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

КОЗАК Аліна Андріївна

Метод класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі / Method for subjective question classification based on a multilayer neural network

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконала студентка групи КНзм-21
А. А. Козак
_____

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

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

ТЕРНОПІЛЬ – 2025

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

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

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
Козак Аліна Андріївна
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи
Метод класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі / Method for subjective question classification based on a multilayer neural network
керівник роботи к.т.н., доцент Х.В. Ліп'яніна- Гончаренко
затверджені наказом по університету від 20 грудня 2024 року № 938.

2. Строк подання студентом закінченої кваліфікаційної роботи 1 грудня 2025 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити
– огляд предметної області та аналіз суб'єктивних і інсинсирних запитань у системах питань-відповідей;
– аналіз класичних алгоритмів класифікації текстових запитань;
– огляд сучасних нейронних мереж і глибинних моделей для класифікації суб'єктивних запитань;
– постановка задачі класифікації суб'єктивних та інсинсирних запитань;
– розробка базової моделі на основі TF-IDF та XGBoost;
– розробка багаторівневої нейронної моделі з SimpleRNN та двонапрямленою GRU;
– реалізація методу та проведення експериментів;
– аналіз і порівняння результатів із базовими підходами.

5. Перелік графічного матеріалу у роботі
– схема архітектури багаторівневої нейронної мережі для класифікації суб'єктивних та інсинсирних запитань;
– графіки порівняльного аналізу ефективності базової та нейромережевих моделей класифікації.

6. Консультанти розділів кваліфікаційної роботи

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

7. Дата видачі завдання 20 грудня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Затвердження теми кваліфікаційної роботи, ознайомлення з літературними джерелами та складання плану роботи.	до 01.01. 2025 р.	
2	Написання 1 розділу кваліфікаційної роботи	до 01.03. 2025 р.	
3	Написання 2 розділу кваліфікаційної роботи	до 20.05.2025 р.	
4	Написання 3 розділу кваліфікаційної роботи	до 28.10. 2025 р.	
5	Представлення попереднього варіанту кваліфікаційної роботи, перевірка та внесення змін керівником	до 11.11.2025 р.	
6	Опрацювання зауважень та представлення завершеного варіанту кваліфікаційної роботи. Підготовка супроводжуючих документів.	до 25.11.2025 р.	
7	Перевірка кваліфікаційної роботи на оригінальність тексту.	до 1.12.2025 р.	
8	Оформлення кваліфікаційної роботи та отримання допуску до захисту	до 04.12.2025 р.	
9	Подання кваліфікаційної роботи до захисту на засіданні атестаційної комісії.	до 14.12. 2025 р.	

Студент _____ А.А. Козак
підпис

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

РЕЗЮМЕ

Кваліфікаційна робота на тему «Метод класифікації суб’єктивних запитань на основі багаторівневої нейронної мережі» на здобуття освітнього ступеня «Магістр» зі спеціальності 122 «Комп’ютерні науки» освітньої програми «Комп’ютерні науки» виконана на 55 сторінках і містить 37 ілюстрацій, додатки та 32 використані джерела.

Мета роботи — розроблення та дослідження методу автоматичної класифікації суб’єктивних та інсинсирних запитань у системах питань-відповідей на основі багаторівневої нейронної мережі для підвищення точності розпізнавання складних риторичних та емоційних формулювань.

Результати дослідження: розроблено багаторівневу нейронну мережу, що поєднує базове TF-IDF-представлення та рекурентні шари SimpleRNN і BiGRU для класифікації суб’єктивних і інсинсирних запитань. Запропонований метод забезпечує підвищення точності класифікації порівняно з базовими підходами та дозволяє ефективно обробляти великі масиви текстових даних із урахуванням класового дисбалансу. Проведено експериментальні дослідження, порівняльний аналіз та візуалізацію результатів, що підтверджують ефективність багаторівневої нейронної архітектури.

