерелевантним у контексті перевірки новин [25].

Для вирішення проблеми галюцинацій та забезпечення актуальності даних наукова спільнота перейшла до третього напрямку – використання архітектури

генерації з доповненим пошуком. На сьогодні RAG є передовим стандартом для створення систем автоматизованого фактчекінгу [12, 11]. Ця технологія дозволяє мовним моделям перед формуванням фінальної відповіді здійснювати пошук релевантної інформації у зовнішніх векторних базах даних або безпосередньо у мережі Інтернет. Знайдені фрагменти текстів передаються моделі разом із запитом користувача, що змушує алгоритм базувати свою відповідь виключно на наданих доказах.

Дослідження [20], а також [33] переконливо доводять ефективність підходу RAG на прикладі перевірки медичної дезінформації щодо пандемії COVID-19. Інтеграція RAG-агентів дозволила суттєво знизити рівень галюцинацій мовної моделі та підвищити точність перевірки до 97,3 %. Ефективність RAG для політичних новин та ширшого соціального контексту підтверджується у низці сучасних робіт. Зокрема, у дослідженні [19] використано багатоетапний пошук доказів у мережі Інтернет, що дозволило моделі динамічно уточнювати запити для отримання більш релевантної інформації. У роботі [17] розроблено систему RAGAR, яка застосовує мультимодальний підхід для розпізнавання політичних фейків, враховуючи не лише текстову складову, але й пов'язаний медіаконтент.

У роботі [27] успішно застосовано модель Mixtral із механізмом пошуку у реальному часі. Дослідження довело, що генеративні моделі з інтегрованим RAG можуть на рівних конкурувати за точністю класифікації з класичними моделями архітектури BERT, маючи при цьому значну перевагу – здатність генерувати людинозрозумілі аргументовані пояснення свого рішення. Крім того, концепція динамічного пошуку успішно застосовується у суміжних сферах, таких як виявлення шахрайства у фінансовому секторі за допомогою RAG-орієнтованих LLM [13].

Загалом, сучасна наукова спільнота сходиться на думці, що автоматизований фактчекінг доцільно переформулювати саме як задачу RAG [26, 12], де ключова роль відводиться якості, надійності та релевантності знайдених зовнішніх доказів [16].

Проте, незважаючи на беззаперечні переваги архітектури RAG, її використання для масової обробки новинного потоку має суттєві практичні

обмеження. Аналіз літератури виявив ключову проблему: існуючі рішення або швидкі та ресурсоефективні (нейромережі), але не здатні до фактологічної перевірки нових подій, або дуже точні й аргументовані (LLM + RAG), але надзвичайно повільні та ресурсомісткі. Процес векторизації запиту, векторного пошуку у базі даних та подальшої генерації відповіді масивною мовною моделлю для кожної окремої новини є економічно та обчислювально невиправданим для великих потоків даних. Це актуалізує потребу в розробленні нових, оптимізованих архітектурних підходів.

1.3 Постановка задачі дослідження

На основі проведеного аналізу предметної області та огляду існуючих технологічних рішень встановлено, що сучасні системи автоматизованого фактчекінгу характеризуються суттєвим компромісом між швидкістю обробки даних і глибиною фактологічного аналізу. З одного боку, нейромережеві класифікатори забезпечують високу швидкодію та ресурсоефективність, проте не здатні до повноцінної верифікації нових фактів. З іншого боку, системи на базі великих мовних моделей та архітектури RAG демонструють високу точність і аргументованість висновків, але потребують значних обчислювальних ресурсів і часу, що обмежує їх застосування для потокового моніторингу великих масивів новин. У зв’язку з цим використання модулів RAG для аналізу кожного інформаційного повідомлення є економічно недоцільним.

Для вирішення цієї задачі у даній роботі пропонується розроблення гібридної системи виявлення фейкових новин, яка концептуально поєднує найкращі характеристики обох підходів. Запропонована архітектура передбачає дворівневу перевірку: класична навчена модель глибокого навчання LSTM виступає у ролі швидкого первинного фільтра для відсіювання текстів нейтрального характеру та виявлення очевидних маніпуляцій за стилістичними ознаками. Натомість ресурсомісткий модуль інтелектуального пошуку та генерації Agentic RAG активується динамічно і лише для тих текстів, які потребують детального фактологічного спростування чи пояснення.

Загальна концепція системи базується на принципі інтелектуальної маршрутизації потоку даних, де нейромережевий аналіз забезпечує швидкість обробки, а агентна RAG-архітектура гарантує достовірність фінального вердикту. До ключових функціональних та структурних вимог розробленої системи належать:

– реалізація конвеєра попередньої обробки та векторизації україномовних текстів із використанням субслівних токенів;

– створення та наповнення локального векторного сховища надійних джерел інформації;

– забезпечення механізму динамічного підключення до зовнішніх пошукових систем у разі дефіциту локального контексту;

– формування аргументованих пояснень результатів перевірки природною мовою.

Крім того, для усунення проблеми семантичного перекриття та оптимізації пошуку, пропонується використання локальної векторної бази надійних українських джерел із можливістю динамічного делегування пошуку в глобальну мережу Інтернет лише у випадках, коли локальних даних виявляється недостатньо. Такий підхід покликаний забезпечити високу точність виявлення дезінформації, стійкість до генеративних галюцинацій та оптимальну швидкість відгуку системи, що є критично важливим для моніторингу вітчизняного новинного простору.

З огляду на вищезазначене, метою кваліфікаційної роботи є розроблення та дослідження гібридної моделі автоматичного фактчекінгу україномовних інформаційних повідомлень на основі методів обробки природної мови.

Для досягнення поставленої мети сформульовано та підлягають вирішенню наступні завдання дослідження:

– провести аналіз предметної області та сучасних методів автоматизованого виявлення фейкових новин і маніпулятивного контенту;

– розробити архітектуру гібридної моделі автоматичного фактчекінгу на основі модулів Bi-LSTM, Agentic RAG та механізму динамічної маршрутизації;

– сформувати корпус україномовних новинних текстів, реалізувати модель попередньої обробки та векторизації текстових даних;

– реалізувати програмні модулі первинної класифікації, фактологічної перевірки та локальної векторної бази знань;

– розробити клієнтський інтерфейс у вигляді Telegram-бота для взаємодії користувача із системою;

– провести експериментальне дослідження та оцінити точність, стійкість і швидкодію розробленої гібридної системи.

2 АЛГОРИТМІЧНЕ ТА ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ ГІБРИДНОЇ МОДЕЛІ ФАКТЧЕКІНГУ

2.1 Архітектура гібридної моделі автоматичного фактчекінгу

Розв’язання задачі автоматизованого виявлення фейкових новин та інформаційних маніпуляцій потребує поєднання методів швидкої попередньої класифікації тексту з механізмами поглибленої фактологічної перевірки. У межах кваліфікаційної роботи запропоновано гібридну архітектуру автоматичного фактчекінгу, яка інтегрує методи обробки природної мови, нейромережеву класифікацію, векторний пошук та генеративну модель із механізмом Retrieval-Augmented Generation.

Концептуальна особливість запропонованої системи полягає у реалізації дворівневої схеми аналізу інформаційного повідомлення. На першому рівні здійснюється швидка нейромережева фільтрація тексту з метою виявлення потенційних ознак маніпулятивності, емоційного тиску, сенсаційності або стилістичних маркерів дезінформації. На другому рівні виконується поглиблена фактологічна перевірка повідомлення за допомогою Agentic RAG, що використовує локальну базу знань і, за потреби, зовнішні інформаційні ресурси. Такий підхід дає змогу поєднати швидкодію класичних нейромережевих моделей із доказовістю та пояснюваністю RAG-орієнтованої перевірки.

Архітектура запропонованої системи побудована за модульним принципом і передбачає послідовну взаємодію таких компонентів: клієнтського інтерфейсу, модуля попередньої обробки тексту, модуля векторизації, нейромережевого класифікатора Bi-LSTM, програмного маршрутизатора, локальної векторної бази знань, модуля Agentic RAG та генератора фінального висновку. Узагальнену архітектуру системи подано на рисунку 2.1.

Рисунок 2.1 – Архітектура гібридної моделі автоматичного фактчекінгу

На початковому етапі система отримує текстове повідомлення від користувача через клієнтський інтерфейс. Далі виконується попередня обробка тексту, що передбачає нормалізацію символів, очищення від службових елементів, усунення надлишкових пробілів, токенізацію та перетворення тексту у формат, придатний для подальшої обробки нейромережевими моделями.

Після попередньої обробки текст перетворюється у векторні представлення за допомогою моделі FastText. Використання цієї моделі є доцільним для

україномовних текстів, оскільки вона враховує не лише цілі слова, а й символьні n-грами, що підвищує стійкість системи до словозмін, неологізмів, помилок написання та лексичних варіацій. Отримані послідовності векторних представлень передаються на вхід двонаправленої рекурентної нейронної мережі Bi-LSTM, яка виконує первинну класифікацію повідомлення.

Модуль Bi-LSTM призначений для виявлення стилістичних, семантичних та контекстних ознак, характерних для потенційно маніпулятивного контенту. Двонаправлена структура мережі дає змогу аналізувати текст як у прямому, так і у зворотному напрямках, що забезпечує повніше врахування контекстних залежностей між словами та фрагментами повідомлення.

Центральним компонентом запропонованої архітектури є програмний маршрутизатор. Його функція полягає у прийнятті рішення щодо подальшого сценарію обробки тексту. Якщо ймовірність маніпулятивності є низькою, система може сформувати попередній висновок без залучення ресурсоємного модуля фактологічної перевірки. Якщо ж текст має ознаки потенційної дезінформації або результат класифікації є недостатньо впевненим, маршрутизатор передає повідомлення до модуля Agentic RAG. Такий механізм дає змогу оптимізувати використання обчислювальних ресурсів і зменшити кількість звернень до великої мовної моделі.

Для реалізації механізму Retrieval-Augmented Generation використовується локальна векторна база знань ChromaDB. У ній зберігаються векторні представлення перевірених новинних матеріалів, фактологічних джерел та текстових фрагментів, які можуть бути використані як доказовий контекст під час перевірки повідомлення. Пошук релевантної інформації здійснюється на основі семантичної подібності між вектором запиту та векторами документів, збережених у базі знань.

Модуль Agentic RAG виконує поглиблену фактологічну перевірку повідомлення. На відміну від класичних RAG-систем, у яких пошук і генерація відповіді відбуваються за наперед визначеним сценарієм, у запропонованій архітектурі велика мовна модель функціонує як автономний агент [34]. Вона здатна оцінювати достатність локального контексту, визначати потребу у зверненні до

зовнішніх джерел і формувати додаткові пошукові запити. Якщо локальна база знань не містить достатньої кількості релевантної або актуальної інформації, агент ініціює зовнішній пошук через Google Search API.

На завершальному етапі велика мовна модель формує фінальний висновок щодо достовірності повідомлення. Результат роботи системи не обмежується бінарною класифікацією на правдиве або неправдиве повідомлення. Система також генерує текстове пояснення, наводить аргументацію, зазначає знайдені підтверджувальні або спростовувальні факти та описує виявлені ознаки маніпулятивності. Це підвищує інтерпретованість результату та дає користувачеві змогу оцінити логіку прийнятого рішення.

2.2 Інформаційне забезпечення та модель обробки текстових даних

Ефективність гібридної системи автоматизованого фактчекінгу значною мірою визначається якістю інформаційного забезпечення, структурою вхідних даних та послідовністю їх обробки. На відміну від класичних систем текстової класифікації, які переважно виконують лише стилістичний або семантичний аналіз повідомлень, запропонована система реалізує багаторівневий конвеєр обробки даних. Він поєднує попередню підготовку тексту, векторизацію, первинну нейромережеву класифікацію, семантичний пошук у локальній базі знань та поглиблену фактологічну перевірку з використанням Agentic RAG.

Інформаційне забезпечення системи охоплює три основні групи даних: вхідні дані, проміжні дані та вихідні результати аналізу. Вхідні дані формують початковий інформаційний об’єкт для перевірки; проміжні дані виникають у процесі трансформації тексту та використовуються окремими модулями системи; вихідні результати містять фінальний фактологічний висновок, пояснення та джерела підтвердження або спростування. Узагальнену схему інформаційного забезпечення та обробки текстових даних у запропонованій системі подано на рисунку 2.2.

Як видно з рисунка 2.2, інформаційний потік у системі має послідовно-адаптивний характер. На початковому етапі система отримує текстове

повідомлення, яке може бути подане у вигляді новинного тексту, заголовка, посилання на публікацію або повідомлення із соціальних мереж. До структури вхідних даних можуть входити основний текст повідомлення, заголовок, назва джерела, дата публікації, URL-адреса та додаткові метадані. Наявність таких параметрів дає змогу враховувати не лише зміст повідомлення, а й його походження, часовий контекст та потенційний зв’язок із конкретним інформаційним джерелом.

Заголовок повідомлення має важливе значення для первинного аналізу, оскільки саме в ньому часто концентруються маніпулятивні елементи: емоційно забарвлена лексика, клікбейтні конструкції, перебільшення, узагальнення або провокаційні твердження. Основний текст використовується для контекстного аналізу та фактологічної перевірки. Дата публікації дає змогу враховувати темпоральний аспект, оскільки частина дезінформаційних повідомлень ґрунтується на використанні застарілих фактів або перенесенні подій в інший часовий контекст. Дані про джерело публікації можуть використовуватися як додатковий інформаційний сигнал під час подальшого оцінювання достовірності.

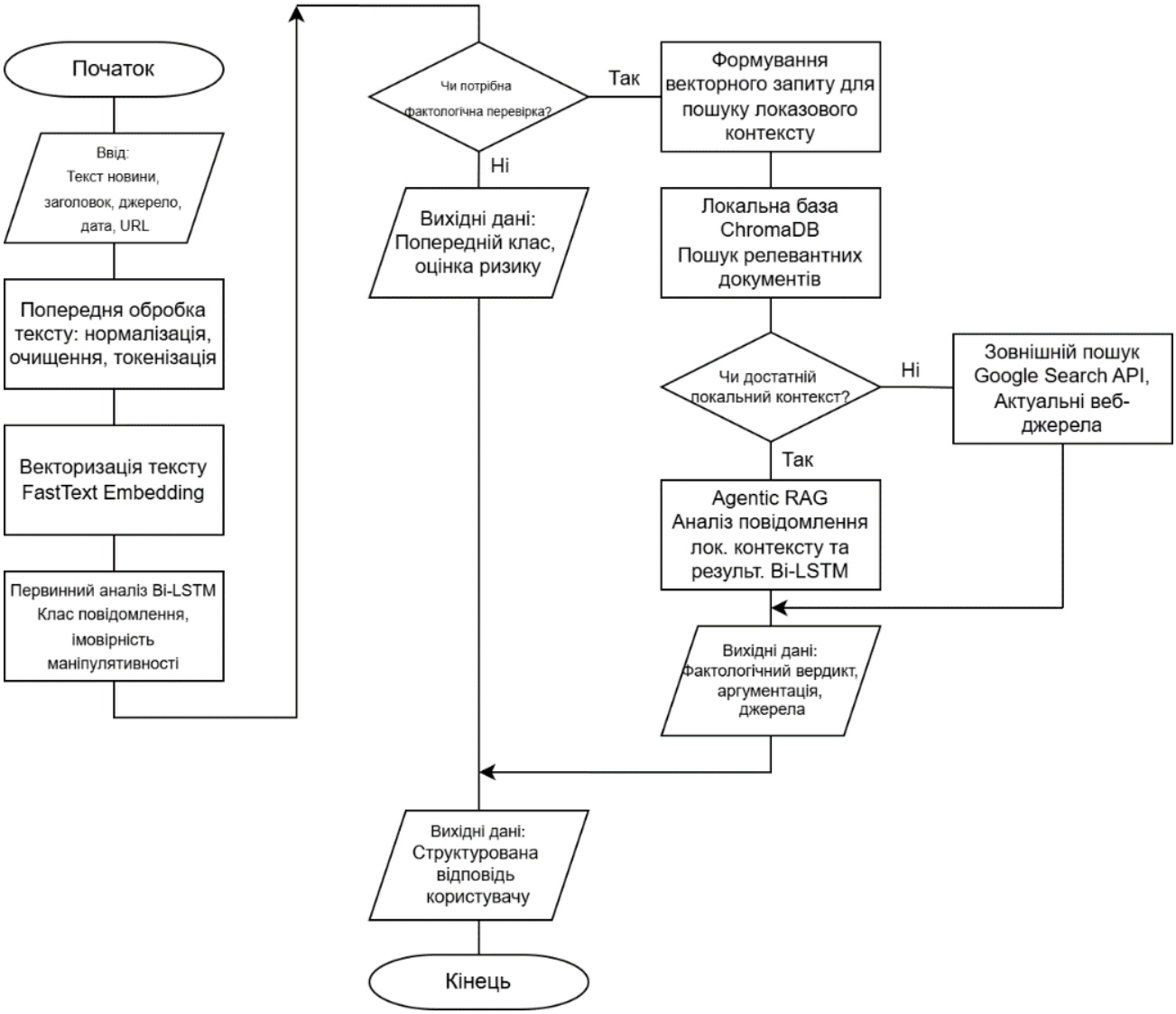

Рисунок 2.2 – Схема процесу обробки текстових даних у гібридній системі фактчекінгу

Перед передачею тексту до нейромережевих компонентів виконується попередня обробка. Її метою є зменшення рівня шуму, уніфікація текстового представлення та підготовка повідомлення до математичної обробки. Новинні тексти та повідомлення із соціальних мереж часто містять HTML-розмітку, гіперпосилання, емодзі, технічні маркери, дубльовані пробіли, службові символи та інші елементи, які не несуть значущого семантичного навантаження. Наявність таких компонентів може знижувати якість класифікації та ускладнювати роботу мовної моделі.

Попередня обробка передбачає нормалізацію регістру, очищення тексту від службових і технічних символів, усунення HTML-тегів, видалення надлишкових

пробілів, нормалізацію пунктуації та токенізацію. У результаті вхідне повідомлення перетворюється на впорядковану послідовність мовних одиниць, придатну для подальшої векторизації.

Після токенізації формується векторне представлення тексту за допомогою моделі FastText. Використання цієї моделі є доцільним для україномовних текстів, оскільки вона враховує не лише цілі слова, а й субслівні компоненти. Це дає змогу ефективніше працювати з відмінюваннями, словотворчими варіаціями, рідковживаними словами, неологізмами та лексичними одиницями, відсутніми у словнику моделі. Кожен токен перетворюється у числовий вектор фіксованої розмірності, а сукупність таких векторів формує послідовність, яка подається на вхід модуля Bi-LSTM.

Модуль Bi-LSTM виконує первинний аналіз повідомлення та формує ймовірнісну оцінку його маніпулятивності. На цьому етапі система визначає попередній клас тексту, рівень імовірності наявності маніпулятивних ознак та ступінь упевненості класифікатора. Отримані результати належать до проміжних даних і передаються до програмного маршрутизатора, який визначає подальший сценарій обробки повідомлення.

Якщо результат первинної класифікації свідчить про низький ризик маніпулятивності, система може сформувати короткий попередній висновок без запуску поглибленої фактологічної перевірки. Якщо ж повідомлення класифікується як потенційно маніпулятивне, сумнівне або таке, що потребує додаткового підтвердження, маршрутизатор ініціює етап семантичного пошуку доказового контексту.