Практичне значення роботи: результати можуть бути використані у системах автоматичної модерації контенту, освітніх платформах, чат-ботах та онлайн-сервісах питань-відповідей для виявлення провокативних чи потенційно токсичних запитань. Метод дозволяє підтримати модераторів, покращити якість взаємодії користувачів та адаптувати стратегії відповіді системи.

Ключові слова: СУБ’ЄКТИВНІ ЗАПИТАННЯ, ІНСИНСИРНІ ЗАПИТАННЯ, ОБРОБКА ПРИРОДНОЇ МОВИ, РЕКУРЕНТНА НЕЙРОМЕРЕЖА, BigrU, TF-IDF, КЛАСИФІКАЦІЯ ТЕКСТУ, NLP.

ABSTRACT

The qualification thesis titled “Method for Subjective Question Classification Based on a Multilayer Neural Network”, submitted for obtaining the Master’s degree in specialty 122 “Computer Science”, is completed on 55 pages and contains 37 illustrations, appendices and 32 references.

The aim of the work is to develop and investigate a method for automatic classification of subjective and insincere questions in question-answering systems using a multilayer neural network to improve recognition accuracy of complex rhetorical and emotional formulations.

Research results: a multilayer neural network combining TF-IDF representation and recurrent layers SimpleRNN and BiGRU was developed for classification of subjective and insincere questions. The proposed method improves classification accuracy compared to baseline approaches and efficiently processes large text corpora considering class imbalance. Experimental studies and comparative analysis confirmed the effectiveness of the multilayer neural architecture.

Practical significance: the results can be applied in automated content moderation systems, educational platforms, chatbots, and online Q&A services for detecting provocative or potentially toxic questions, supporting moderators, improving user interaction quality, and adapting system response strategies.

Keywords: SUBJECTIVE QUESTIONS, INSINCERE QUESTIONS, NATURAL LANGUAGE PROCESSING, RECURRENT NEURAL NETWORK, BiGRU, TF-IDF, TEXT CLASSIFICATION, NLP.

ЗМІСТ

ВСТУП

Актуальність теми. Стрімке зростання обсягів користувацького контенту в системах питань-відповідей, соціальних мережах та форумах призводить до загострення проблеми автоматичної модерації запитань, що містять упередження, маніпуляції або мову ворожнечі. Значну частку такого контенту становлять суб’єктивні та зокрема інсинсирні запитання, які формулюються не з метою отримати чесну відповідь, а для провокування конфліктів, підтримки стереотипів чи легітимації дискримінаційних наративів. У цих умовах розроблення ефективних методів автоматичної класифікації суб’єктивних запитань є ключовою передумовою підвищення якості комунікації в онлайнових спільнотах та зменшення навантаження на модераторів.

Сучасні підходи до аналізу суб’єктивності здебільшого орієнтовані на тональність висловлювань або виявлення токсичних коментарів, тоді як специфіка запитань як окремого типу дискурсивних одиниць вивчена значно гірше. Інсинсирні запитання поєднують риси суб’єктивної оцінки та риторичних стратегій (іронія, навідні формулювання, узагальнення щодо соціальних груп), що ускладнює їхнє автоматичне розпізнавання традиційними методами машинного навчання на основі поверхневих ознак. Це зумовлює необхідність застосування багаторівневих нейронних архітектур, які здатні одночасно моделювати лексичні, контекстуальні та прагматичні характеристики тексту.

Стан розвитку глибинного навчання в галузі оброблення природної мови відкриває нові можливості для побудови таких моделей. Рекурентні нейронні мережі, зокрема SimpleRNN та двонапрямлені GRU, дають змогу моделювати послідовність слів у запитанні та враховувати довгострокові залежності, тоді як класичні методи на кшталт TF-IDF у поєднанні з ансамблевими алгоритмами (наприклад, XGBoost) можуть виконувати роль інтерпретованого базового рівня. Поєднання цих підходів у єдиній багаторівневій архітектурі дозволяє підвищити якість класифікації та забезпечити гнучкий баланс між точністю, обчислювальною складністю та можливістю практичного впровадження.