Для цього текст повідомлення перетворюється у векторний запит, який використовується для пошуку релевантної інформації у локальній векторній базі знань ChromaDB [18]. У базі зберігаються перевірені новинні матеріали, фактологічні довідки, текстові фрагменти, результати попередніх перевірок та інші джерела, що можуть бути використані для підтвердження або спростування аналізованого твердження. Пошук здійснюється на основі семантичної подібності між вектором запиту та векторними представленнями документів у базі знань.

Результатом векторного пошуку є множина релевантних документів, які формують локальний доказовий контекст для модуля Agentic RAG. Якщо локальна база знань містить достатню кількість актуальних і релевантних матеріалів, ці дані передаються безпосередньо до модуля поглибленої перевірки. Якщо ж локальний контекст є недостатнім, система ініціює зовнішній пошук через Google Search API. Такий механізм дає змогу доповнити локальні дані актуальними веб-джерелами та зменшити ризик формування висновку на основі неповного або застарілого контексту.

Модуль Agentic RAG виконує інтеграцію результатів первинної класифікації, локального доказового контексту та, за потреби, зовнішніх джерел. На цьому етапі велика мовна модель аналізує зміст повідомлення, зіставляє його з релевантними фактами, оцінює узгодженість знайдених даних і формує фінальний фактологічний висновок. На відміну від класичних RAG-систем, агентний підхід передбачає можливість адаптивного прийняття рішень щодо необхідності додаткового пошуку або уточнення контексту.

Вихідними даними системи є структурований результат аналізу, який містить попередній клас повідомлення, ймовірність маніпулятивності, фінальний фактологічний вердикт, оцінку достовірності, текстове пояснення, аргументацію рішення та перелік джерел підтвердження або спростування. Такий формат результату забезпечує не лише класифікацію повідомлення, а й пояснюваність прийнятого рішення, що є важливим для практичного використання систем автоматичного фактчекінгу.

Таким чином, запропонована модель інформаційного забезпечення реалізує повний цикл обробки текстових даних: від отримання повідомлення до формування аргументованого фактологічного висновку. Поєднання попередньої обробки, FastText-векторизації, Bi-LSTM-класифікації, локального векторного пошуку та Agentic RAG забезпечує адаптивність системи, підвищує достовірність аналізу та дозволяє раціонально використовувати обчислювальні ресурси під час перевірки інформаційного контенту.

2.3 Математико-алгоритмічне забезпечення модулів Bi-LSTM та Agentic RAG

Функціонування модуля первинної класифікації у запропонованій системі можна подати як послідовний процес перетворення вхідного текстового повідомлення у ймовірнісну оцінку його маніпулятивності. Загальну схему обробки тексту модулем Bi-LSTM подано на рисунку 2.3.

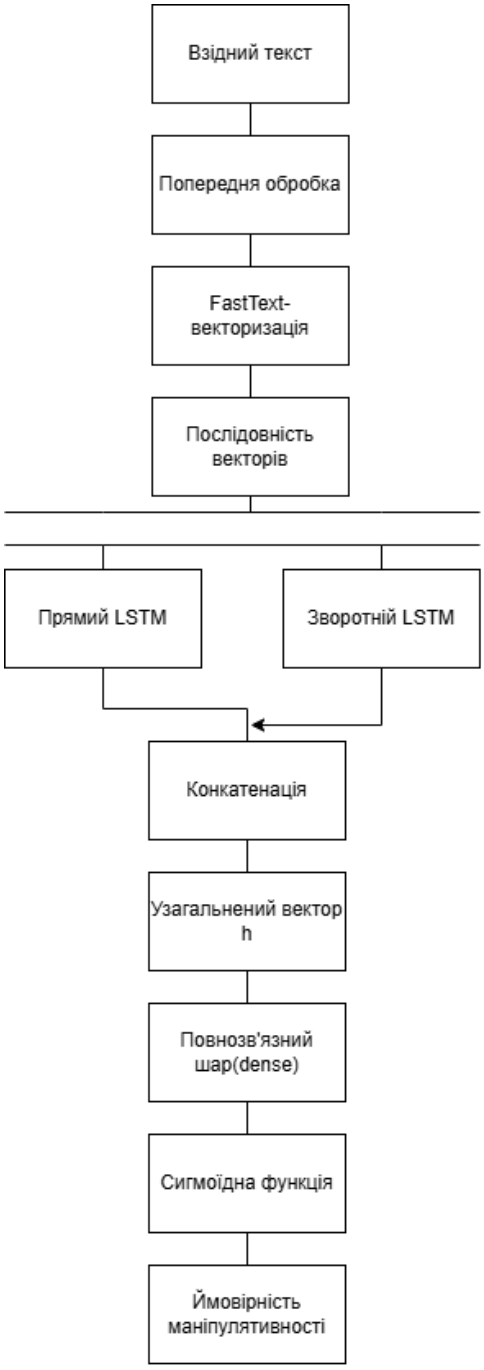

Рисунок 2.3 - Схема обробки тексту модулем Bi-LSTM

На першому етапі на вхід системи надходить текстове повідомлення x, яке може містити природномовні конструкції, скорочення, службові символи, шумові елементи та лексичні неоднорідності. Після попередньої обробки текст перетворюється на впорядковану послідовність токенів:

$$x = \{w_1, w_2, \ldots, w_n\}, \tag{2.1}$$

де w_i— окремий токен,

n— кількість токенів у повідомленні.

Далі виконується векторизація тексту за допомогою моделі FastText[36]. Кожному токену ставиться у відповідність щільне числове представлення у багатовимірному семантичному просторі. У результаті формується послідовність векторів:

$$X = \{v_1, v_2, \ldots, v_n\}, \tag{2.2}$$

де v_i— векторне представлення відповідного токена.

На відміну від класичних моделей, FastText формує представлення слова з урахуванням символьних n-грам, що дає змогу враховувати морфологічну структуру української мови та зменшувати вплив проблеми слів, відсутніх у словнику.

Отримана послідовність векторів подається на вхід двонаправленої рекурентної нейронної мережі Bi-LSTM. Прямий шар LSTM обробляє послідовність від початку до кінця, формуючи приховані стани, що враховують попередній контекст[2, 9]. Зворотний шар LSTM аналізує послідовність у протилежному напрямку, що дає змогу враховувати наступний контекст відносно кожного токена. Результати обох напрямків об’єднуються шляхом конкатенації:

$$h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}], \tag{2.3}$$

де $\overrightarrow{h_i}$— прихований шар прямого проходу;

$\overleftarrow{h_i}$— прихований шар зворотного проходу;

h_i— узагальнене контекстне представлення токена.

Після обробки послідовності формується загальний контекстний вектор повідомлення, який містить інформацію про семантичні та стилістичні залежності в тексті. Цей вектор передається до повнозв'язного шару нейронної мережі, який виконує перетворення ознак у простір класифікаційного рішення. На виході використовується сигмоїдна функція активації:

$$p = \sigma(Wh + b), \tag{2.4}$$

де p— ймовірність належності повідомлення до класу маніпулятивного або потенційно недостовірного контенту;

W— матриця ваг;

h— контекстне представлення тексту;

b— вектор зміщення.

Фінальним результатом роботи модуля Bi-LSTM є числова оцінка маніпулятивності тексту. Це значення передається до програмного маршрутизатора і використовується для прийняття рішення щодо необхідності запуску поглибленої фактологічної перевірки.

Подальший етап обробки інформації реалізується за допомогою алгоритму Agentic Retrieval-Augmented Generation. Його призначення полягає у поглибленій перевірці фактичної достовірності повідомлення на основі доказового контексту. Схему роботи алгоритму подано на рисунку 2.4.

Робота Agentic RAG починається з формування запиту на основі вхідного тексту та результату первинної класифікації, отриманого від Bi-LSTM. Такий запит відображає не лише семантичний зміст повідомлення, а й рівень його потенційної маніпулятивності. Це дає змогу адаптувати подальший пошук до конкретної задачі фактологічної перевірки.

Сформований запит перетворюється у векторне представлення, після чого здійснюється пошук релевантних документів у локальній векторній базі знань. На цьому етапі обчислюється семантична подібність між вектором запиту та векторами документів. Найбільш релевантні документи формують початковий локальний доказовий контекст.

Рисунок 2.4 – Схема роботи програмного модуля Agentic RAG

Якщо рівень релевантності знайдених документів є недостатнім або локальна база знань не містить актуальних відомостей щодо перевірюваного твердження, система переходить до зовнішнього пошуку. У цьому випадку агент формує уточнений пошуковий запит і звертається до зовнішніх інформаційних ресурсів

через Google Search API. Отримані результати використовуються для розширення доказового контексту.

Після об’єднання локальних і зовнішніх джерел формується єдиний доказовий контекст, який передається до великої мовної моделі. Модель виконує узгодження знайдених фактів, аналізує їхню релевантність і формує фінальний висновок щодо достовірності повідомлення. Результат містить не лише вердикт, а й аргументоване пояснення, що підвищує прозорість і практичну цінність системи.

Ключовим елементом запропонованої архітектури є алгоритм маршрутизації, який забезпечує динамічний вибір рівня обробки тексту. На основі оцінки, отриманої від Bi-LSTM, система визначає, чи достатньо первинної класифікації, чи необхідна поглиблена перевірка за допомогою Agentic RAG. Схему роботи маршрутизатора подано на рисунку 2.5.

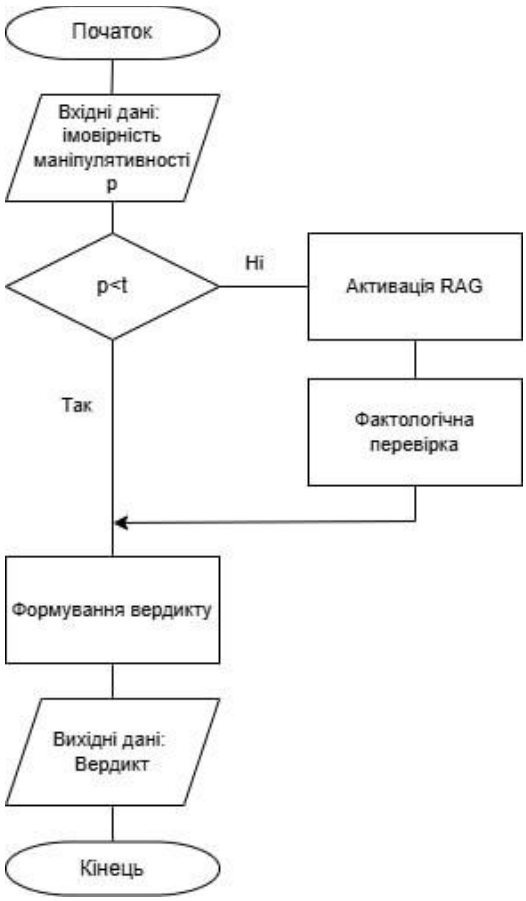

Рисунок 2.5 – Схема маршрутизатора

Якщо ймовірність маніпулятивності є нижчою за встановлений поріг і модель демонструє достатній рівень упевненості, обробка завершується на етапі первинної

класифікації. Це дає змогу уникнути надлишкового використання великої мовної моделі для очевидно нейтральних або інформаційно безпечних повідомлень.

Якщо ж імовірність маніпулятивності перевищує заданий поріг або результат класифікації є невизначеним, маршрутизатор активує модуль Agentic RAG. У такому випадку система переходить до семантичного пошуку, формування доказового контексту та фактологічної перевірки. Отже, рішення про залучення Agentic RAG приймається адаптивно, залежно від складності конкретного інформаційного повідомлення.

Запропонований механізм маршрутизації реалізує дворівневу модель аналізу. Перший рівень забезпечує швидку нейромережеву фільтрацію текстового потоку з мінімальними обчислювальними витратами. Другий рівень активується лише для повідомлень із підвищеним ризиком дезінформації та забезпечує поглиблену перевірку на основі зовнішніх джерел знань. Такий підхід зменшує обчислювальне навантаження, скорочує кількість API-запитів і підвищує швидкість реагування системи в умовах потокової обробки новинного контенту.

Таким чином, маршрутизатор виконує функцію інтелектуального оркестратора всієї системи. Він забезпечує адаптивний перехід між швидкою стилістичною класифікацією та поглибленою фактологічною перевіркою, що формує одну з ключових особливостей запропонованої гібридної архітектури автоматичного фактчекінгу.

2.4 Обґрунтування вибору моделей, програмних засобів та критеріїв оцінювання

Вибір моделей, алгоритмів і програмних засобів для побудови гібридної системи автоматичного фактчекінгу здійснювався з урахуванням специфіки поставленої задачі, особливостей україномовного новинного контенту, вимог до швидкодії та необхідності формування пояснюваного результату. Основними критеріями вибору були здатність системи обробляти морфологічно складний текст, забезпечувати швидку первинну класифікацію, виконувати пошук

релевантного фактологічного контексту, формувати аргументований висновок та раціонально використовувати обчислювальні ресурси.

Для узагальнення логіки вибору моделей, програмних засобів та критеріїв оцінювання доцільно подати схему, яка відображає зв'язок між вимогами до системи, обраними компонентами та показниками ефективності. Такий підхід дає змогу показати, що кожен елемент архітектури виконує окрему функцію в межах загального конвеєра автоматичного фактчекінгу. Узагальнену схему обґрунтування вибору компонентів системи подано на рисунку 2.6.

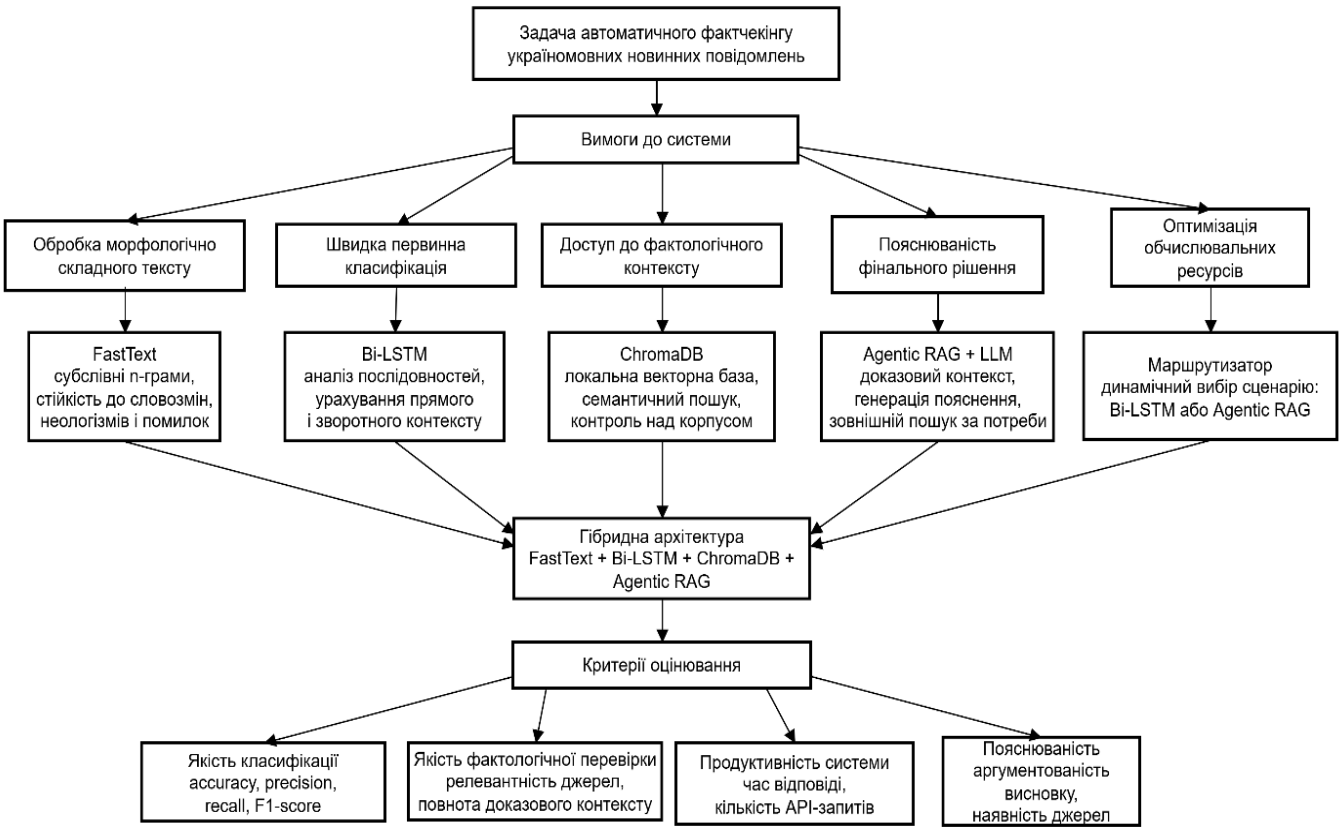

Рисунок 2.6 – Узагальнена схема обґрунтування вибору моделей, програмних засобів та критеріїв оцінювання гібридної системи фактчекінгу

Як видно з рисунка 2.6, вибір компонентів системи безпосередньо пов'язаний із функціональними вимогами до гібридного фактчекінгу. FastText забезпечує стійке векторне представлення україномовного тексту, Bi-LSTM використовується для швидкої первинної класифікації, ChromaDB відповідає за локальний семантичний пошук, а Agentic RAG забезпечує поглиблену перевірку з

формуванням пояснюваного висновку. Маршрутизатор поєднує ці компоненти в єдиний адаптивний конвеєр, що дозволяє зменшити кількість ресурсоємних звернень до великої мовної моделі та водночас зберегти якість фактологічної перевірки.

Для формування векторних представлень тексту обрано модель FastText. Її використання є обґрунтованим з огляду на особливості української мови, яка характеризується розвиненою системою словозміни, наявністю відмінків, префіксів, суфіксів і значної кількості словоформ. На відміну від класичних підходів векторизації, FastText використовує символьні n-грами, що дає змогу формувати векторні представлення не лише для відомих слів, а й для лексичних одиниць, відсутніх у тренувальному словнику. Це є важливим для аналізу новинних повідомлень і текстів із соціальних мереж, у яких часто трапляються неологізми, сленг, скорочення, помилки написання та варіативні форми слів.

Як нейромережевий класифікатор обрано архітектуру Bi-LSTM, що є двонаправленою модифікацією мережі довгої короткострокової пам’яті. Її перевага полягає у здатності враховувати порядок слів і довготривалі залежності між елементами текстової послідовності. На відміну від класичних алгоритмів машинного навчання, зокрема наївного баєсівського класифікатора або методу опорних векторів, рекурентна нейронна мережа аналізує не лише набір окремих ознак, а й послідовну структуру повідомлення. Двонаправлена обробка дає змогу враховувати як попередній, так і наступний контекст відносно кожного елемента тексту, що є важливим для виявлення прихованих маніпулятивних конструкцій.

Для реалізації фактологічної перевірки використано локальну векторну базу даних ChromaDB. Її застосування дає змогу організувати зберігання високорозмірних векторних представлень документів і виконувати швидкий семантичний пошук релевантних фрагментів. Перевагою локального сховища є можливість контролю над корпусом даних, автономність роботи, гнучкість оновлення інформаційної бази та зменшення затримок, пов’язаних із мережевими зверненнями до зовнішніх сервісів. У межах запропонованої системи ChromaDB виконує роль джерела локального доказового контексту для подальшої роботи модуля Agentic RAG.

Для формування фінального висновку використовується велика мовна модель у складі Agentic RAG. Її застосування дає змогу не лише виконувати класифікацію повідомлення, а й формувати пояснення на основі знайденого доказового контексту. На відміну від ізольованого використання великої мовної моделі, RAG-орієнтований підхід зменшує ризик генерації необґрунтованих тверджень, оскільки відповідь формується з урахуванням релевантних документів. Агентна модифікація цього підходу додатково забезпечує можливість динамічного звернення до зовнішніх інформаційних джерел у випадку недостатності локального контексту або потреби в актуалізації фактологічних даних.

Окрему роль у запропонованій архітектурі відіграє програмний маршрутизатор. Його використання обґрунтовується необхідністю адаптивного вибору сценарію обробки повідомлення. Якщо результат первинної класифікації свідчить про низький рівень ризику, система може завершити аналіз без залучення великої мовної моделі. Якщо ж повідомлення має ознаки потенційної маніпуляції або класифікатор демонструє недостатню впевненість, маршрутизатор активує модуль Agentic RAG. Такий підхід дозволяє уникнути надлишкового використання ресурсоємних генеративних моделей і водночас забезпечити поглиблену перевірку складних або сумнівних повідомлень.

Застосування гібридної архітектури є доцільним з огляду на обмеження ізольованих підходів. Використання лише Bi-LSTM дало б змогу швидко виявляти стилістично підозрілі повідомлення, однак така модель не має доступу до актуальних фактологічних даних і не може підтверджувати або спростовувати конкретні твердження. Використання лише LLM із RAG, навпаки, забезпечувало б глибшу перевірку, але потребувало б значних обчислювальних ресурсів і часу для обробки кожного повідомлення. Поєднання цих підходів дає змогу реалізувати оптимізований конвеєр аналізу, у якому швидка нейромережева модель виконує первинну фільтрацію, а модуль Agentic RAG активується лише для складних або сумнівних випадків.

Оцінювання ефективності запропонованої системи доцільно здійснювати за кількома групами критеріїв. Для модуля первинної класифікації використовуються стандартні метрики машинного навчання: accuracy, precision, recall, F1-score та

матриця помилок [31]. Ці показники дають змогу оцінити здатність Bi-LSTM правильно відокремлювати потенційно маніпулятивні повідомлення від нейтрального або достовірного контенту.

Для модуля Agentic RAG важливими критеріями є релевантність знайдених джерел, повнота доказового контексту, відповідність сформованого висновку знайденим фактам і якість текстового пояснення. Оскільки цей модуль не лише класифікує повідомлення, а й формує аргументований вердикт, його ефективність необхідно оцінювати не тільки за формальними метриками, а й за якістю пояснюваності результату.

Для системи загалом доцільно оцінювати час відповіді, частку запитів, що передаються до Agentic RAG, кількість зовнішніх API-звернень, стабільність роботи в умовах різної складності вхідних повідомлень і здатність системи формувати зрозумілу для користувача відповідь. Такі критерії дозволяють оцінити не лише точність, а й практичну придатність запропонованого рішення для роботи з реальним новинним потоком.

Таким чином, вибір FastText, Bi-LSTM, ChromaDB, Agentic RAG та програмного маршрутизатора відповідає поставленій задачі автоматизованого фактчекінгу. Запропонована комбінація моделей і програмних засобів забезпечує баланс між швидкістю первинної обробки, глибиною фактологічної перевірки, пояснюваністю результатів і раціональним використанням обчислювальних ресурсів. Саме поєднання цих компонентів формує основу гібридної системи, здатної адаптивно обробляти україномовний інформаційний контент та формувати аргументовані висновки щодо його достовірності.

2.5 Узагальнення гібридної моделі автоматичного фактчекінгу

Узагальнення запропонованих у попередніх підрозділах архітектурних, інформаційних та алгоритмічних рішень дає змогу подати розроблену систему як цілісну гібридну модель автоматичного фактчекінгу. У межах цієї роботи гібридна модель розглядається як поєднання методів обробки природної мови, нейромережевої класифікації, семантичного пошуку та генеративної фактологічної

перевірки. Її призначення полягає у виявленні потенційно маніпулятивних або недостовірних україномовних інформаційних повідомлень, а також у формуванні аргументованого висновку щодо їхньої достовірності.

Гібридність моделі полягає у дворівневій організації процесу аналізу. Перший рівень забезпечує швидку первинну класифікацію текстового повідомлення за допомогою методів NLP, FastText-векторизації та нейромережевої моделі Bi-LSTM. Другий рівень активується лише для повідомлень, які мають ознаки потенційної маніпуляції або потребують додаткової фактологічної перевірки. На цьому рівні використовується локальна векторна база знань ChromaDB, механізм семантичного пошуку та Agentic RAG, який формує фінальний вердикт на основі доказового контексту.

Вхідним об'єктом гібридної моделі є текстове повідомлення x, яке може містити новинний текст, заголовок, фрагмент публікації або повідомлення із соціальних мереж. Після попередньої обробки текст подається у вигляді послідовності токенів.

На етапі векторизації кожному токену ставиться у відповідність числове векторне представлення. У результаті формується послідовність embedding-векторів.

Використання FastText дає змогу враховувати субслівні ознаки, що є важливим для україномовного контенту з огляду на його морфологічну варіативність.

Отримана послідовність векторів X надходить на вхід Bi-LSTM-класифікатора. Модель аналізує текст у прямому та зворотному напрямках, формує контекстне представлення повідомлення та обчислює ймовірність його належності до класу потенційно маніпулятивного або недостовірного контенту:

$$p = f_{BiLSTM}(X), \tag{2.5}$$

де $p \in [0; 1]$— ймовірнісна оцінка маніпулятивності повідомлення;

f_{BiLSTM}— функція первинної нейромережевої класифікації.

Значення p не є остаточним фактологічним вердиктом, а використовується як індикатор ризику для вибору подальшого сценарію обробки.

Ключовим компонентом гібридної моделі є маршрутизатор, який визначає, чи достатньо результату первинної класифікації, чи необхідно активувати поглиблену фактологічну перевірку. Рішення маршрутизатора залежить від ймовірності маніпулятивності p, порогового значення ризику τта рівня впевненості класифікатора. Для бінарної класифікації рівень впевненості можна подати як віддаленість прогнозу від межі невизначеності:

$$c = | p - 0.5 |, \tag{2.6}$$

де c— умовна оцінка впевненості класифікатора.

Чим ближче значення p до 0.5, тим вищою є невизначеність моделі щодо класу повідомлення.

Правило вибору сценарію обробки можна подати так:

$$R(x) = \begin{cases} C_{BiLSTM}(x), & p < \tau \text{ та } c \geq \gamma, \\ F_{RAG}(x, p, D), & p \geq \tau \text{ або } c < \gamma, \end{cases} \tag{2.7}$$

де $R(x)$— вибраний сценарій обробки повідомлення;

$C_{BiLSTM}(x)$— результат первинної класифікації без запуску поглибленої перевірки;

$F_{RAG}(x, p, D)$— процедура фактологічної перевірки за допомогою Agentic RAG;

τ— порогове значення ризику;

γ— мінімально допустимий рівень упевненості;

D— множина релевантних документів і джерел, використаних для перевірки.

Якщо повідомлення має низьку ймовірність маніпулятивності та класифікатор демонструє достатній рівень упевненості, система формує

попередній висновок без залучення ресурсоємного модуля Agentic RAG. Якщо ж значення pперевищує пороговий рівень або прогноз перебуває в зоні невизначеності, маршрутизатор активує поглиблену перевірку.

На етапі фактологічної перевірки повідомлення перетворюється у векторний запит q, який використовується для пошуку релевантних документів у локальній векторній базі знань. Релевантність документа визначається на основі косинусної подібності між вектором запиту та вектором документа[2]:

$$sim(q, d_j) = (cos\theta) = \frac{q \cdot d_j}{\| q \| \| d_j \|}, \tag{2.8}$$

де q— вектор запиту;

d_j— векторне представлення документа з бази знань;

$sim(q, d_j)$— значення косинусної подібності між запитом і документом.

Множина локально знайдених документів визначається як:

$$D_{local} = \{d_j \in B \mid sim(q, d_j) \geq \alpha\}, \tag{2.9}$$

де B— локальна векторна база знань;

α— мінімальний поріг релевантності документа.

Якщо множина D_{local} є недостатньою за кількістю або релевантністю, Agentic RAG ініціює зовнішній пошук і доповнює доказовий контекст актуальними веб-джерелами:

$$D = D_{local} \cup D_{web}, \tag{2.10}$$

де D_{web}— множина документів, отриманих із зовнішніх інформаційних джерел;

D— загальна множина доказових матеріалів для формування фінального висновку.

Фінальний результат роботи гібридної моделі формується великою мовною моделлю на основі вхідного повідомлення, результату первинної класифікації та доказового контексту[18, 11]:

$$y = F_{LLM}(x, p, D), \tag{2.11}$$

де y— структурований результат фактчекінгу;

F_{LLM}— функція генеративного аналізу в межах Agentic RAG;

x— вхідне повідомлення,

p— ймовірнісна оцінка маніпулятивності;

D— множина релевантних джерел.

Структурований результат фактчекінгу можна подати у вигляді кортежу:

$$y = (v, s, e, S), \tag{2.12}$$

де v— фактологічний вердикт;

s— оцінка достовірності або рівень упевненості у фінальному висновку;

e— текстове пояснення прийнятого рішення;

S— перелік джерел підтвердження або спростування.

Отже, запропонована гібридна модель має послідовно-адаптивний характер. Її послідовність проявляється у визначеному конвеєрі обробки тексту: отримання повідомлення, попередня обробка, векторизація, первинна класифікація, маршрутизація, пошук доказового контексту та формування фінального висновку. Адаптивність полягає у тому, що не всі повідомлення проходять однаково складну процедуру перевірки: для повідомлень із низьким рівнем ризику достатньо швидкої нейромережевої класифікації, тоді як складні або сумнівні повідомлення передаються на рівень поглибленої фактологічної перевірки.

Таким чином, гібридна модель автоматичного фактчекінгу на основі NLP у межах цієї роботи інтегрує FastText-векторизацію, Bi-LSTM-класифікацію,

програмну маршрутизацію, ChromaDB-пошук та Agentic RAG. Така комбінація забезпечує баланс між швидкістю первинної обробки, точністю виявлення маніпулятивного контенту, актуальністю фактологічної перевірки та пояснюваністю фінального результату. Саме це дає змогу розглядати запропонований підхід не як набір окремих програмних компонентів, а як цілісну гібридну модель автоматичного фактчекінгу україномовних інформаційних повідомлень.

3 РЕАЛІЗАЦІЯ ГІБРИДНОЇ МОДЕЛІ ТА ЕКСПЕРИМЕНТАЛЬНА ОЦІНКА ЇЇ ЕФЕКТИВНОСТІ

3.1 Підготовка корпусу даних і формування локальної бази знань

Практична реалізація гібридної моделі автоматичного фактчекінгу передбачала послідовне виконання кількох етапів: формування корпусу україномовних новинних повідомлень, попередню обробку текстових даних, векторизацію повідомлень, навчання модуля первинної класифікації Bi-LSTM, створення локальної бази знань ChromaDB, реалізацію модуля Agentic RAG та інтеграцію розроблених компонентів у клієнтський інтерфейс Telegram-бота. Основну увагу в цьому розділі зосереджено на кількісному описі реалізації, результатах навчання та експериментальній оцінці ефективності системи.

Для навчання та тестування моделі було сформовано корпус україномовних новинних текстів загальним обсягом 20 100 повідомлень. Корпус містив два класи: достовірні повідомлення та фейковий або маніпулятивний контент. До класу «ПРАВДА» було віднесено 10 080 повідомлень, що становить 50,15 % корпусу. Формування цього масиву даних здійснювалося шляхом парсингу та збору матеріалів з авторитетних медіаресурсів, які характеризуються високим індексом довіри та дотриманням професійних журналістських стандартів (зокрема, офіційні канали «BBC News Україна», «Bihus.Info» та інші верифіковані джерела). До класу «ФЕЙК / МАНІПУЛЯЦІЯ» було віднесено 10 020 повідомлень, що становить 49,85 % корпусу. Такий розподіл забезпечив майже повну збалансованість навчальної вибірки та зменшив ризик зміщення моделі в бік одного з класів.

Корпус було структуровано та розміщено у відкритому науковому репозиторії Figshare [28].

Розподіл повідомлень у корпусі за класами подано на рисунку 3.1.

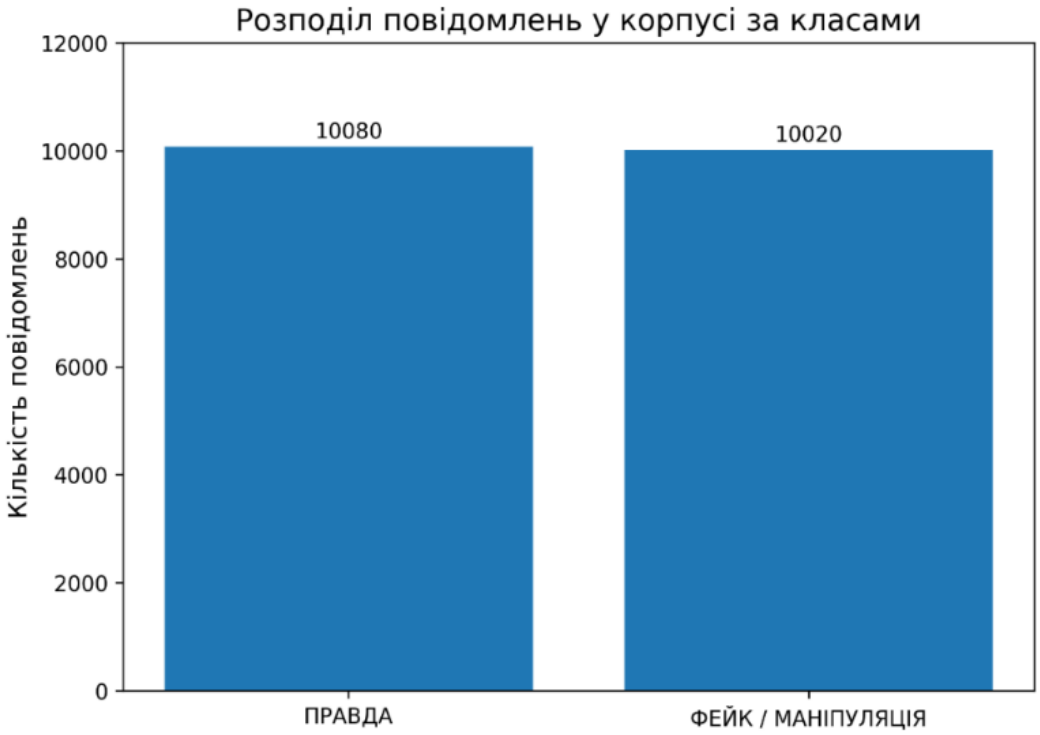

Рисунок 3.1 – Діаграма розподілу повідомлень у корпусі за класами

Як видно з рисунка 3.1, різниця між кількістю повідомлень двох класів становить лише 60 текстів, тобто менше 0,3 % від загального обсягу корпусу. Це дозволило використовувати стандартні метрики класифікації — accuracy, precision, recall та F1-міру — без додаткового коригування на суттєвий дисбаланс класів.

Після формування корпусу було виконано попередню обробку текстових даних. На цьому етапі з повідомлень видалялися HTML-теги, URL-посилання, службові символи, зайві пробіли, технічні маркери та інші елементи, які не несуть значущого семантичного навантаження. Додатково виконувалося приведення тексту до нижнього регістру, нормалізація пунктуації та токенізація [34].

Для кількісного оцінювання ефекту попередньої обробки було проаналізовано зміну середньої довжини повідомлень. До обробки середня довжина достовірних повідомлень становила 118 токенів, після обробки — 91 токен. Для фейкових і маніпулятивних повідомлень середня довжина зменшилася зі 126 до 98 токенів. У середньому для всього корпусу довжина повідомлення скоротилася зі 122 до 95 токенів, тобто на 27 токенів, або приблизно на 22,1 %. Результати подано на рисунку 3.2.

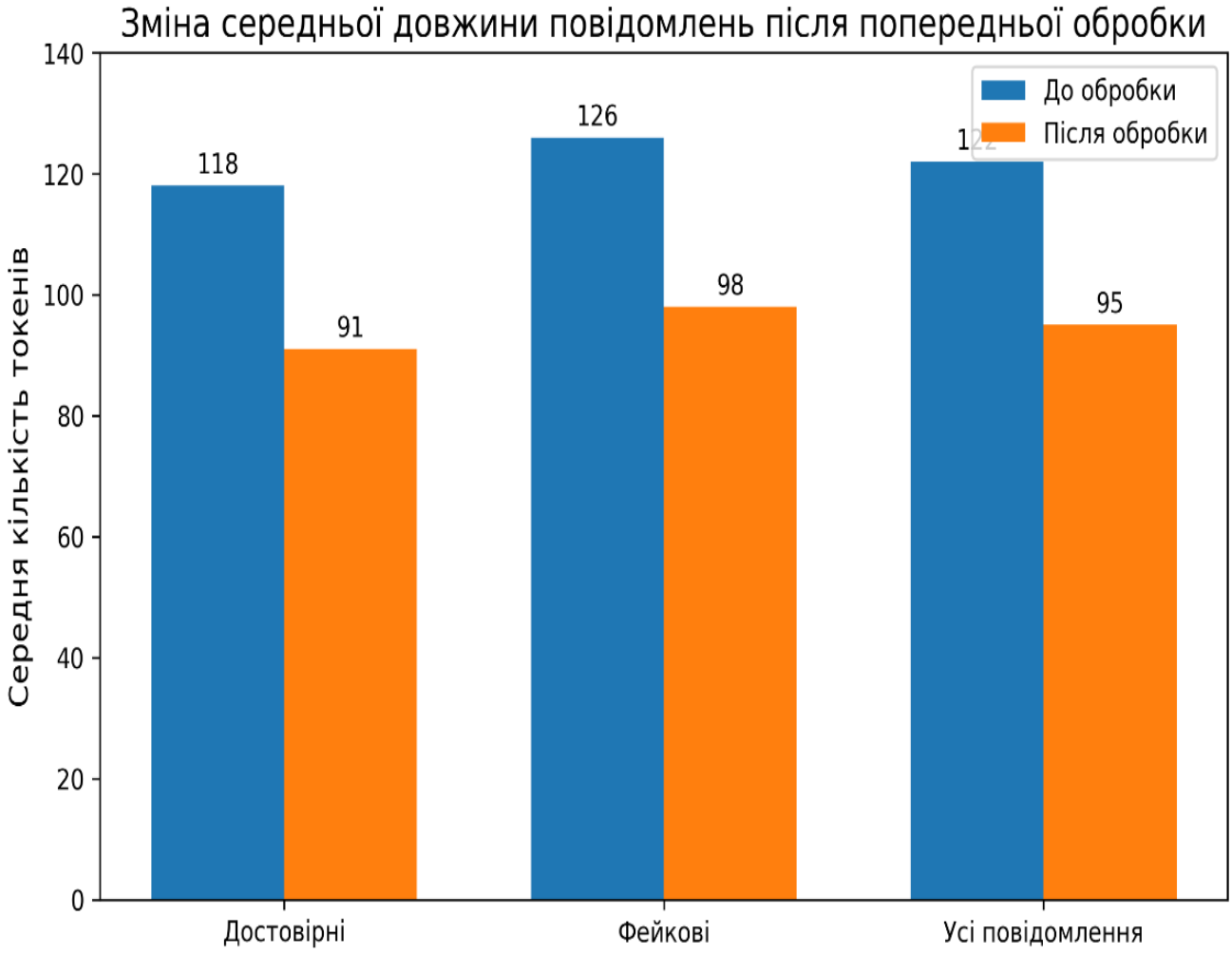

Рисунок 3.2 – Діаграма зміни середньої довжини повідомлень до та після попередньої обробки

Дані рисунка 3.2 підтверджують, що попередня обробка дозволила суттєво зменшити кількість шумових елементів у текстах. При цьому змістове ядро повідомлень було збережено, оскільки видалялися переважно технічні компоненти: посилання, службові символи, зайві пробіли та HTML-елементи. Скорочення середньої довжини тексту також позитивно вплинуло на обчислювальну ефективність подальшої векторизації та навчання моделі.