Додатковою мотивацією до вибору теми є її відповідність сучасним тенденціям розвитку комп'ютерних наук та потребам індустрії. Інтернет-платформи активно впроваджують алгоритми автоматичного відсіювання небажаного контенту, однак переважно зосереджуються на повідомленнях і коментарях, тоді як модерація запитань залишається менш формалізованою. Запропонований у роботі метод класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі може бути інтегрований у системи питань-відповідей, освітні платформи та чат-боти, підвищуючи безпечність і інформативність цифрового середовища.

Метою кваліфікаційної роботи є розроблення та дослідження методу класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі, який забезпечує підвищення якості виявлення інсинсирних запитань у великих корпусах даних у порівнянні з базовими підходами.

Для досягнення поставленої мети в роботі сформульовано та розв'язано такі основні завдання:

1. проаналізувати особливості суб'єктивних та інсинсирних запитань у сучасних системах питань-відповідей і визначити їхні лінгвістичні та прагматичні маркери;

2. здійснити огляд наукових підходів до виявлення суб'єктивності, інсинсирності й токсичності текстів, зокрема методів класичного машинного навчання та глибинних нейронних архітектур;

3. дослідити можливості моделей глибинного навчання для класифікації запитань і виконати характеристику релевантних корпусів даних, обравши репрезентативний датасет для експериментів;

4. спроєктувати базову модель класифікації інсинсирних запитань на основі TF-IDF-представлення тексту та алгоритму XGBoost і оцінити її якість;

5. розробити та реалізувати багаторівневу нейронну архітектуру, що поєднує шар ембедингів із рекурентними моделями SimpleRNN та двонапрямленою GRU, для послідовнісного моделювання тексту запитань;

6. здійснити експериментальне порівняння базової моделі та рекурентних архітектур за метриками accuracy та F1-score на валідаційній вибірці з урахуванням класового дисбалансу;

7. проаналізувати отримані результати, визначити переваги та обмеження запропонованого методу й сформулювати рекомендації щодо його практичного використання та подальшого розвитку.

Об’єкт і предмет дослідження. Об’єктом дослідження є процес автоматизованої класифікації запитань у системах питань-відповідей та інших текстових онлайн-сервісах.

Предметом дослідження є методи і моделі глибинного машинного навчання, зокрема багаторівневі нейронні мережі з рекурентними шарами, призначені для класифікації суб’єктивних та інсинсирних запитань на основі їхнього текстового представлення.

Методи дослідження. У роботі використано комплекс теоретичних і прикладних методів: аналіз і синтез наукових публікацій у галузі оброблення природної мови та модерації контенту; методи статистичного аналізу даних і класичного машинного навчання (TF-IDF-векторизація, ансамблеві моделі класифікації); методи глибинного навчання для послідовнісних даних (SimpleRNN, Bidirectional GRU); експериментальне моделювання з використанням відкритих корпусів запитань; кількісна оцінка якості моделей за стандартними метриками (accuracy, F1-score, precision, recall) та аналіз впливу класового дисбалансу.

Наукова новизна одержаних результатів. Наукова новизна роботи полягає в подальшому розвитку підходів до класифікації суб’єктивних та інсинсирних запитань шляхом побудови багаторівневої нейронної архітектури, яка поєднує базову статистичну модель TF-IDF + XGBoost із послідовнісними рекурентними моделями SimpleRNN та двонапрямленою GRU.

Практичне значення одержаних результатів. Практичне значення роботи полягає у можливості інтеграції запропонованого методу класифікації суб’єктивних запитань у системи модерації контенту та інтелектуальні модулі

платформ питань-відповідей. Розроблений програмний прототип дає змогу автоматично виявляти інсинсирні формулювання, що містять потенційно токсичний або провокативний зміст, і може бути використаний для підтримки рішень модераторів, попереднього фільтрування запитань чи адаптації стратегії відповіді в діалогових системах. Отримані результати можуть бути застосовані також у навчальному процесі при викладанні дисциплін, пов’язаних з інтелектуальним аналізом даних та обробленням природної мови.