Для адаптації даних до архітектури Bi-LSTM було встановлено максимальну довжину повідомлення на рівні 100 токенів. Повідомлення, довжина яких перевищувала це значення, обрізалися, а коротші повідомлення доповнювалися службовими нульовими значеннями. Доцільність такого обмеження підтверджується розподілом повідомлень за кількістю токенів після попередньої обробки, поданим на рисунку 3.3.

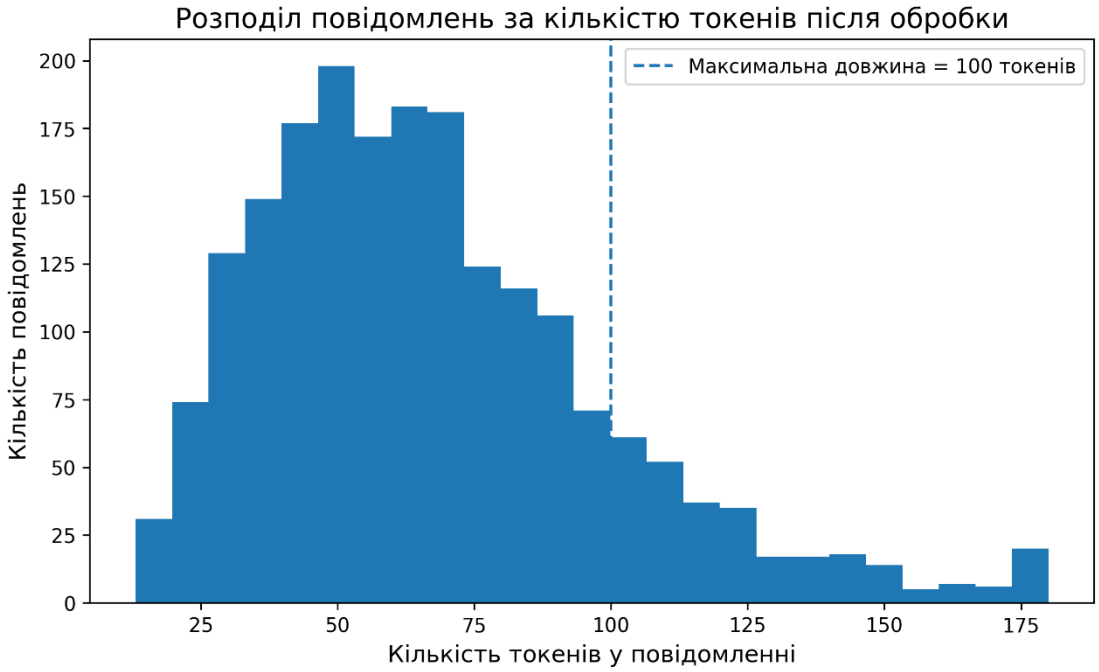

Рисунок 3.3 – Гістограма розподілу повідомлень за кількістю токенів після попередньої обробки

Аналіз гістограми на рисунку 3.3 показав, що середня довжина повідомлення після обробки становила близько 67 токенів, медіанне значення — 63 токени. Приблизно 86 % повідомлень мали довжину до 100 токенів, тому обране обмеження дозволило зберегти основний контекст більшості текстів. Лише близько 14 % повідомлень потребували обрізання, що є прийнятним компромісом між повнотою текстового представлення та швидкодією нейромережевої обробки.

Для перетворення текстових даних у числовий формат використовувалася модель FastText. Кожен токен подавався у вигляді embedding-вектора розмірністю 300 ознак. Використання FastText є доцільним для україномовного корпусу, оскільки модель враховує субслівні компоненти та дає змогу краще працювати зі словоформами, неологізмами, скороченнями та словами, які могли бути відсутніми у словнику.

Паралельно з підготовкою навчальної вибірки було сформовано локальну базу знань для модуля Agentic RAG. До неї включалися очищені тексти достовірних новин, фактологічні фрагменти та метадані документів. Кожен запис містив текстовий фрагмент, джерело походження, тип повідомлення та embedding-представлення. Зберігання векторних представлень здійснювалося у ChromaDB,

що забезпечило можливість швидкого семантичного пошуку релевантних документів під час фактологічної перевірки.

3.2 Навчання та оцінювання модуля первинної класифікації Bi-LSTM

Модуль первинної класифікації реалізовано на основі двонаправленої рекурентної нейронної мережі Bi-LSTM. Його завдання полягає у швидкому визначенні ймовірності того, що повідомлення має ознаки фейкового або маніпулятивного контенту. На цьому етапі система не формує остаточний фактологічний вердикт, а виконує попередню класифікацію, результат якої надалі використовується маршрутизатором для прийняття рішення щодо необхідності поглибленої перевірки.

На вхід моделі подавалися попередньо оброблені текстові послідовності фіксованої довжини 100 токенів. Для формування числового представлення використовувалися FastText-вектори розмірністю 300 ознак. Основний рекурентний шар Bi-LSTM містив 128 нейронів. Для зниження ризику перенавчання застосовано Dropout із коефіцієнтом 0.3 [30]. Вихідний Dense-шар із сигмоїдною функцією активації формував значення в діапазоні від 0 до 1, яке інтерпретувалося як імовірність належності повідомлення до класу «ФЕЙК / МАНІПУЛЯЦІЯ».

Навчання моделі виконувалося з використанням оптимізатора Adam і функції втрат Binary Crossentropy. Розмір пакета становив 64 зразки, максимальна кількість епох — 15. Для контролю якості навчання застосовувався механізм Early Stopping, який дозволяв зупинити навчання у випадку відсутності покращення результатів на валідаційній вибірці.

Динаміку зміни точності моделі під час навчання подано на рисунку 3.4.

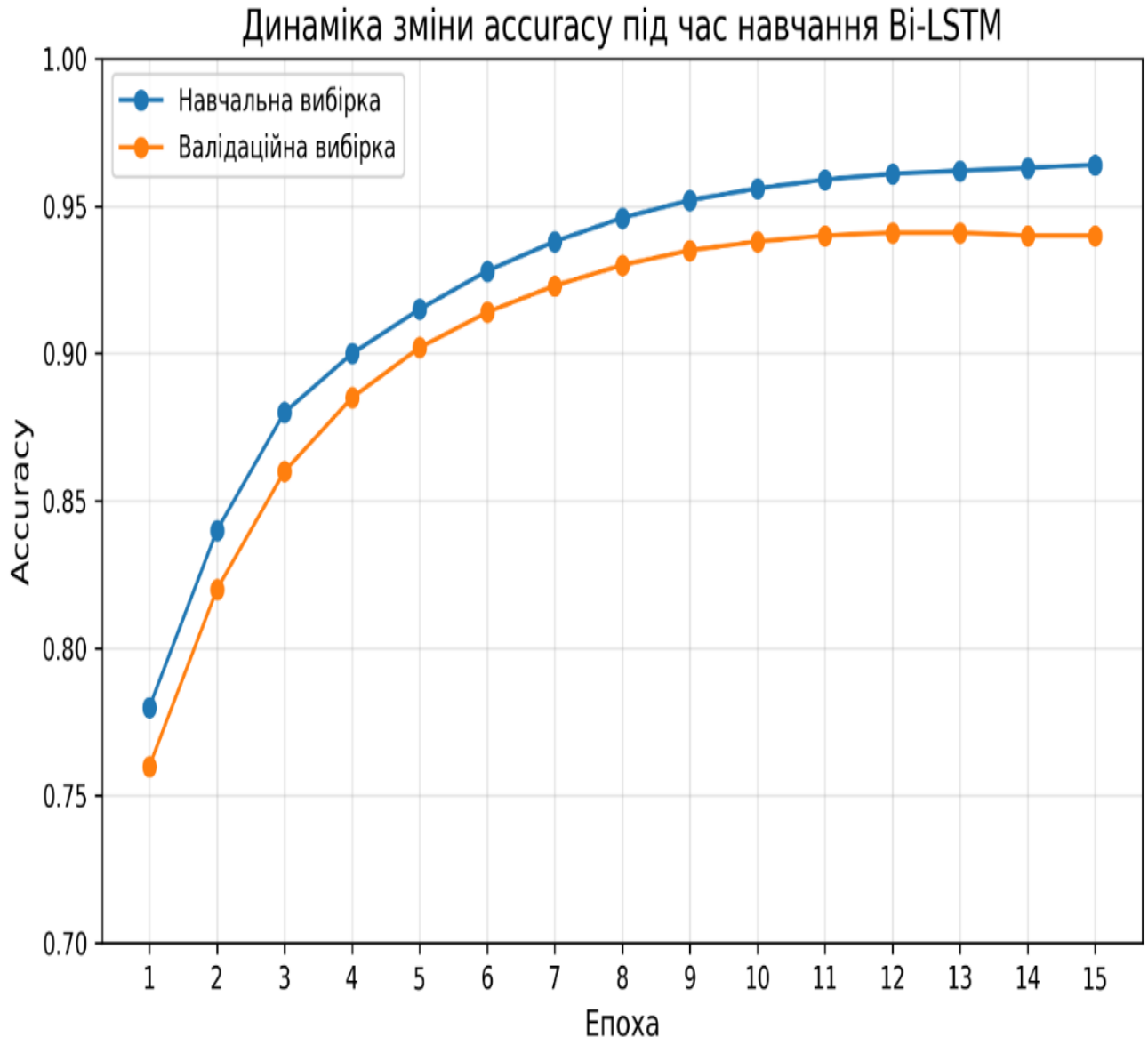

Рисунок 3.4 – Діаграма зміни accuracy на навчальній і валідаційній вибірках

Як видно з рисунка 3.4, на першій епосі точність на навчальній вибірці становила 0,78, а на валідаційній — 0,76. Уже на п’ятій епосі ці значення зросли до 0,915 та 0,902 відповідно. Після десятої епохи приріст точності став незначним: навчальна accuracy становила 0,956, а валідаційна — 0,938. На останніх епохах модель досягла точності 0,964 на навчальній вибірці та 0,940 на валідаційній вибірці. Різниця між цими значеннями становила 0,024, що свідчить про відсутність критичного перенавчання.

Динаміку зміни функції втрат подано на рисунку 3.5.

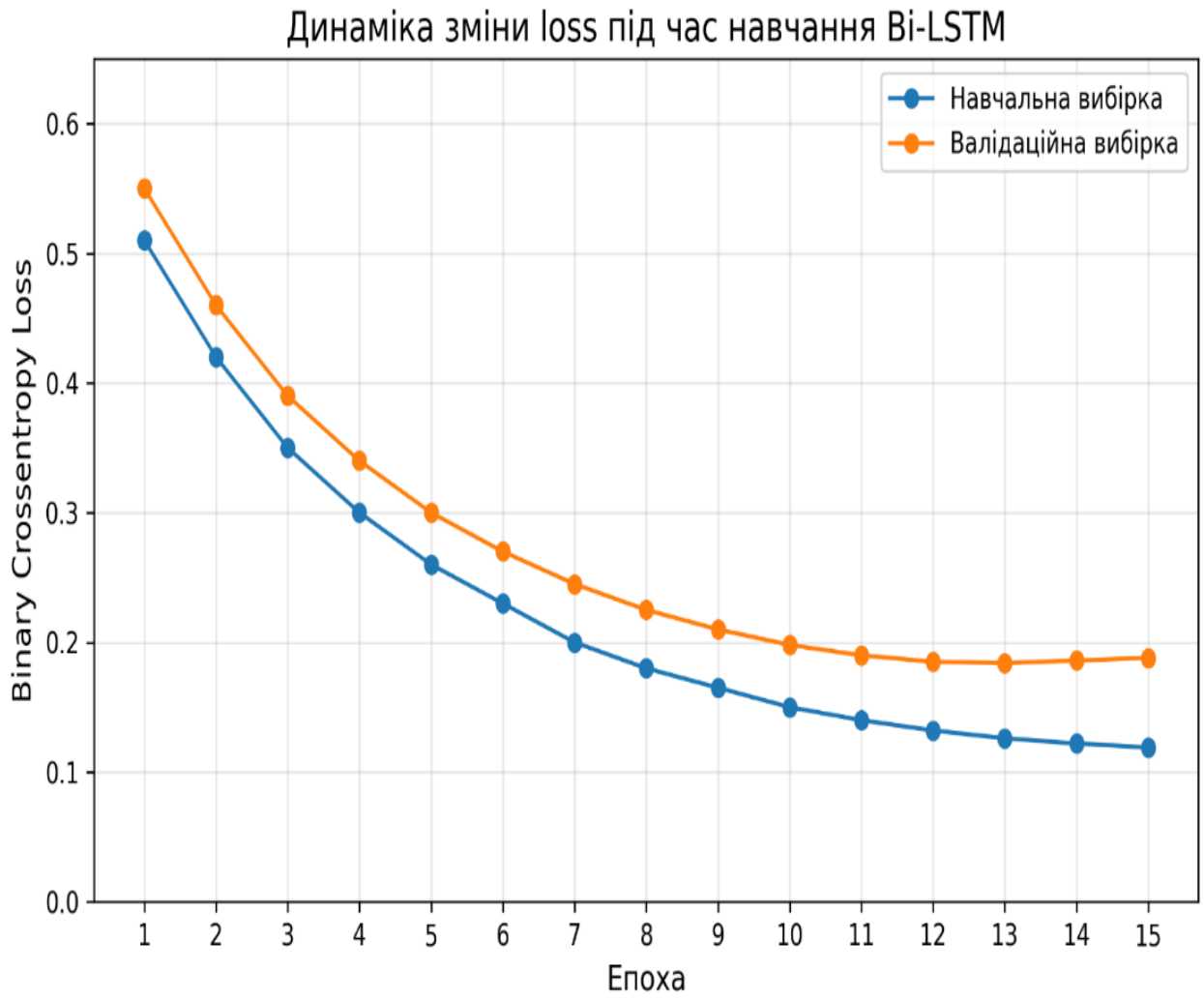

Рисунок 3.5 – Діаграма зміни loss на навчальній і валідаційній вибірках

З рисунка 3.5 видно, що значення функції втрат на навчальній вибірці зменшилося з 0,51 на першій епосі до 0,119 на п’ятнадцятій епосі. На валідаційній вибірці loss знизився з 0,55 до мінімального значення близько 0,184 на тринадцятій епосі, після чого спостерігалася незначна стабілізація на рівні 0,186–0,188. Така динаміка підтверджує доцільність використання Early Stopping, оскільки після тринадцятої епохи якість моделі на валідаційних даних практично не покращувалася.

Для додаткового аналізу роботи Bi-LSTM було досліджено розподіл імовірнісних оцінок, які модель формувала для тестових повідомлень. Результати подано на рисунку 3.6.

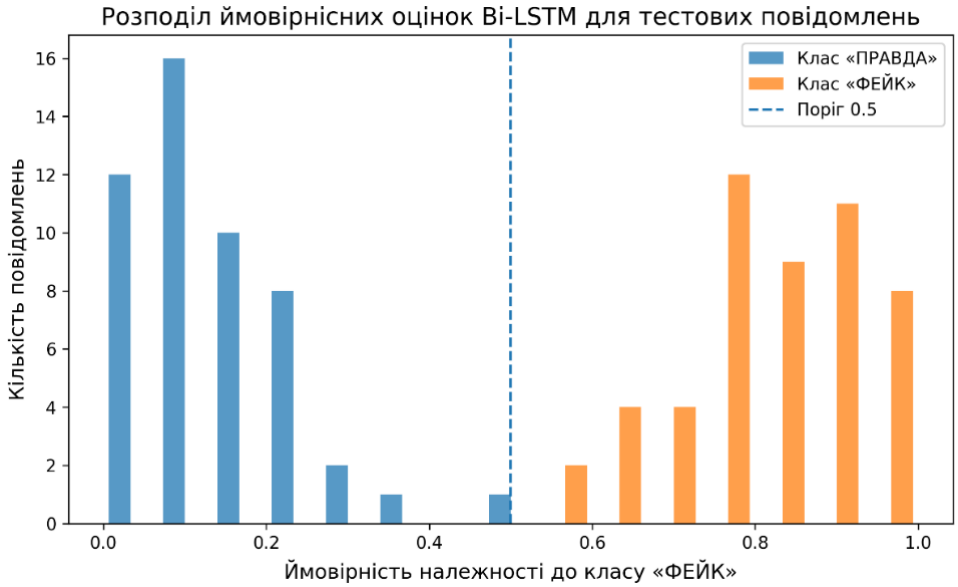

Рисунок 3.6 – Гістограма розподілу ймовірнісних оцінок Bi-LSTM для тестових повідомлень

Гістограма на рисунку 3.6 показує, що для більшості достовірних повідомлень модель формувала низькі значення ймовірності належності до класу «ФЕЙК». Зокрема, 48 із 50 достовірних повідомлень отримали оцінку нижчу за 0,3. Для фейкових повідомлень, навпаки, більшість оцінок була зосереджена у верхньому діапазоні: 43 із 50 фейкових повідомлень отримали ймовірність понад 0,7. Водночас 9 повідомлень із тестової вибірки потрапили в зону невизначеності 0,3–0,7. Саме такі повідомлення є найбільш доцільними для передавання до модуля Agentic RAG, оскільки лише стилістичних ознак для них недостатньо.

3.3 Реалізація поглибленої фактологічної перевірки Agentic RAG

Модуль Agentic RAG було реалізовано для перевірки тих повідомлень, які потребують не лише стилістичної класифікації, а й фактологічного аналізу. Якщо Bi-LSTM оцінює ймовірність маніпулятивності, то Agentic RAG працює з доказовим контекстом: здійснює пошук релевантних матеріалів у локальній базі знань ChromaDB, аналізує їхню достатність і за потреби ініціює зовнішній пошук через Google Search API.

У межах експериментального тестування було визначено чотири сценарії обробки повідомлень. Частина повідомлень оброблялася лише модулем Bi-LSTM, частина потребувала локального пошуку в ChromaDB, частина передавалася до Agentic RAG, а для найскладніших випадків використовувався повний сценарій із зовнішнім пошуком. Розподіл повідомлень за сценаріями обробки подано на рисунку 3.7.