Апробація результатів дослідження. Основні теоретичні положення роботи й практичні результати дослідження доповідалися й обговорювалися на студентської науково-практичної конференції «Інтелектуальні інформаційні технології в прикладних дослідженнях» (IITAR-2025), яка відбулася в місті Тернополі 27–29 травня 2025 року та IX Міжнародної студентської наукової конференції «Модернізація та сучасні українські і світові наукові дослідження» 14 листопада 2025 в місті Житомир, Україна.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАВДАННЯ КЛАСИФІКАЦІЇ СУБ'ЄКТИВНИХ ЗАПИТАНЬ

1.1 Особливості суб'єктивних запитань у сучасних системах питань-відповідей

Сучасні системи питань-відповідей (question answering, QA) працюють не лише з фактологічними запитами, а й з широким спектром комунікативних намірів користувачів – від пошуку думок і порад до емоційної підтримки в онлайнових спільнотах [1–3]. У таких середовищах, як Quora, StackExchange чи тематичні форуми, частка суб'єктивних запитань є суттєвою, а їхня коректна класифікація безпосередньо впливає на якість контенту, навантаження на модераторів і задоволеність користувачів [2]. Саме тому завдання автоматичного розпізнавання суб'єктивних і, зокрема, інсинсирних запитань сьогодні розглядається як критично важливий компонент інтелектуальних систем підтримки дискусій.

Під суб'єктивним запитанням зазвичай розуміють таке, відповідь на яке пов'язана з особистими оцінками, вподобаннями, переконаннями або емоціями, а не з перевірюваними фактами [4–6]. У подібних запитаннях користувач очікує множину можливих відповідей, що відображають різні точки зору, а не єдину «правильну» відповідь [3]. На відміну від цього, об'єктивні запитання спрямовані на отримання конкретної інформації з фіксованою і, як правило, верифікованою істинністю. Таке розрізнення є базовим для побудови типології запитань у QA-системах.

Окрему підгрупу суб'єктивних запитань становлять інсинсирні (insincere) запитання, які формулюються не з метою чесно дізнатися думку співрозмовників, а для провокування конфлікту, поширення ненависті, підтримки стереотипів або маніпуляції аудиторією [7–9]. Такі запитання можуть містити приховані звинувачення, риторичні конструкції, мову ворожнечі або відверті образи [21–24]. Для платформ на кшталт Quora чи Reddit наявність

інсинсирних запитань є джерелом токсичних обговорень і загрозою для іміджу сервісу, тому автоматичне виявлення таких формулювань набуває стратегічного значення.

З точки зору лінгвістики суб'єктивні запитання характеризуються підвищеною частотою оціночної лексики, прикметників із позитивною або негативною конотацією, інтенсивіфікаторів, модальних дієслів та займенників першої особи [4, 5]. Наприклад, конструкції «найкращий», «жахливий», «чи варто мені», «як ти ставишся» сигналізують про запит на оцінку або пораду, а не на фактологічну інформацію. Вони часто містять емоційно забарвлені іменники (наприклад, «зрада», «зловживання»), що прямо пов'язують запитання з суб'єктивною інтерпретацією ситуації.

Семантичним маркером суб'єктивності виступає також фокус на переживаннях або мотивації учасників події, а не на її об'єктивних параметрах [4]. Замість уточнення «коли відбулася подія X» суб'єктивне запитання формулюється як «чому люди відреагували на X саме так» або «чи правильно вчинив герой Y». Такі конструкції вимагають від відповідачів надання аргументованої думки, інтерпретації чи моральної оцінки, що суттєво ускладнює автоматичну обробку контенту.

На прагматичному рівні суб'єктивні запитання часто реалізують непрямі мовленнєві акти – прохання, критику або схвалення, замасковані під нейтральну форму питання [5]. Наприклад, «чи не є політика Z абсолютно неефективною?» вже містить вбудовану позицію автора і провокує відповіді в межах заданої рамки. Подібні «наведені» запитання є особливо проблемними для автоматичних систем, оскільки вимагають аналізу імплікатур, риторичних прийомів і політичного або культурного контексту.