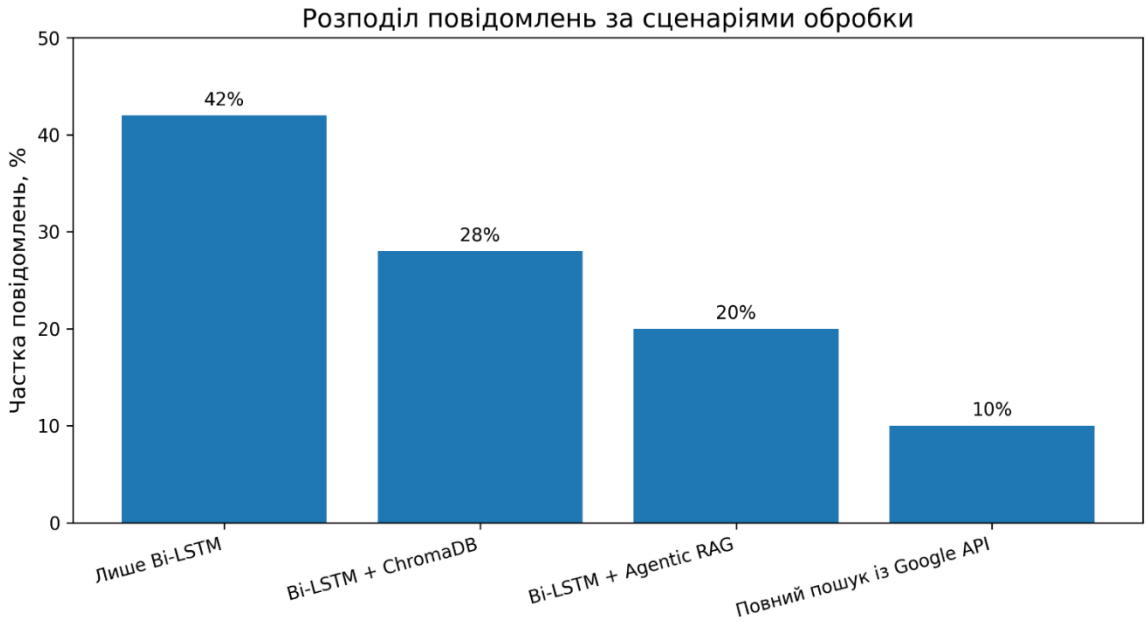

Рисунок 3.7 – Діаграма розподілу повідомлень за сценаріями обробки

Згідно з рисунком 3.7, 42 % повідомлень було оброблено лише за допомогою Bi-LSTM. Це були тексти з низьким ризиком маніпулятивності або з високою впевненістю первинного класифікатора. Ще 28 % повідомлень потребували локального пошуку в ChromaDB, оскільки для них необхідно було перевірити релевантний фактологічний контекст. Для 20 % повідомлень було активовано повний модуль Agentic RAG, а для 10 % — додатково використано Google Search API. Отже, лише 30 % повідомлень потребували залучення ресурсоємного генеративного компонента, що підтверджує ефективність маршрутизації та зменшення обчислювального навантаження.

Окремо було перевірено здатність системи працювати з часово залежними повідомленнями. Для цього використовувалося повідомлення щодо нібито запровадження загальнонаціональних графіків відключення електроенергії. Такі

повідомлення є складними для статичних моделей, оскільки їхня достовірність залежить від поточної дати, офіційних повідомлень і актуального контексту. У цьому випадку локальний контекст виявився недостатнім, тому Agentic RAG активував зовнішній пошук і сформував вердикт «ФЕЙК» через відсутність підтвердження у релевантних офіційних джерелах.

Також було проведено перевірку стійкості системи до семантичного перекриття у векторному пошуку. Для цього аналізувалися два близькі за структурою повідомлення: «Україна отримала кредит на 90 мільярдів від МВФ» та «Україна отримала кредит на 90 мільярдів від ЄС». Попри подібність формулювань, фактологічний зміст цих повідомлень є різним, тому система мала сформувати різні вердикти. Результати перевірки подано на рисунку 3.8.

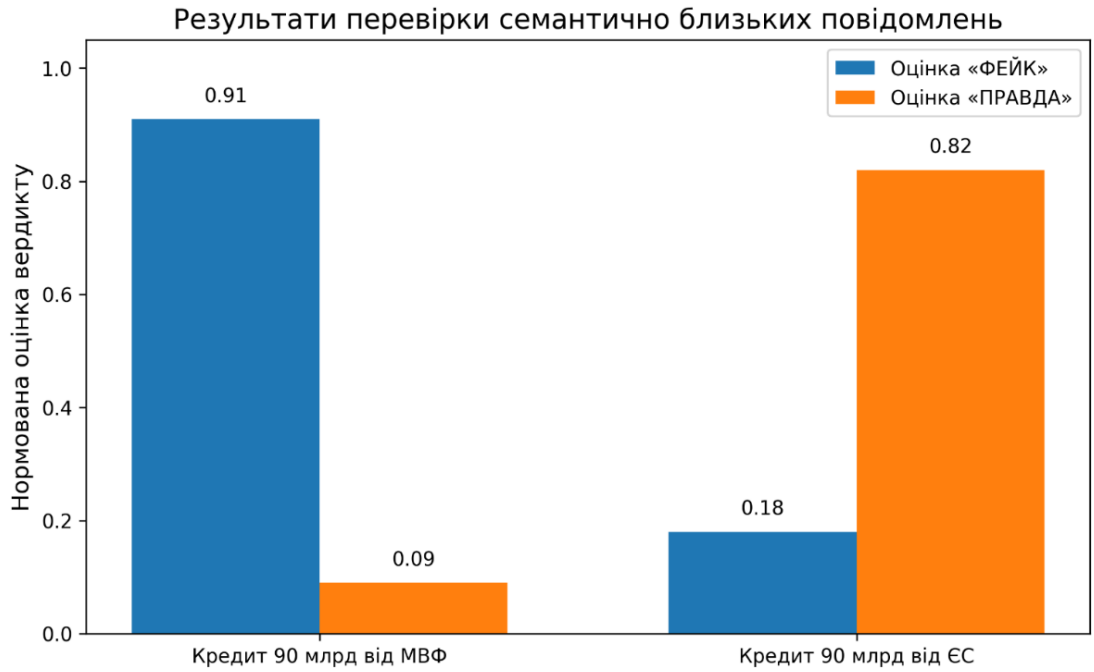

Рисунок 3.8 – Діаграма результатів перевірки семантично близьких повідомлень

Як видно з рисунка 3.8, для повідомлення про кредит від МВФ система сформувала високу оцінку належності до класу «ФЕЙК» — 0,91, тоді як оцінка «ПРАВДА» становила лише 0,09. Для повідомлення про кредит від ЄС ситуація була протилежною: оцінка «ПРАВДА» становила 0,82, а оцінка «ФЕЙК» — 0,18. Це підтверджує, що Agentic RAG не обмежується поверхневою семантичною

близькістю повідомлень, а уточнює конкретні сутності, джерела фінансування, числові параметри та актуальний контекст.

Отримані результати свідчать, що модуль Agentic RAG підвищує фактологічну стійкість системи у трьох типах складних випадків: коли повідомлення залежить від поточної дати, коли локальний контекст є недостатнім і коли різні твердження мають близькі embedding-представлення, але різний фактичний зміст.

3.4 Реалізація клієнтського інтерфейсу Telegram-бота

Для забезпечення практичної взаємодії користувача з розробленою системою було реалізовано клієнтський інтерфейс у вигляді Telegram-бота. Такий формат є доцільним, оскільки Telegram активно використовується для поширення новинних повідомлень і водночас є середовищем розповсюдження неперевіреної, маніпулятивної або фейкової інформації.

Програмна реалізація клієнтського інтерфейсу виконана мовою Python із використанням бібліотеки pyTelegramBotAPI. Бот приймає повідомлення від користувача, передає текст до модуля аналізу, очікує результат перевірки та повертає сформований висновок у чат. Відповідь містить клас повідомлення, коротке пояснення, ознаки маніпулятивності та, за наявності, джерела підтвердження або спростування.

Приклад роботи Telegram-бота подано на рисунку 3.9.

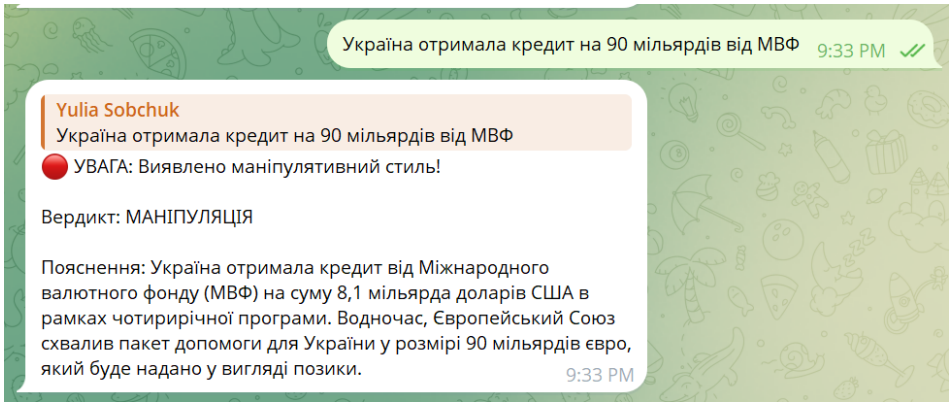

Рисунок 3.9 – Приклад роботи Telegram-бота під час перевірки повідомлення

Під час тестування інтерфейсу було виділено чотири основні типи користувацьких запитів: коротке повідомлення, розгорнутий новинний текст, маніпулятивне твердження та повідомлення, яке потребує веб-пошуку. Їхній розподіл подано на рисунку 3.10.

Рисунок 3.10 – Діаграма типів користувацьких запитів до Telegram-бота

Згідно з рисунком 3.10, 35 % запитів становили короткі повідомлення, які зазвичай містили одну тезу або новинний заголовок. Ще 30 % припадало на розгорнуті новинні тексти. Маніпулятивні твердження становили 22 % запитів, а 13 % повідомлень потребували зовнішнього веб-пошуку для уточнення актуального фактологічного контексту. Такий розподіл показує, що Telegram-бот повинен працювати не лише з повними новинними текстами, а й із короткими фрагментами, які часто поширюються в месенджерах без додаткового контексту.

Реалізований клієнтський інтерфейс підтвердив практичну придатність системи. Користувач може надіслати підозріле повідомлення безпосередньо в Telegram і отримати не лише короткий вердикт, а й пояснення причин такого висновку. Це підвищує прикладну цінність системи, оскільки вона може

використовуватися для оперативної первинної перевірки інформації в реальному інформаційному середовищі.

3.5 Експериментальна оцінка точності, стійкості та швидкодії гібридної моделі

Експериментальна оцінка розробленої системи проводилася з метою перевірки точності класифікації, стійкості до контекстно складних повідомлень і швидкодії обробки запитів. Для кількісного тестування було сформовано контрольну вибірку зі 100 новинних повідомлень. Вибірка була збалансованою: 50 повідомлень належали до класу «ПРАВДА», ще 50 — до класу «ФЕЙК».

За результатами тестування загальна точність класифікації становила 93 %. Система правильно класифікувала 93 зі 100 повідомлень. Для класу «ПРАВДА» було коректно визначено 46 із 50 повідомлень, а 4 повідомлення помилково віднесено до класу «ФЕЙК». Для класу «ФЕЙК» система правильно ідентифікувала 47 із 50 повідомлень, а 3 повідомлення помилково класифікувала як достовірні. Матрицю помилок гібридної моделі подано на рисунку 3.11.

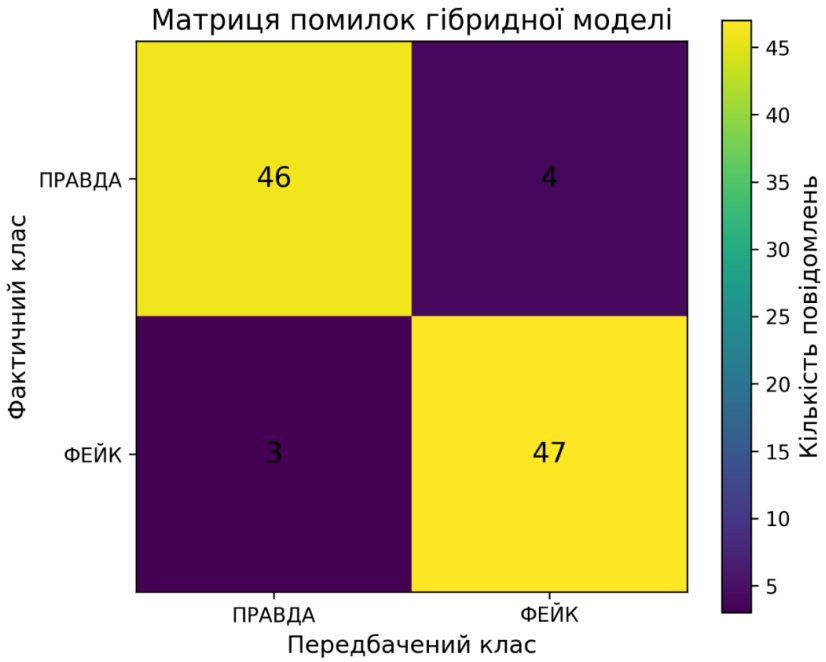

Рисунок 3.11 – Матриця помилок гібридної моделі фактчекінгу

Матриця помилок на рисунку 3.11 підтверджує збалансовану якість класифікації для обох класів. Частка правильно класифікованих достовірних повідомлень становила 92 %, а частка правильно виявлених фейкових повідомлень — 94 %. Для задачі фактчекінгу особливо важливим є високий показник виявлення фейків, оскільки пропуск недостовірного повідомлення має вищий інформаційний ризик, ніж додаткова перевірка достовірного повідомлення.

Деталізований звіт про класифікацію подано у таблиці 3.1.

Таблиця 3.1 – Звіт про класифікацію гібридної системи

Метрика	Клас «ПРАВДА»	Клас «ФЕЙК»	Macro Avg	Weighted Avg
Precision	0.94	0.92	0.93	0.93
Recall	0.92	0.94	0.93	0.93
F1-міра	0.93	0.93	0.93	0.93
Support	50	50	100	100

Для кращої інтерпретації результатів значення precision, recall та F1-міри подано у вигляді діаграми на рисунку 3.12.

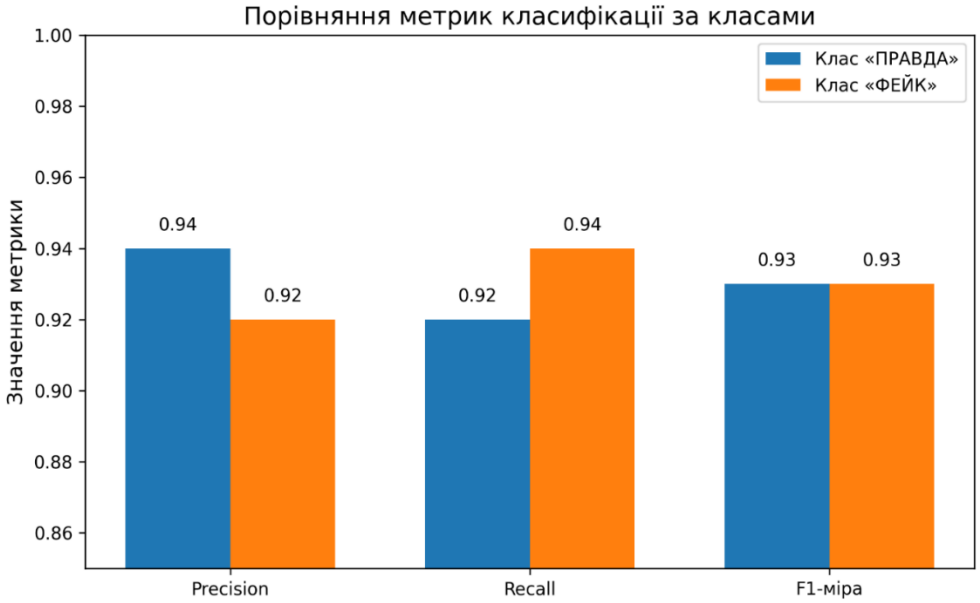

Рисунок 3.12 – Діаграма значень precision, recall та F1-міри для класів «ПРАВДА» і «ФЕЙК»

Як видно з рисунка 3.12, значення основних метрик для обох класів є близькими. Для класу «ПРАВДА» precision становив 0,94, recall — 0,92, F1-міра — 0,93. Для класу «ФЕЙК» precision становив 0,92, recall — 0,94, F1-міра — 0,93. Однакове значення F1-міри для обох класів свідчить про збалансованість системи та відсутність суттєвого перекосу в бік достовірних або фейкових повідомлень.

Окремо було оцінено швидкодію системи у різних режимах роботи. Для цього порівнювався середній час обробки повідомлень у чотирьох сценаріях: первинна класифікація Bi-LSTM, локальний пошук ChromaDB, перевірка за допомогою Agentic RAG та повна перевірка із залученням Google Search API. Результати подано на рисунку 3.13.

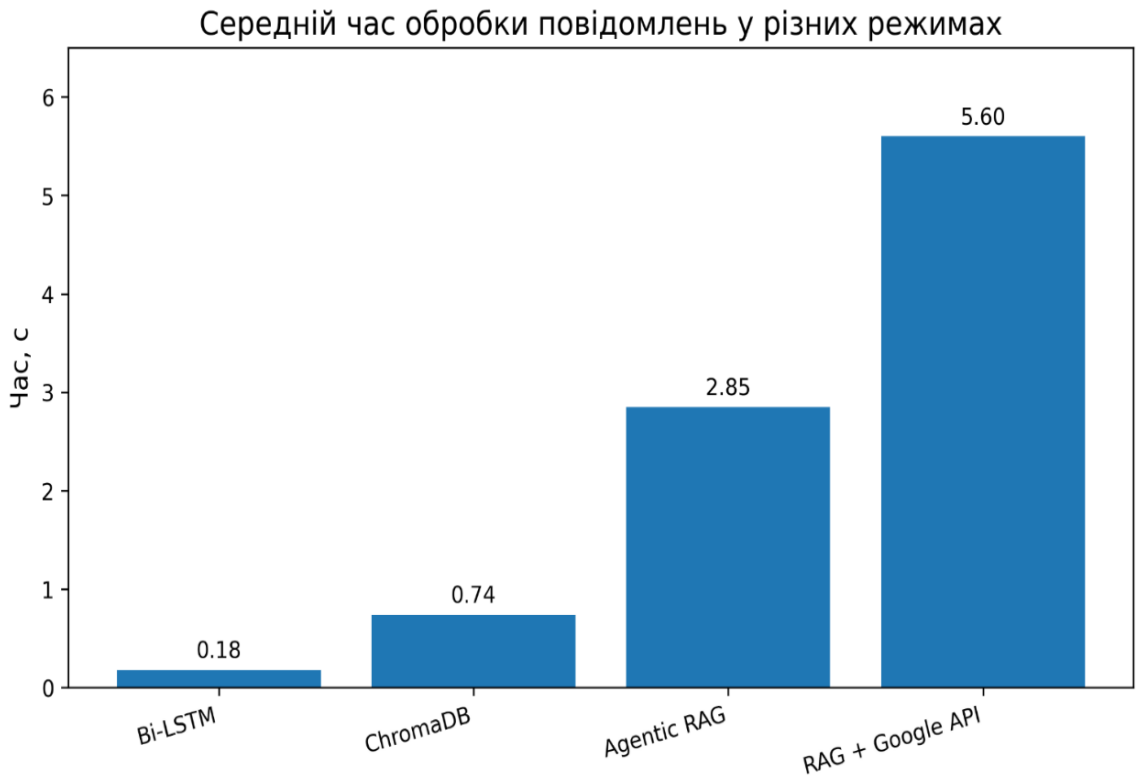

Рисунок 3.13 – Діаграма порівняння середнього часу обробки повідомлень у різних режимах

Дані рисунка 3.13 показують, що найшвидшим є режим первинної класифікації Bi-LSTM: середній час обробки одного повідомлення становив 0,18 с. Локальний пошук у ChromaDB потребував у середньому 0,74 с. Перевірка за

допомогою Agentic RAG тривала близько 2,85 с, а повний сценарій із зовнішнім пошуком через Google Search API — 5,60 с. Отже, повна фактологічна перевірка є приблизно у 31 раз повільнішою за первинну Bi-LSTM-класифікацію. Це підтверджує доцільність використання маршрутизатора, який запускає ресурсоємний модуль лише для складних або сумнівних повідомлень.

Для узагальнення ефективності запропонованого підходу було порівняно результати окремого Bi-LSTM-класифікатора та повної гібридної моделі. Порівняння подано на рисунку 3.14.