Інсинсирні запитання, на відміну від нейтрально суб'єктивних, характеризуються навмисним порушенням норм доброзичливої комунікації. Їхнім прихованим завданням є загострення конфлікту, створення враження легітимності упереджених суджень або нормалізація мови ворожнечі [21–23]. Такі запитання часто містять узагальнення на кшталт «чому всі представники

групи G є…», використання стереотипних ярликів та образливих назв груп, що зближує їх із завданнями виявлення hate speech [9, 10, 23].

Для модераторів онлайн-платформ суб'єктивні та інсинсирні запитання мають різний статус. Перші зазвичай дозволені та навіть бажані, оскільки стимулюють дискусію й обмін досвідом; другі підлягають обмеженню або видаленню відповідно до політики спільноти [7, 21]. Тому модель класифікації повинна не лише відрізняти суб'єктивні запитання від об'єктивних, а й виділяти в межах суб'єктивних ті, що порушують етичні та правові норми.

З алгоритмічної точки зору суб'єктивні запитання є більш неоднорідними за формою, ніж фактологічні: вони допускають вільний порядок слів, широке використання розмовної лексики, емодзі, сленгу, а також багатомовних та код-міксових фрагментів [9, 23]. Це ускладнює використання поверхневих ознак на кшталт n-грам або шаблонів регулярних виразів і вимагає моделей, здатних уловлювати довгострокові залежності та латентні семантичні структури.

Крім лінгвістичних, для суб'єктивних запитань важливими є й позатекстові характеристики – інформація про автора, історію його активності, реакції спільноти, часовий та тематичний контекст [2, 26–28]. Наприклад, одне й те саме запитання, поставлене в різних спільнотах або в різні історичні моменти, може мати різний ступінь конфліктогенності. Однак у більшості відкритих корпусів, зокрема Quora Insincere Questions Classification, ці метадані або відсутні, або суттєво обмежені, що стимулює розвиток методів, орієнтованих переважно на текстову складову [7, 8].

Особливістю суб'єктивних запитань є також висока міжанотаторська варіативність під час розмітки. Навіть експерти можуть по-різному трактувати межу між щирим, але різким запитанням та інсинсирною провокацією [21, 23]. Це призводить до появи шуму в мітках, що негативно впливає на навчання моделей та вимагає використання методів, стійких до помилкових анотацій, наприклад, робастних функцій втрат або баганотаторських схем з агрегацією голосів [8].

З точки зору етики штучного інтелекту класифікація суб’єктивних запитань пов’язана з ризиками цензури та упередженості моделей [9, 23]. Надмірно агресивне віднесення запитань до класу «інсинсирних» може призвести до приглушення легітимної критики або вразити свободу вираження поглядів, тоді як надто «м’яка» модель не забезпечить достатнього рівня захисту від мови ворожнечі. Тому важливо враховувати баланс між precision та recall для «проблемного» класу й проводити регулярний аудит моделей на предмет упередженості щодо певних груп користувачів.

У контексті багатомовних спільнот постає питання переносимості моделей суб’єктивності між мовами. Вираження думок, ввічливості чи іронії суттєво залежить від граматичних та культурних особливостей конкретної мови [23]. Це особливо актуально для україномовних платформ, де поєднуються українська, російська, англійська та суржик, а стандартні англомовні лексикони суб’єктивності показують обмежену ефективність.

Для вирішення описаних проблем дедалі частіше застосовуються багаторівневі нейронні мережі, які поєднують рівень лексичних ембедингів, рекурентні або трансформерні шари та вихідні класифікаційні блоки [12–16]. Такі архітектури дозволяють моделі автоматично виявляти релевантні патерни суб’єктивності – від окремих слів до дискурсивних структур – без ручного конструювання ознак. У той же час вони потребують великих розмічених корпусів і ретельного контролю за перенавчанням, особливо в умовах значного класового дисбалансу, притаманного задачі інсинсирних запитань [8].