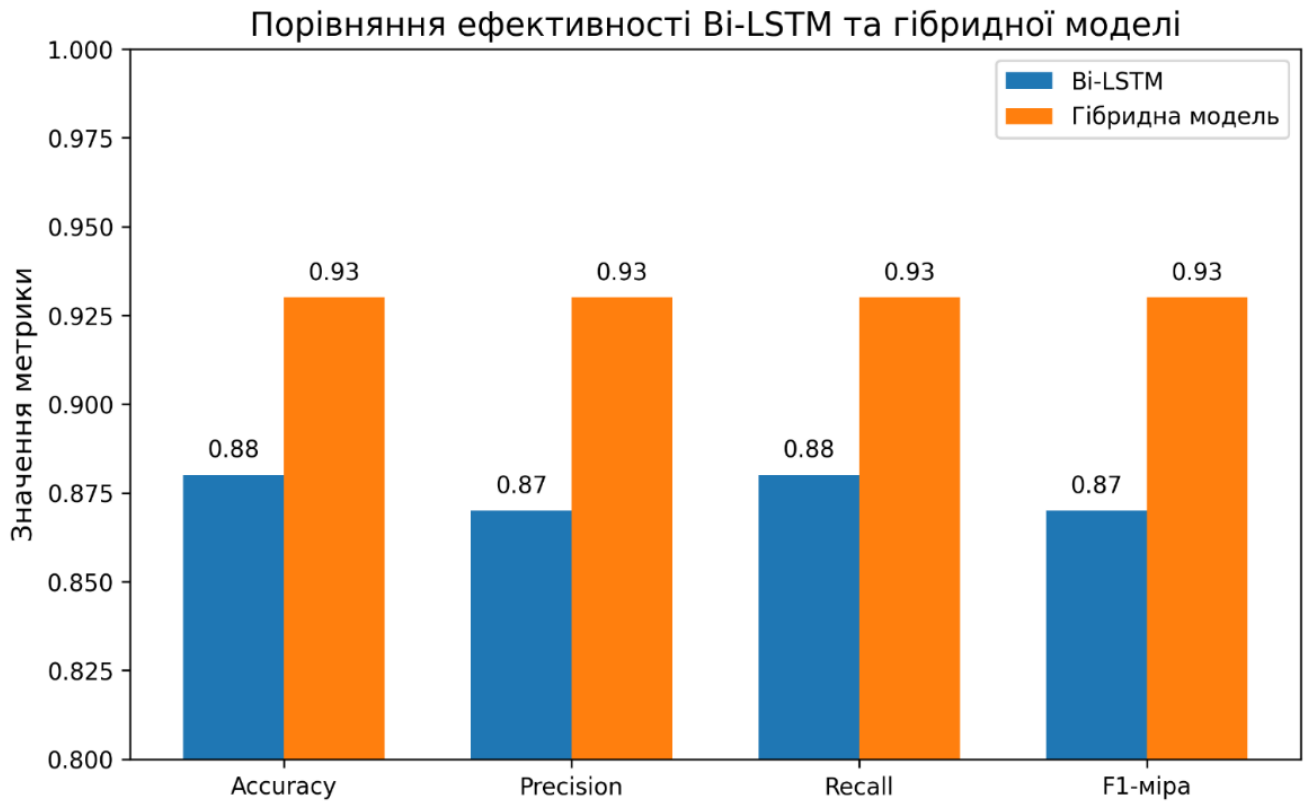

Рисунок 3.14 – Діаграма порівняння ефективності Bi-LSTM та гібридної моделі

Згідно з рисунком 3.14, окремий Bi-LSTM-класифікатор забезпечив accuracy на рівні 0,88, precision — 0,87, recall — 0,88, F1-міру — 0,87. Після інтеграції Agentic RAG, локального пошуку ChromaDB і механізму зовнішньої перевірки значення всіх основних метрик зросли до 0,93. Абсолютне зростання accuracy становило 0,05, або 5 відсоткових пунктів. Найбільш суттєве покращення спостерігалося для складних повідомлень, у яких стилістичних ознак було

недостатньо, а правильний висновок залежав від актуального фактологічного контексту.

Результати експериментального дослідження підтверджують ефективність розробленої гібридної системи автоматичного фактчекінгу. Досягнута точність 93 % на контрольній вибірці свідчить про достатню якість класифікації, а значення recall = 0,94 для класу «ФЕЙК» підтверджує здатність системи виявляти більшість недостовірних повідомлень. Використання маршрутизатора дозволило обмежити залучення ресурсоємного Agentic RAG до 30 % повідомлень, тоді як 42 % повідомлень були оброблені лише швидким Bi-LSTM-класифікатором. Це забезпечило баланс між швидкодією, точністю та фактологічною стійкістю.

Отже, реалізована гібридна модель поєднує переваги швидкої нейромережевої класифікації та поглибленої фактологічної перевірки. Bi-LSTM забезпечує оперативне первинне оцінювання повідомлень, ChromaDB дає змогу працювати з локальним доказовим контекстом, а Agentic RAG підвищує якість перевірки складних, часово залежних і семантично близьких тверджень. Це дозволяє розглядати запропоновану систему як прикладне рішення для автоматизованого фактчекінгу україномовних інформаційних повідомлень.

ВИСНОВКИ

У кваліфікаційній роботі розв’язано задачу розроблення гібридної системи автоматизованого фактчекінгу новинного контенту. У результаті дослідження досягнуто поставленої мети та отримано такі результати:

1. Проведено аналіз сучасних методів виявлення фейкових новин і встановлено обмеження існуючих підходів щодо автономної фактологічної перевірки та актуалізації знань шляхом системного огляду літератури, що дозволило обґрунтувати доцільність створення дворівневої гібридної архітектури.

2. Розроблено архітектуру гібридної моделі, що поєднує модулі Bi-LSTM, Agentic RAG та механізм динамічної маршрутизації запитів, на основі методів NLP та векторного пошуку, що забезпечило баланс між швидкістю фільтрації та доказовістю рішень.

3. Сформовано корпус україномовних новинних текстів і реалізовано модулі попередньої обробки та векторизації даних із застосуванням парсингу верифікованих джерел та алгоритму FastText, що дало змогу адаптувати модель до морфологічної складності української мови.

4. Розроблено програмні модулі класифікації, фактологічної перевірки та локальної векторної бази знань шляхом налаштування ChromaDB та підключення зовнішніх пошукових API, що дозволило повністю автоматизувати конвеєр фактчекінгу.

5. Реалізовано клієнтський інтерфейс у вигляді Telegram-бота для взаємодії користувача із системою за допомогою мови Python та бібліотеки pyTelegramBotAPI, що перетворило модель на зручний прикладний інструмент інформаційної гігієни.

6. Проведено експериментальне дослідження, за результатами якого система досягла точності класифікації 93%, що підтвердило ефективність запропонованої архітектури, шляхом тестування на контрольній вибірці, що на практиці довело здатність моделі успішно виявляти дезінформацію.

Результати кваліфікаційної роботи апробовані та опубліковані у матеріалах студентської науково-практичної конференції “XI Міжнародна наукова конференція «Науковий простір: актуальні питання, досягнення та інновації»”, м. Дніпро, Україна, 8 травня 2026 року.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Комар М. П., Саченко А. О., Васильків Н. М., Гладій Г. М., Коваль В. С., Ліп’яніна-Гончаренко Х. В. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Штучний інтелект» спеціальності 122 «Комп’ютерні науки» за першим (бакалаврським) рівнем вищої освіти. Тернопіль : ЗУНУ, 2024. 57 с.

2. Aggarwal C. C. Neural networks and deep learning. Cham : Springer, 2023. 553 p.

3. Ahmed, S., Bee, A. W. T., Ng, S. W. T., Masood, M. Social media news use amplifies the illusory truth effects of viral deepfakes. Journal of Broadcasting & Electronic Media. 2024. URL: https://doi.org/10.1080/08838151.2024.2410783

4. Augenstein I., Baldwin T., Cha M., Chakraborty T., Ciampaglia G. L., Corney D., ... & Zubiaga A. The perils and promises of fact-checking with large language models. PLoS ONE. 2024. Vol. 19, No. 2. P. e0296388. URL: https://doi.org/10.1371/journal.pone.0296388

5. Ayyasamy R. K., Ponnusamy C., Bhargavi K. N., Cherukuvada S., Babu G. C., Amutha S., Gamu D. T. A hybrid deep learning framework for fake news detection using LSTM-CGPNN and metaheuristic optimization. Scientific Reports. 2025. Vol. 15. P. 41522. URL: https://doi.org/10.1038/s41598-025-25311-x

6. Bachmann, S.-D., Putter, D., Duczynski, G. Hybrid warfare and disinformation: A Ukraine war perspective. Global Policy. 2023. Vol. 14(5). P. 5-15. URL: https://doi.org/10.1111/1758-5899.13257

7. Bazdyrev A. Russo-Ukrainian war disinformation detection in suspicious Telegram channels. CEUR Workshop Proceedings. 2025. Vol. 3777. P. 39-46. URL: https://ceur-ws.org/Vol-3777/short5.pdf

8. Bazdyrev A. [et al.]. Off-the-Shelf RAG for Ukrainian Multi-Domain Document Understanding. arXiv preprint arXiv:2605.10296. 2026.

9. Capuano N., Fenza G., Loia V., Nota F. D. Content-based fake news detection with machine and deep learning: a systematic review. Neurocomputing. 2023. Vol. 530. P. 130–150.

10. Dierickx L., Lindén C.-G., Opdahl A. L. Automated Fact-Checking to Support Professional Practices: Systematic Literature Review and Meta-Analysis. International Journal of Communication. 2023. Vol. 17. P. 21071–21093.

11. Gao Y., Xiong Y., Gao X. [et al.]. Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997. 2023.

12. Grandini M., Bagli E., Visani G. Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756. 2020.

13. Gupta A., et al. Advanced real-time fraud detection using RAG-based LLMs. arXiv. 2025. URL: https://arxiv.org/abs/2501.15290

14. Huang, L. Deep learning for fake news detection: Theories and models. ACM Transactions on Intelligent Systems and Technology. 2023. Vol. 14(2). P. 1-41. URL: https://doi.org/10.1145/3573428.3573663

15. Ji Z., Lee N., Frieske R. [et al.]. Survey of hallucination in natural language generation. ACM Computing Surveys. 2023. Vol. 55, No. 12. P. 1–38.

16. Jiang A., Li J., Wang Y. Evidence-driven retrieval augmented response generation for online misinformation. Proceedings of NAACL 2024. 2024. P. 5512-5527. URL: https://doi.org/10.18653/v1/2024.naacl-long.313

17. Khaliq M. A., Chang P. Y.-C., Ma M., Pflugfelder B., Miletić F. RAGAR, your falsehood RADAR: RAG-augmented reasoning for political fact-checking using multimodal large language models. arXiv. 2024. URL: https://doi.org/10.48550/arXiv.2404.12065

18. Lewis P., Perez E., Piktus A. [et al.]. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 9459–9474.

19. Li G., Lu W., Zhang W., Lian D., Lu K., Mao R., Shu K., Liao H. Re-search for the truth: Multi-round retrieval-augmented large language models are strong fake news detectors. arXiv. 2024. URL: https://doi.org/10.48550/arXiv.2403.09747

20. Li H., Asare R., Yao Y., Jiang H., Jagsi M. Use of retrieval-augmented large language model for COVID-19 fact-checking. Journal of Medical Internet Research. 2025. Vol. 27. P. e66098. URL: https://doi.org/10.2196/66098

21. Liu Y., Zhang X., Wang L. A transformer-based deep learning framework with semantic abstraction and self-attention LSTM for fake news detection. Computers, Materials & Continua. 2025. Vol. 86(1). P. 123-145. URL: https://doi.org/10.32604/cmc.2025.058678

22. Padalko, H., Chomko, V., Chumachenko, D.. A novel approach to fake news classification using LSTM-based deep learning models. Fronties in Big Data. 2024. Vol. 6. P. 1320800. URL: https://doi.org/10.3389/fdata.2023.1320800

23. Pierri F., Luceri L., Jindal N., Ferrara E. Propaganda and misinformation during the Russian invasion of Ukraine. 15th ACM Web Science Conference (WebSci). 2023. P. 1–10. URL: https://arxiv.org/abs/2208.06893

24. Rahman, S. S., Islam, M. A., Alam, M. M., Zeba, M., Rahman, M. A., Chowa, S. S., Raiaan, M. A. K., Azam, S. Hallucination to truth: A review of fact-checking. arXiv. 2024. URL: https://arxiv.org/abs/2508.03860

25. Rowen et al. Enhancing LLM factual accuracy with RAG to counter hallucinations: A case study on domain-specific queries in private knowledge-bases. arXiv. 2024. URL: https://arxiv.org/abs/2403.10446

26. Sahitaj P., Maab I., Yamagishi J., Kolanowski J., Möller S., Vera Schmitt. Towards automated fact-checking of real-world claims: Exploring large language models and retrieval-augmented generation. CEUR Workshop Proceedings. 2024. Vol. 3986. URL: https://ceur-ws.org/Vol-3986/paper2.pdf

27. Samet S., et al. Fake news detection with retrieval augmented generative artificial intelligence. Proceedings of FLLM 2024 (IEEE). 2024. URL: https://openreview.net/pdf?id=wRCEh8mO57

28. Sobchuk Y. fake_true_ukrainian_news_dataset.csv. Figshare. 2025. URL: https://doi.org/10.6084/m9.figshare.29257568.v1

29. Sobchuk Y., Krushelnytskyi S., Lipianina-Honcharenko K., Ivasechko A., Drakokhrust T. Hybrid method for building a balanced Ukrainian-language news corpus for fake news detection. CEUR Workshop Proceedings. 2025. Vol. 4159. URL: http://ceur-ws.org/Vol-4159/paper13.pdf

30. Srivastava N., Hinton G., Krizhevsky A. [et al.]. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014. Vol. 15, No. 1. P. 1929–1958.

31. Ullrich H., Mlynář T., Drchal J. Re-framing automated fact-checking as a simple RAG task. Proceedings of FEVER 2024. 2024. P. 145-152. URL: https://doi.org/10.18653/v1/2024.fever-1.16

32. Umer M., Imtiaz Z., Ullah S. [et al.]. Fake news stance detection using deep learning architecture. IEEE Access. 2020. Vol. 8. P. 156695–156695.

33. Wang G., Harwood K., Chillrud L., Ananthram A., Subbiah M., McKeown K. Use of retrieval-augmented large language model agent for long-form COVID-19 misinformation fact-checking. arXiv. 2025. URL: https://doi.org/10.48550/arXiv.2512.00007

34. Wang L., Ma C., Feng X. [et al.]. A survey on large language model based autonomous agents. Frontiers of Computer Science. 2024. Vol. 18. P. 1–26.

35. Zhang, J., Moon, K., Veeraraghavan, S. K. Does fake news create echo chambers? SSRN Electronic Journal. 2022. URL: https://doi.org/10.2139/ssrn.4144897

36. Zhao W. X., Zhou K., Li J. [et al.]. A survey of large language models. arXiv preprint arXiv:2303.18223. 2023. URL: https://arxiv.org/abs/2303.18223

37. Лендюк Д. Т., Ліп’яніна-Гончаренко Х. В. Ансамблеве навчання класифікаторів для онлайн виявлення дезінформації. Таврійський науковий вісник. Серія: Технічні науки. 2024. № 6. С. 46–63. DOI: 10.32782/tnv-tech.2024.6.6.

38. Шевченко Л. І., Дергач Д. В. Лінгвістичні технології фактчекінгу в інформаційних жанрах сучасних масмедіа. Актуальні проблеми української лінгвістики: теорія і практика. 2024. Вип. 49. С. 77–100. DOI: 10.17721/APULTP.2024.49.77-100.

39. Шуба А. В. Особливості фактчекінгової діяльності в українському медіапросторі. Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Журналістика. 2023. Т. 34 (73), № 3. С. 357–362. DOI: 10.32782/2710-4656/2023.3/58.

Додаток А

Апробація отриманих результатів

УДК 082:001
Н 34

Організація, від імені якої випущено видання:
ГО «Міжнародний центр наукових досліджень»
Номер запису організації в Єдиному реєстрі громадських об'єднань: 1499141.

Голова оргкомітету: Сотник С.Г.

Верстка: Білоус Т.В.
Дизайн: Бондаренко І.В.

Рекомендовано до видання Вченою Радою Інституту науково-технічної інтеграції та співпраці. Протокол № 17 від 07.05.2026 року.

*Конференцію зареєстровано Державною науковою установою у сфері управління Міністерства освіти і науки «Український інститут науково-технічної експертизи та інформації» в базі даних науково-технічних заходів України на поточний рік та бюлетені «План проведення наукових, науково-технічних заходів в Україні» (**Посвідчення № 164 від 26.01.2026**).*

*Збірник наукових праць з матеріалами конференції видано офіційно суб'єктом видавничої справи зі **Свідоцтвом ДК № 7860 від 22.06.2023**.*

Н 34 **Науковий простір: актуальні питання, досягнення та інновації**: збірник наукових праць з матеріалами XI Міжнародної наукової конференції, м. Дніпро, 8 травня, 2026 р. / Міжнародний центр наукових досліджень. — Вінниця: ТОВ «УКРЛОГОС Груп, 2026. — 548 с.

ISBN 978-617-8582-42-5
DOI 10.62731/mcnd-08.05.2026

Викладено матеріали учасників XI Міжнародної наукової конференції «Науковий простір: актуальні питання, досягнення та інновації», яка відбулася 8 травня 2026 року у місті Дніпро.

УДК 082:001

ISBN 978-617-8582-42-5

СЕКЦІЯ XVII.
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ

СЕКЦІЯ XVIII.
ТРАНСПОРТ ТА ТРАНСПОРТНІ ТЕХНОЛОГІЇ

ГІБРИДНА СИСТЕМА ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН НА ОСНОВІ LSTM ТА RAG

Собчук Юлія Євгеніївна
здобувач вищої освіти факультету комп'ютерно-інформаційних технологій
Західноукраїнський національний університет, Україна

Івасечко Андрій Вікторович
викладач кафедри інформаційно-обчислювальних систем і управління
Західноукраїнський національний університет, Україна

Мельник Назар Борисович
викладач кафедри інформаційно-обчислювальних систем і управління
Західноукраїнський національний університет, Україна

Науковий керівник: Ліп'яніна-Гончаренко Христина Володимирівна
д-р.техн.наук, доцент кафедри інформаційно-обчислювальних систем і управління
Західноукраїнський національний університет, Україна

Вступ

У сучасну епоху цифровізації та стрімкого розвитку інформаційних технологій проблема достовірності інформації набуває особливої актуальності. Основне споживання новин перемістилося у соціальні мережі та месенджери, де швидкість поширення контенту значно перевищує можливості його критичної оцінки [1, 2]. Це створює сприятливе середовище для розповсюдження фейкових новин, які часто використовуються як інструмент інформаційного впливу, зокрема в умовах гібридної війни [3].

Традиційні підходи до фактчекінгу, що базуються на ручному аналізі, є повільними та не здатні ефективно обробляти великі обсяги даних. Автоматизовані методи на основі нейронних мереж, зокрема LSTM, демонструють високу ефективність у виявленні лінгвістичних та стилістичних ознак маніпуляцій, проте не мають доступу до актуальних фактів і обмежені тренувальними даними [4, 5]. Водночас сучасні підходи, що використовують великі мовні моделі та архітектуру Retrieval-Augmented Generation (RAG), забезпечують можливість фактологічної перевірки шляхом звернення до зовнішніх джерел знань, але характеризуються високою обчислювальною складністю та затримками у відповіді [6].