Нарешті, у діалогових системах та чат-ботах класифікація запитань за суб’єктивністю дозволяє адаптувати стратегій відповіді. Для фактологічних запитів система може звертатися до баз знань або пошуку, тоді як для суб’єктивних – пропонувати збалансовану аргументацію, посилатися на різні точки зору або попереджати про потенційно образливий характер контенту [11, 12]. Таким чином, розуміння особливостей суб’єктивних та інсинсирних запитань є ключовою передумовою для побудови безпечних та корисних систем питань-відповідей.

Узагальнюючи, суб’єктивні запитання в сучасних QA-системах характеризуються складною лінгвістичною структурою, високою прагматичною насиченістю та значною соціально-етичною вагою. Їхнє автоматичне розпізнавання вимагає моделей, що поєднують глибоке семантичне представлення тексту із здатністю виявляти тонкі риторичні та дискурсивні маркери, характерні для інсинсирної комунікації. Саме в такому контексті запропонований метод класифікації суб’єктивних запитань на основі багаторівневої нейронної мережі є обґрунтованим та актуальним напрямом досліджень.

1.2 Огляд наукових підходів до виявлення суб’єктивності та інсинсирності запитань

Перші роботи з виявлення суб’єктивності були зосереджені здебільшого на рівні речення та документу і виконувалися в рамках завдань аналізу тональності та opinion mining [5, 6]. Дослідники використовували лексикони суб’єктивної лексики, частотні списки емоційних прикметників, а також ручні правила, які описували типові патерни, наприклад «I think», «in my opinion» тощо. Попри простоту, такі підходи виявилися чутливими до мовних варіацій, ідіом, сарказму та контексту-залежних значень слів, що обмежило їхню придатність для автоматичної модерації запитань у великих спільнотах [4, 5].

Подальший розвиток відбувався в напрямі статистичного машинного навчання, де суб’єктивність стала розглядатися як задача бінарної або багатокласової класифікації текстів [4, 33]. У роботах цього етапу використовувалися bag-of-words і n-грамні представлення, які подавали речення або запитання у вигляді високовимірних розріджених векторів [6]. Для побудови моделей широко застосовувалися алгоритми SVM, логістична регресія, наївний Баєс і дерева рішень, що демонстрували задовільну точність на відносно

невеликих корпусах, але погано масштабувалися на сучасні масиви користувацького контенту.

Спеціалізовані дослідження суб'єктивності саме в запитаннях з'явилися в контексті систем спільнотних запитань-відповідей (Community Question Answering, CQA). Зокрема, було запропоновано класифікувати запитання на суб'єктивні, об'єктивні та соціальні, комбінуючи текстові ознаки, метадані та структуру взаємодій у спільноті [2]. Інша лінія робіт запропонувала визначати суб'єктивність як окремий аспект наміру автора, що дозволяє відрізняти запити на факти, думки або досвід користувачів [3, 34]. У цих підходах використовувалися як поверхневі лінгвістичні ознаки, так і спеціальні індикатори дискурсивних ролей (наприклад, «прохання поради», «оцінка продукту»).

Важливою віхою став перехід до розподілених векторних представлень слів (word embeddings), таких як Word2Vec, GloVe та fastText [17–19]. Ці моделі дозволили відображати лексичні одиниці в компактний латентний простір, де семантично подібні слова розташовані близько одне до одного. Завдяки цьому стало можливим застосовувати до задачі суб'єктивності нейронні мережі, які спираються не на окремі ознаки, а на континуальні представлення, здатні узагальнювати знання про лексику поза межами тренувального корпусу.

Паралельно для класифікації коротких текстів, зокрема запитань і твітів, активно досліджувалися згорткові нейронні мережі (CNN) та різні варіанти рекурентних нейронних мереж (RNN) [13, 15, 16]. CNN демонструють високу ефективність у виявленні локальних n-грамних шаблонів, тоді як RNN, LSTM і GRU краще моделюють довгострокові залежності в послідовності слів [13, 15]. Для задач виявлення суб'єктивності це означає можливість одночасно враховувати як локальні індикатори (окремі оціночні слова), так і глобальну структуру запитання (контрастні конструкції, риторичні запитання, сарказм).