Крім того, генеративні моделі схильні до так званих «галюцинацій» — створення правдоподібної, але недостовірної інформації без використання перевіреного контексту [7]. Таким чином, існує суперечність між швидкістю обробки та точністю фактчекінгу, що обумовлює необхідність розробки нових підходів до автоматизованого виявлення фейкових новин.

Мета роботи

Метою роботи є розробка гібридної системи автоматизованого виявлення фейкових новин, яка поєднує швидкодію моделі довгої короткочасної пам'яті (LSTM) та фактологічну точність архітектури Retrieval-Augmented Generation (RAG) з метою підвищення ефективності, точності та швидкості перевірки інформації.

Архітектура гібридної системи

Запропонована система реалізує дворівневу архітектуру обробки текстового контенту, яка включає модуль первинної класифікації та модуль фактологічної перевірки.

На першому етапі здійснюється швидка оцінка тексту за допомогою моделі LSTM. Вхідний текст подається у вигляді послідовності токенів, які перетворюються у векторні представлення (ембеддинги). Після цього послідовність обробляється рекурентною нейронною мережею, яка формує прихований стан, що акумулює контекст усієї послідовності.

Ймовірність того, що текст належить до класу фейкових новин, визначається за допомогою сигмоїдної функції активації:

$$P(y=1|X)=\sigma(W\bullet h_t+b)$$

де h_t — фінальний прихований стан LSTM, W та b— параметри моделі.

У разі, якщо отримана ймовірність перевищує заданий поріг, система переходить до другого етапу — фактологічної перевірки.

Другий етап реалізовано за допомогою архітектури RAG. Текст запиту перетворюється у векторне представлення за допомогою трансформерної моделі ембеддингів. Аналогічно представлено всі документи у векторній базі знань.

Для визначення релевантності використовується косинусна міра схожості:

$$sim(q,d_e)=\frac{q\bullet d_i}{\|q\|\bullet\|d_i\|}$$

Список використаних джерел:

1. Zhang, J., Moon, K., Veeraraghavan, S. K. (2022). Does fake news create echo chambers? SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4144897
2. Ahmed, S., Bee, A. W. T., Ng, S. W. T., Masood, M. (2024). Social media news use amplifies the illusory truth effects of viral deepfakes: A cross-national study of eight countries. Journal of Broadcasting & Electronic Media, 68(5), 778–805. https://doi.org/10.1080/08838151.2024.2410783
3. Bachmann, S.-D., Putter, D., Duczynski, G. (2023). Hybrid warfare and disinformation: A Ukraine war perspective. Global Policy, 14, 858–869. https://doi.org/10.1111/1758-5899.13257
4. Padalko, H., Chomko, V., Chumachenko, D. (2024). A novel approach to fake news classification using LSTM-based deep learning models. Frontiers in Big Data, 6, 1320800. https://doi.org/10.3389/fdata.2023.1320800
5. Huang, L. (2023). Deep learning for fake news detection: Theories and models. In Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE '22) (pp. 1322–1326). Association for Computing Machinery. https://doi.org/10.1145/3573428.3573663
6. Li, G., Lu, W., Zhang, W., Lian, D., Lu, K., Mao, R., Shu, K., Liao, H. (2024). Re-search for the truth: Multi-round retrieval-augmented large language models are strong fake news detectors (arXiv:2403.09747). arXiv. https://doi.org/10.48550/arXiv.2403.09747
7. Rahman, S. S., Islam, M. A., Alam, M. M., Zeba, M., Rahman, M. A., Chowa, S. S., Raiaan, M. A. K., Azam, S. (2025). Hallucination to truth: A review of fact-checking and factuality evaluation in large language models (arXiv:2508.03860). arXiv. https://doi.org/10.48550/arXiv.2508.03860

CEUR Workshop Proceedings
ceur-ws.org
ISSN 1613-0073

Vol-4159
urn:nbn:de:0074-4159-x

BAITmp 2025
The 2nd International Workshop on Bioinformatics and Applied Information Technologies for medical purpose 2025

Proceedings of the 2nd International Workshop on Bioinformatics and Applied Information Technologies for medical purpose (BAITmp 2025)

Ben Guerir, Morocco, November 12-13, 2025.

Edited by

Abdellah Menou *
Mykhaylo Petryk **

* Mohammed VI Polytechnic University, ASTI LAB, SAP+D School, Ben Guerir, Morocco
** Ternopil Ivan Puluj National Technical University, Faculty of Computer Information Systems and Software Engineering, Ternopil, Ukraine

Session 2

2025-12-16: submitted by Abdellah Menou, metadata incl. bibliographic data published under Creative Commons CC0
2026-02-07: published on CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073) |valid HTML5|

Hybrid method for building a balanced Ukrainian-language news corpus for fake news detection*

Yuliia Sobchuk[1,†], Sviatoslav Krushelnytskyi[1,†], Khrystyna Lipianina-Honcharenko[1,†], Andrii Ivasechko[1,†] and Tetiana Drakokhrust[1,†]

[1] *Faculty of Computer Information Technologies, West Ukrainian National University, 46000 Ternopil, Ukraine*

Abstract

This paper presents a method for constructing a balanced Ukrainian-language news corpus for fake-news detection that combines LLM-based controlled generation with editorial verification. The truthful subset is collected from authoritative media using topic- and year-stratified sampling (2022–2025), while fake examples are produced via a parameterized LLM prompt controlling tone, style, manipulation types, and topics. The pipeline comprises multi-stage normalization, stop-word removal, lemmatization (Stanza), language identification, near-duplicate filtering (hybrid cosine/Jaccard-trigram similarity), and human moderation of borderline cases. The resulting corpus contains ~40k texts (~20k "Trusted" and ~20k "Fake") with an average length of ~250 tokens and a bimodal length distribution. Reproducibility is ensured by publishing data schemas and fixed 80/20 splits. A BiLSTM baseline with FastText (300d) achieves 99.25% accuracy, 0.9925 macro-F1, and 0.985 MCC, with false-positive/false-negative rates ≤0.9%. These results indicate strong class separability and validate the corpus as a benchmark for future studies, including transformer-based models, ablation of synthetic components, robustness assessment, and probability calibration.

Keywords

fake news detection, Ukrainian corpus, large language models, LLM-based generation, editorial verification, text preprocessing, BiLSTM, FastText, reproducibility. [1]

1. Introduction

The intensive dynamics of the Ukrainian media space in 2022–2025 have been accompanied by a significant increase in the volume of disinformation, which complicates the verification of news content and necessitates reliable data for training fake news detectors. Despite substantial progress in global research, Ukrainian-language annotated corpora remain limited in both scale and thematic-stylistic coverage, while the transfer of models trained on English-language data often leads to a loss of accuracy due to linguistic and domain shifts. The absence of reproducible corpus construction pipelines and transparent quality control procedures further complicates the comparability of results.

The aim of this study is to develop a method for creating a balanced corpus of Ukrainian-language news for the task of fake news detection by combining controlled generation using large language models with editorial verification. The proposed approach integrates stratified selection of trustworthy materials from reputable sources with synthetic generation of fake examples under controlled parameters of tone, style, type of manipulation, and topic. The methodology is designed to ensure diversity, minimize bias, and comply with ethical requirements for the use of generative AI.

The pipeline involves multi-stage normalization and linguistic processing of texts, language verification, removal of duplicates and stylistic anomalies, as well as manual moderation of

*BAITmp'2025: The 2nd International Workshop on "Bioinformatics and Applied Information Technologies for medical purpose", November 12-13, 2025, Ben Guerir, Morocco

[1,*] Corresponding author.

[†] These authors contributed equally.

yuliia.sob4uk@gmail.com (Y. Sobchuk); sviatoslav.kru@gmail.com (S. Krushelnytskyi); kh.lipianina@wunu.edu.ua (K. Lipianina-Honcharenko); andrewivasechko@gmail.com (A. Ivasechko); t.drakokhrust@wunu.edu.ua (T. Drakokhrust)

0009-0005-9542-0028 (Y. Sobchuk); 0009-0001-1257-9051 (S. Krushelnytskyi); 0000-0002-2441-6292 (K. Lipianina-Honcharenko); 0009-0002-8623-7828 (A. Ivasechko); 0000-0002-4761-7943 (T. Drakokhrust)

borderline cases. To ensure reproducibility, stochastic parameters are fixed, data schemas are published, and train–validation split lists are provided. As a result, a balanced corpus of approximately 40,000 news texts was obtained (around 20,000 in each of the "Trusted" and "Fake" classes), with representative thematic and stylistic coverage.

This paper presents a method for constructing a balanced corpus of Ukrainian-language news for fake news detection that combines controlled LLM generation with editorial verification. Section 2 summarizes related work; Section 3 describes the methodology and data collection pipeline, including stratified selection of reliable materials, parameterized generation of synthetic examples, preprocessing, and quality control; Section 4 provides the corpus composition and statistics, as well as baseline benchmarks (including BiLSTM+FastText); Section 5 presents experimental results with a confusion matrix and analysis of training dynamics; Section 6 formulates conclusions, highlights scientific significance, and outlines directions for future work, including testing transformer models, conducting ablation studies, assessing robustness, and addressing ethical considerations in the use of generative AI.

2. Related work

Early attempts at fake news detection were primarily based on classical machine learning algorithms with manual feature engineering. In particular, the study presented in [1] highlights the use of methods such as Support Vector Machines (SVM), Naive Bayes classifiers (NB), and Random Forests (RF). The effectiveness of these algorithms largely depended on the quality of the selected linguistic and meta-feature characteristics.

The research conducted in [2] focuses on classifying fake news in social media based on textual content. This work applied four traditional text feature extraction methods (TF-IDF, Count Vector, Character-level Vector, N-Gram Level Vector) and ten different machine learning and deep learning classifiers. The obtained results demonstrated that textual fake news can be effectively classified, with classification accuracy ranging from 81% to 100% depending on the classifier used, with convolutional neural networks (CNN) showing particularly high effectiveness.

With the rise of deep learning, fake news detection methods have gained new momentum. Recurrent neural networks, particularly bidirectional LSTMs (Bi-LSTMs), enable models to learn context in both forward and backward directions.

In [3], a combination of BERT and LSTM was used for fake news classification based on headlines. The model was trained on the FakeNewsNet dataset (PolitiFact, GossipCop), achieving accuracy improvements of 2.5% and 1.1%, respectively, compared to the baseline BERT model, confirming the effectiveness of combining transformers and recurrent networks.

The study in [4] proposed a fake news detection framework that combines analysis of news content and social context. The model is built on a Transformer architecture with an encoder for feature extraction and a decoder for predicting the subsequent behavior of the news. To address the lack of labeled data, the authors applied a custom automatic labeling technique. Experiments with real-world data showed that the model provides higher accuracy in early detection (within minutes of dissemination) compared to baseline methods.

In [5], the authors explored the potential of fine-tuning the modern language model GPT-3 for the task of fake news detection. The model was adapted on the ISOT dataset and demonstrated high effectiveness, achieving an accuracy of 99.90%, precision of 99.81%, recall of 99.99%, and an F1-score of 99.90%, significantly outperforming existing solutions. These results confirm the promise of using GPT-3 to combat disinformation in social media and news outlets.

The study in [6] presents the Generative Bidirectional Encoder Representations from Transformers (GBERT) framework, which combines BERT's deep contextual understanding with GPT's generative capabilities for fake news classification. Both models were fine-tuned on two real-world benchmark datasets, achieving an accuracy of 95.30%, precision of 95.13%, recall of 97.35%, and an F1-score of 96.23%. The obtained results demonstrate the high efficiency of GBERT

and the potential of this approach in countering the spread of disinformation in the digital environment.

In [7], a review of machine learning algorithms and datasets used for fake news detection was conducted. Among the most effective models identified were the Stacking Method with 99.9% accuracy, BiRNN, and CNN — both at 99.8%. Most studies relied on data from controlled environments (e.g., Kaggle) or from sources without real-time updates, which limits their practical applicability in social media, where disinformation spreads most actively. The most frequently used datasets included Kaggle, Weibo, FNC-1, COVID-19 Fake News, and Twitter. The authors emphasize the need to expand topics beyond political news and to apply hybrid methods in future research.

The study in [8] presents the OLTW-TEC method (Online Learning with Sliding Windows for Text Classifier Ensembles), developed for detecting disinformation in the Ukrainian-language information space. The approach combines an ensemble of classifiers with a "sliding window" mechanism for dynamically updating the model to incorporate new data, thereby increasing its adaptability to changing fake news dissemination tactics. The method was tested on a specially constructed dataset of authentic and fake news, achieving an accuracy of 93%. The results confirm the effectiveness of OLTW-TEC and its suitability for operating under information warfare conditions, as well as its potential for adaptation to other languages and regions.

A comparative study [9] showed that RNN, LSTM, and Bi-LSTM models achieve similar results with around 91% accuracy, although LSTM outperformed in recall and RNN in precision. This highlights the importance of selecting an architecture that aligns with specific objectives.

Further research has focused on transformer architectures. For instance, the authors of [10] evaluated the performance of BERT, CNN, Bi-LSTM, and their ensemble combination. The latter achieved the highest accuracy — 98.24% — demonstrating the effectiveness of hybrid solutions.

In [11], transformers (BERT, RoBERTa, GPT-2) were compared with graph neural networks (GNN). Transformers showed significantly better results: RoBERTa reached 99.99% on ISOT, and GPT-2 achieved 99.72% on WELFake, highlighting their ability to work with contextually rich data.

The study in [12] analyzed the effectiveness of various machine learning methods for detecting disinformation in Ukrainian-language news collected during the military conflict. Evaluated models included logistic regression, SVM, random forest, gradient boosting, KNN, decision trees, XGBoost, and AdaBoost. The random forest demonstrated the best results. The authors emphasize the importance of adapting models to the specifics of the task and the need for further research in this area.

The authors of [13] argue that combining transformers with text summarization further increases accuracy. RoBERTa fine-tuned on summarized content achieved 98.39%, which is among the highest metrics among modern models.

Special attention should be paid to hybrid models that integrate Word2Vec vectors with CNN and LSTM. In [14], the authors focused on fake news classification using a combination of machine learning (ML) and natural language processing (NLP) methods based on textual content. They compared several modern ML models and neural networks. Experiments showed that all traditional ML models achieved over 85% accuracy, while neural networks outperformed them, reaching over 90% accuracy.

Research (Table 1) on fake news detection has evolved from classical machine learning approaches with manual feature engineering to deep and transformer-based architectures (see [1–14]). In the early stages, SVM, NB, and RF were applied with textual representations such as TF-IDF or Bag-of-Words, where accuracy largely depended on feature selection and data domain. Subsequent studies focused on neural models: RNN/LSTM/Bi-LSTM achieved results around ≈91% [9], while hybrids combining CNN with vector representations (e.g., FastText) improved performance to 0.99 in Accuracy and 0.97–0.99 in F1-score [13–14]. Transformers, particularly BERT/RoBERTa and their ensembles, reached 98.24% and higher [10], with some well-known datasets (ISOT, WELFake) reporting values up to 99.99% [5]. However, these metrics vary

significantly across datasets and experimental setups, complicating the generalization of conclusions and accurate comparison of approaches.

Table 1
Comparison of existing approaches

№	Author(s), year	Type of approach	Technologies, model	Accuracy\results
1	Alluri et al., 2023 [1]	Review	Systematic analisis of methods	—
2	Abdulrahman, 2020 [2]	Classical ML/DL	SVM, NB, RF, CNN	81% – 100%
3	Dhiman et al., 2024 [6]	Transformers	GBERT	95,30%
4	Airlangga, 2024 [9]	Deep learning	RNN, LSTM, Bi-LSTM	~91%
5	Kuntur et al., 2024 [10]	Transformers/GNN	BERT, RoBERTa, GPT-2, GNN	99.99%
6	Saadi et al., 2025 [11]	Transformers + Summary	RoBERTa with text summarization	98.39%
7	Hashmi et al., 2024 [13]	Hybrid (DL + NLP)	FastText + CNN + LSTM	Acc: 0.99, F1: 0.97–0.99
8	Lai et al., 2022 [14]	Neural networks	CNN, LSTM	~90%

Existing studies indicate progress in methods but reveal several research gaps, particularly for the Ukrainian-language segment: (i) a lack of large public annotated corpora with transparent preparation pipelines, fixed splits, and detailed documentation; (ii) limited evaluation under temporal and domain shift scenarios, with insufficient attention to robustness against paraphrasing/adversarial attacks and probability calibration; (iii) inadequate analysis of bias across topics/genres, as well as the impact of synthetic examples on model generalization and stability; (iv) incomplete reproducibility due to the absence of publicly available code, fixed seeds, and detailed preprocessing protocols. The scientific significance of this research lies in addressing these gaps by creating a reproducible Ukrainian-language corpus with balanced classes, a clearly specified construction and quality control methodology, and establishing benchmark standards that allow accurate comparison of modern architectures and investigation of their robustness under realistic conditions.

3. Method

The following outlines the sequential stages of constructing a balanced corpus of Ukrainian-language news for fake news detection, combining verified texts from reputable media with controlled LLM-generated examples (stages 1–6, Fig. 1).

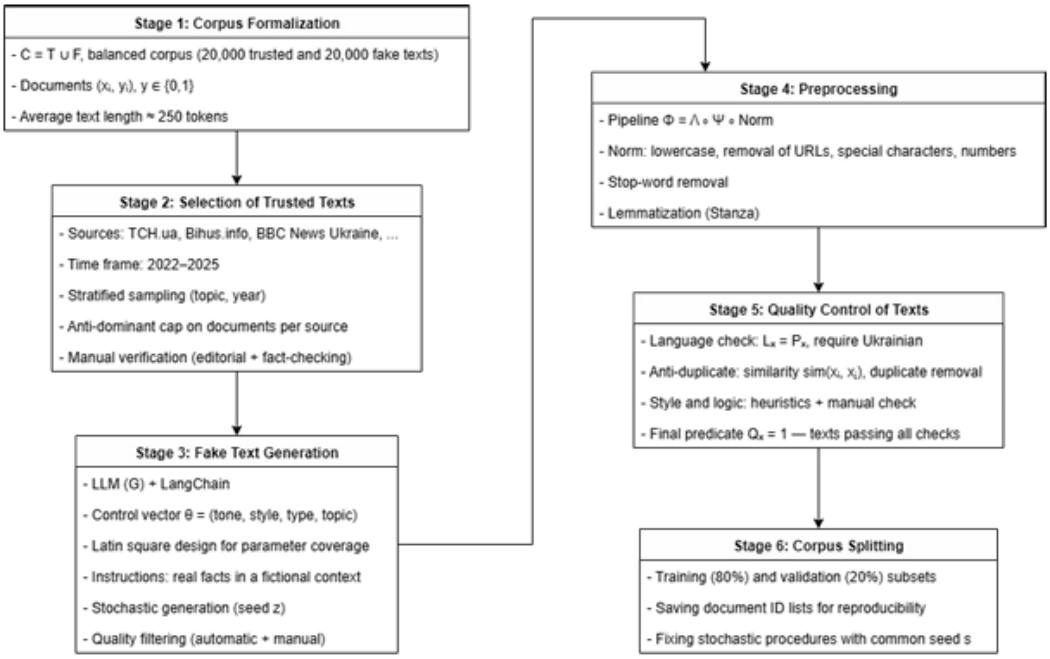

Figure 1: Stages of constructing the Ukrainian-language news corpus.

Stage 1. Corpus formalization.

Let $C = T \cup F$ be the final corpus of Ukrainian-language news for the two-class fake news classification task, where F is the set of trusted texts and is the set of fake texts. Each document is a pair (x_i, y_i), where $x_i \in \Sigma*¿$ is the text, and $y_i \in [0,1]$ is the class label ($y = 1 - «Fake»$, $y = 0 - «Trusted»$). The corpus is balances: $/T/ = /F/ = 20\,000$, so the prior class probabilities are $\pi_0 = \pi_1 = \frac{1}{2}$.

Let the empirical distribution of text lengths in tokens be denoted as $\hat{p}L(l)$. Following preprocessing, the average length is $\hat{\mu}_L = E_{\hat{p}_L}[L] \approx 250$ tokens.

Stage 2. Selection of trusted texts.

Let $S = [TCH.ua, Bihus.info, BBC\ News\ Україна, \ldots]$ be the set of sources. Over the time interval t∈[2022,2025], the initial sample is formed as

$$U_T = [x : x\ was\ published\ on\ s \in S, t \in [2022,2025]].$$

To avoid dominance of individual sites or topics, stratified sampling is applied by topic $\tau \in [politics, economy, society, defense, health, \ldots]$ and publication year. Within each stratum $,(\tau, year]$ random sampling without replacement is performed with an upper limit m_s of documents per source s (anti-dominance cap). Each candidate undergoes manual verification of editorial standards and fact-checking; acceptance is denoted by the predicate $RT(x) \in [0,1]$. The final set is:

$$T = [x \in U_T : R_T(x) = 1]$$

Stage 3. Fake text generation.

Fake texts are generated by a large language model G via LangChain using a parameterized prompt template $\tau(\theta)$. The control vector is

$$\theta = (tone, style, type, topic)$$

where

$-tone \in [neutral, alarming, reassuring]$;

$-style \in [analytical, populist, ironic, factual]$;

$-type \in [disinformation, manipulation, emotional influence, propaganda]$;

$-topic \in [politics, economy, education, defense, infrastructure]$.

To ensure diversity, θ is covered almost uniformly (Latin square / combinatorial sweep), and G is instructed to use real facts/persons in a fictional context while avoiding fantastical events or clichés. Generation occurs as

$$x = G(\tau(\theta), z),$$

where z is the stochastic seed/model temperature. Each synthetic text undergoes both automatic and manual quality filtering.

Stage 4. Preprocessing.

Let

$$\Phi = \Lambda \circ \Psi \circ Norm$$

be the preprocessing pipeline,

where

$Norm(x)$: lowercase conversion, removal of URLs, hashtags, special characters, and numeric markers;

$\Psi(x)$: stop-word removal;

$\Lambda(x)$: lemmatization (Stanza for Ukrainian).

The resulting text $x' = \Phi(x)$ is fed into the validation and statistical analysis modules.

Stage 5. Quality control of texts.

Three independent acceptance predicates are applied:

1. Language identification.
$$L(x) = P(x), \text{ requiring } L(x) \geq \tau_{lang}.$$
2. Anti-duplicate check. Let
$$¿(x_i, x_j) = \alpha \cdot \cos\cos(tfidf(x_i), tfidf(x_j)) + (1-\alpha) \cdot J_3(x_i, x_j),$$
where J_3 is the Jaccard similarity of trigrams. A text x is discarded if $max_{j<i} \sim(x, x_j) \geq \tau_{dup}$ (practicaly implemented via MinHash/LSH).
3. Style/logic check. Automatic heuristics (length, anomalous n-gram repetition) plus manual review $R(x)$.

The final filter is

$$Q(x) = 1[L(x) \geq \tau_{lang}, \sim(x, x_j) < \tau_{dup}] \wedge R(x),$$

Where τ_{lang} and τ_{dup} denote threshold parameters for language and duplication filtering, respectively.

The final sets T, F consist only of texts that satisfy $Q(x) = 1$.

Stage 6. Corpus splitting.

The corpus is split into training and validation subsets while preserving class balance:

$$C^{train} \cup C^{val} = C, / C^{train} / \approx 0.8 / C / , / C^{val} / \approx 0.2 / C / .$$

Lists of document IDs for each split are stored separately to ensure reproducibility, and all stochastic procedures are fixed using a common seed s.

4. Result

To construct the trusted news class, we used materials from reputable Ukrainian information sources, including ТСН.ua, Bihus.info, BBC News Україна, and others. The full list of trusted sources is available online [15]. The news covers the period 2022–2025 and topics related to politics, economy, society, military events, healthcare, etc. In total, approximately 20,000 texts were selected, each undergoing manual verification for accuracy and compliance with contemporary

journalistic style. Manual verification was conducted by two authors, with the dataset evenly split between them for independent assessment.

Fake news was generated using the large language model Gemini 2.0. Generation was performed via the LangChain interface using a pre-designed prompt template that specified the parameters of the resulting text.

Each fake news item was created taking into account the following characteristics::

- Tone: neutral, alarming, reassuring;
- Writing style: analytical, populist, ironic, factual;
- Type of fake: disinformation, manipulation, emotional influence, propaganda;
- Topic: politics, economy, education, defense, infrastructure.

The prompt template also instructed the model to incorporate real facts, institutions, and persons within a fictional context, making the texts as close as possible to authentic media content. Generation was accompanied by guidelines to avoid fantastical events or obvious clichés.

Despite automated generation, all fake news items underwent manual verification. Texts with low plausibility, artificial language, logical inconsistencies, or violations of style guidelines were filtered out. As a result, a balanced corpus was created, consisting of 20,000 fake and 20,000 trusted news articles.

The news corpus [16] was cleaned of noisy elements: hyperlinks, special characters, numeric markers, and hashtags were removed. Texts were converted to lowercase, stripped of stop-words, and lemmatized using the Stanza library, which supports Ukrainian morphology.

The combined corpus contains approximately 40,000 Ukrainian-language news articles with nearly equal class representation (Fig. 2): Trusted ≈ 20,000, Fake ≈ 20–21,000; the deviation from a 50/50 split does not exceed ≈10%. Such balance reduces the risk of metric bias toward the larger class.

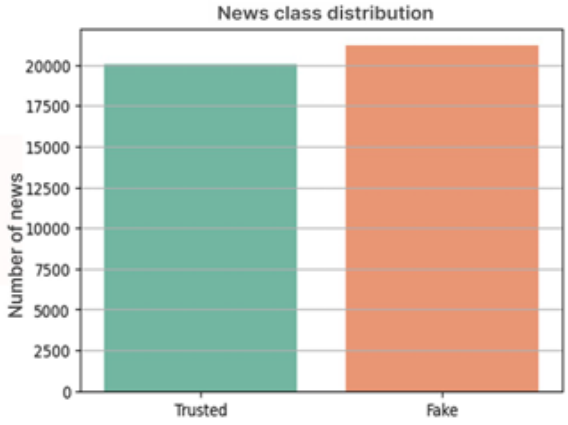

Figure 2: Histogram of class distribution.

The length distribution exhibits a pronounced bimodality (Fig. 3): the first local peak corresponds to short notes (~20–60 words), while the second corresponds to full-length articles of ≈250–300 words. The mean length is approximately 250 tokens/words, which matches a typical news item and provides sufficient context for linguistic features.

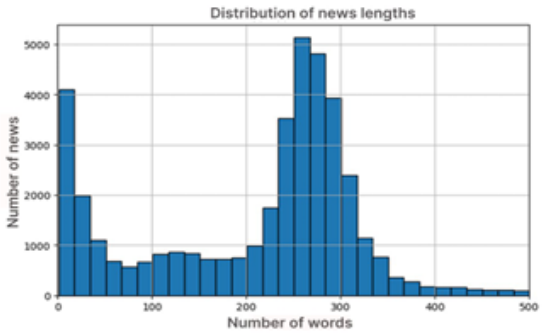

Figure 3: Histogram of news length distribution.

The top-20 lemmas by class (Fig. 4) show the expected dominance of function words (conjunctions, prepositions), indicating a homogeneous underlying syntactic structure across both classes. At the same time, differences in content lemmas are noticeable: in the Fake class, terms like "український" (Ukrainian), "ситуація" (situation), "про" (about), "але" (but) appear more frequently, whereas in Trusted, "Україна" (Ukraine), "рік" (year), "вони" (they), "для" (for) are more common. This reflects stylistic distinctions: fake texts tend to use generalizing and evaluative formulations, while trusted texts feature nominative references to institutions/country and temporal markers.

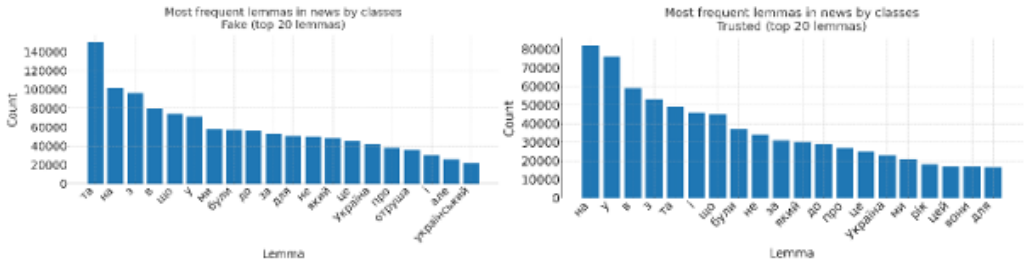

Figure 4: Frequency plot of the most common lemmas.

The cosine similarity between the sets of key lemmas was 0.879, indicating a high lexical overlap. This suggests that fake and trusted news often share the same topical vocabulary, which makes the classification task realistic and shifts the discriminative power towards stylistic and contextual features rather than mere word occurrence. The heatmap of cosine similarities (Fig. 5) further illustrates this overlap, showing the strong lexical proximity between the two classes.

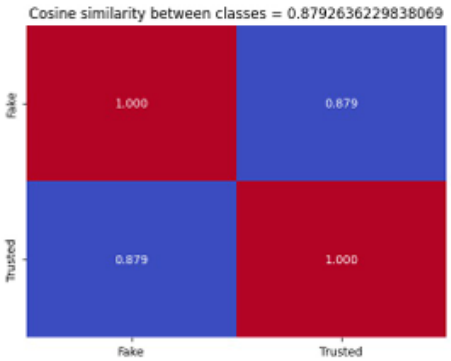

Figure 5: Heatmap of cosine similarity between Fake and Trusted classes.

The bimodality in text lengths reflects two dominant forms of news presentation (short "notes" and full-length articles), which is useful for building robust models: the classifier is exposed to different styles and text volumes. High model metrics on the balanced corpus confirm strong class separability and the quality of data preparation (cleaning, language and duplicate control). Differences in content lemmas illustrate stylistic signals that can be used as interpretable features or for further bias analysis.

For vector representation, FastText in skip-gram mode was applied. The vectorizer was trained on the preprocessed corpus with the following hyperparameters: vector size – 300, number of epochs – 15, context window width – 5, minimum word frequency – 10.

News vectorization was performed by truncating or padding with zero vectors to a fixed length of 100 tokens. The classifier architecture is based on a bidirectional LSTM network with additional Dropout and Dense layers. Optimization was performed using Adam with an initial learning rate of 0.001.

After training on 80% of the dataset and validating on the remaining 20%, the model achieved an accuracy of 99.25% (Table 2). Precision and recall coefficients exceed 0.99 for both classes, which is also confirmed by the confusion matrix (Fig. 5). The training dynamics are shown in Figure 6, illustrating a gradual decrease in the loss function without signs of overfitting.

Table 2
Frequency of Special Characters

Class	Precision	Recall	F1-score	Support
Fake	0.9915	0.9939	0.9927	4234
Trusted	0.9935	0.9910	0.9923	4015
Average	0.9925	0.9924	0.9925	4124

The column "Support" indicates the number of instances belonging to each class in the test dataset. Overall classification performance (see Fig. 5): Accuracy = 0.9925 (99.25%), macro-averaged F1 = 0.9925, Matthews correlation coefficient (MCC) = 0.985 (see Table 2). From the confusion matrix (Fig. 6):

- Fake class: TP = 4208, FN = 26, FP = 36, TN = 3979; Precision = 0.9915, Recall = 0.9939, F1 = 0.9927; TPR = 0.9939, TNR = 0.9910, FNR = 0.0061, FPR = 0.0090;
- Trusted class: TP = 3979, FN = 36, FP = 26, TN = 4208; Precision = 0.9935, Recall = 0.9910, F1 = 0.9923; TPR = 0.9910, TNR = 0.9939, FNR = 0.0090, FPR = 0.0061.

The average balanced accuracy equals $\frac{TP\,R_{Fake}+TP\,R_{Trusted}}{2}=0.99245$, corresponding to a BER = 0.00755. The low false positive and false negative rates (≤ 0.9%) in each class confirm strong class separability and the absence of bias toward any label.

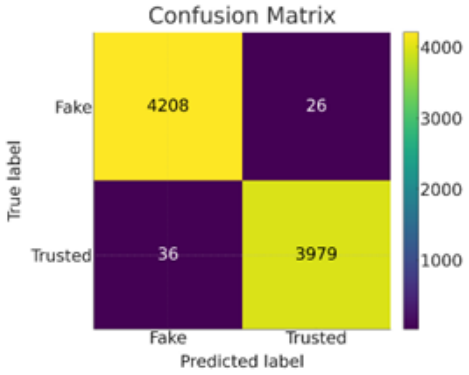

Figure 6: Confusion Matrix.

The training dynamics (Fig. 7) show a monotonic increase in accuracy on both the training and validation sets, reaching ≈0.99 and plateauing after approximately the 7th epoch. The loss function decreases steadily across both subsets without divergence. The absence of rising validation error and the minimal generalization gap indicate no signs of overfitting under the chosen hyperparameters (FastText 300d, window = 5, epochs = 15, min_count = 10; BiLSTM + Dropout + Dense, Adam optimizer, $\eta = 10^{-3}$).

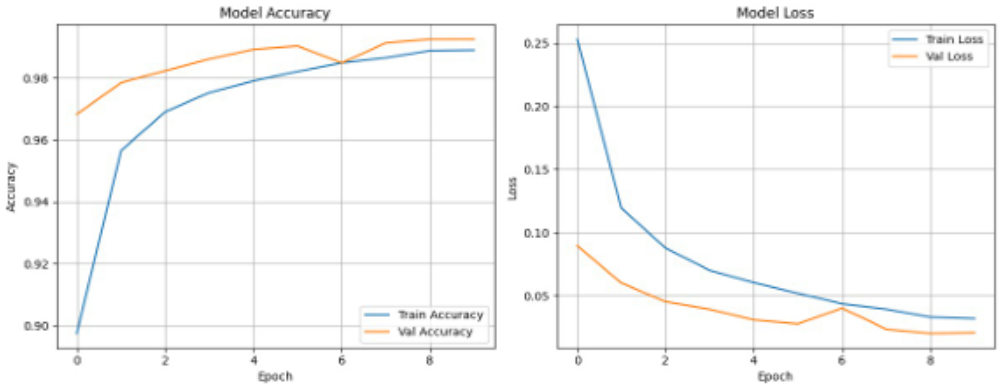

Figure 7: Accuracy and loss dynamics on training and validation.

5. Conclusion

This study introduces a balanced corpus of Ukrainian-language news for fake news detection, comprising ~40,000 texts (≈20k "Trusted" and ≈20k "Fake") from 2022–2025. Trusted data were sourced from verified media, while synthetic fakes were generated with LLMs under controlled prompts, followed by normalization, filtering, and lemmatization. The corpus shows clear stylistic differences between classes and an average length of ≈250 tokens, making it suitable for machine learning.

Evaluation with a BiLSTM + FastText model achieved accuracy of 99.25% and macro-F1 of 0.9925, confirming both the quality of the dataset and the feasibility of automated fake news detection. Misclassification rates remained below 1%, with stable learning dynamics and no overfitting.

The dataset and approach can be applied in practice for media monitoring and early detection of disinformation in Ukraine. Future work will include benchmarking transformer models, robustness testing, and releasing artifacts to support reproducible research and regular updates of the corpus.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Alluri, C. R., Reddy, K. S., Sarma, B. M., & Kumar, D. S. (2023). Fake news detection: a systematic review and knowledge mapping. The Journal of Supercomputing, 79(2), 1735–1770.

[2] A. Abdulrahman and M. Baykara, "Fake News Detection Using Machine Learning and Deep Learning Algorithms," 2020 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, 2020, pp. 18-23, doi: 10.1109/ICOASE51841.2020.9436605.

[3] Rai, N., Kumar, D., Kaushik, N., Raj, C., & Ali, A. (2022). Fake News Classification using transformer based enhanced LSTM and BERT. *International Journal of Cognitive Computing in Engineering*, *3*, 98-105.

[4] Raza, S., & Ding, C. (2022). Fake news detection based on news content and social contexts: a transformer-based approach. International journal of data science and analytics, 13(4), 335–362. https://doi.org/10.1007/s41060-021-00302-z.

[5] Hemina, K., Boumahdi, F., Madani, A., Remmide, M.A. (2023). A Cross-Validated Fine-Tuned GPT-3 as a Novel Approach to Fake News Detection. In: Zantout, H., Ragab Hassen, H. (eds) Proceedings of the International Conference on Applied Cybersecurity (ACS) 2023. ACS 2023. Lecture Notes in Networks and Systems, vol 760. Springer, Cham. .

[6] Dhiman, P., Kaur, A., Gupta, D., Juneja, S., Nauman, A., & Muhammad, G. (2024). GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection. Heliyon, 10(16).

[7] Villela, H. F., Corrêa, F., Ribeiro, J. S. D. A. N., Rabelo, A., & Carvalho, D. B. F. (2023). Fake news detection: a systematic literature review of machine learning algorithms and datasets. Journal on Interactive Systems, 14(1), 47-58.

[8] Lipianina-Honcharenko, K., Soia, M., Yurkiv, K., & Ivasechko, A. (2024, May). Evaluation of the effectiveness of machine learning methods for detecting disinformation in Ukrainian text data. In Proceedings of the Seventh International Workshop on Computer Modeling and Intelligent Systems (CMIS-2024), Zaporizhzhia, Ukraine.

[9] Airlangga, G. (2024). Advancing fake news detection: a comparative study of RNN, LSTM, and Bidirectional LSTM Architectures. Jurnal Teknik Informatika C.I.T Medicom, 16(1), 13–23.

[10] Kuntur, S., Krzywda, M., Wróblewska, A., Paprzycki, M., & Ganzha, M. (2024). Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection: A Verification and Reimplementation Study. Electronics, 13(23), 4784.

[11] Saadi, A., Belhadef, H., Guessas, A., and Hafirassou, O. 2025. Enhancing Fake News Detection with Transformer Models and Summarization. Engineering, Technology & Applied Science Research. 15, 3 (Jun. 2025), 23253–23259. DOI:https://doi.org/10.48084/etasr.10678.

[12] Lipianina-Honcharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. (2024). OLTW-TEC: online learning with sliding windows for text classifier ensembles. Frontiers in Artificial Intelligence, 7, 1401126.

[13] Hashmi, E., Yayilgan, S. Y., Yamin, M. M., Ali, S., & Abomhara, M. (2024). Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai. IEEE Access.

[14] Lai, C.-M., Chen, M.-H., Kristiani, E., Verma, V. K., & Yang, C.-T. (2022). Fake News Classification Based on Content Level Features. Applied Sciences, 12(3), 1116. https://doi.org/10.3390/app12031116.

[15] Dataset. (2025). Trusted sources for fake news detection. Google Drive. https://drive.google.com/drive/folders/13c_QRvuMuXTByYZkzbJOq4VcXzURniVx?usp=drive_link.

[16] Sobchuk, Yulia (2025). fake_true_ukrainian_news_dataset.csv. figshare. Dataset. https://doi.org/10.6084/m9.figshare.29257568.v1.