Окремий напрям становлять роботи, спрямовані на класифікацію запитань за типом у системах автоматичного відповідання – наприклад, розподіл на фактологічні, визначальні, процедурні та opinion-запитання [1, 26]. У таких системах суб'єктивність фігурує як одна з категорій або як допоміжний сигнал

для вибору стратегії пошуку відповіді. Зазвичай для цієї мети комбінуються ручні шаблони, статистичні класифікатори та синтаксичні ознаки, отримані з дерев залежностей.

Становлення задачі виявлення інсинсирних запитань відбулося на хвилі досліджень hate speech, offensive language та токсичної поведінки в соціальних мережах [21–24]. У цих роботах тексти класифікувалися на образливі, ненависницькі, дискримінаційні тощо, використовуючи лексичні списки, n-грами, спеціальні маркери нецензурної лексики та імена груп [21, 22]. Хоча прямим об’єктом аналізу були переважно твіти або коментарі, методологічно ці підходи виявилися близькими до задачі інсинсирних запитань, де токсичність виражається у формі провокативних формулювань.

Подальші дослідження у сфері hate speech продемонстрували переваги глибоких нейронних мереж – CNN, BiLSTM, BiGRU – над класичними методами, особливо за наявності достатньо великих корпусів [9, 10, 23]. Порівняльні дослідження показали, що двонапрямлені рекурентні архітектури краще уловлюють контекстуальні підказки й можуть враховувати взаємодію між віддаленими словами, що критично для виявлення завуальованих образ [9, 31, 35]. Для задачі інсинсирних запитань це означає, що модель здатна оцінювати не лише локальні маркери агресії, а й загальну риторичну структуру запитання.

Поява корпусу Quora Insincere Questions Classification надала дослідникам масштабний датасет із понад мільйоном англомовних запитань, позначених як «щирі» та «інсинсирні» [7, 8]. На його основі було запропоновано численні підходи – від класичних ансамблів на основі TF-IDF і XGBoost до глибоких рекурентних мереж із увагою та трансформерних моделей BERT-типу [8, 24, 32]. Окремі роботи демонструють, що використання попередньо навчених мовних моделей разом із спеціалізованими шарами уваги дозволяє суттєво покращити F1-міру для «інсинсирного» класу порівняно з базовими моделями [8, 32].

У той же час суттєвим викликом для всіх підходів залишається класовий дисбаланс: частка інсинсирних запитань у корпусі Quora становить близько 6 %, що призводить до зміщення моделей у бік «безпечного» передбачення класу 0 [7,

8]. У літературі пропонуються різні стратегії – від надсемплінгу «рідкісного» класу та використання методу SMOTE до застосування зважених функцій втрат і спеціальних метрик оптимізації, таких як macro-F1 або Fβ із підвищеною вагою для небажаного класу [9, 23]. У дослідженнях на Quora-корпусі показано, що коректний вибір метрики та порога бінаризації ймовірностей може суттєво змінити баланс між точністю і повнотою виявлення інсинсирних запитань [8].

Окрему групу складають роботи, які накладають завдання суб'єктивності та інсинсирності на багатозадачні архітектури. Наприклад, одночасне навчання моделі на класифікацію тональності, токсичності й суб'єктивності дозволяє спільно використовувати ознаки й покращувати узагальнювальну здатність моделі [9, 23]. Інші підходи доповнюють класичну класифікацію prediction-модулями для оцінки емоцій, виявлення сарказму чи визначення тематики запитання [10]. Такі мультимодульні системи фактично реалізують багаторівневі нейронні мережі, де кожен рівень відповідає за окремий аспект дискурсу.

На найновішому етапі розвитку активно досліджується застосування великих мовних моделей (LLM) – GPT-подібних архітектур, інструкційно налаштованих трансформерів – до класифікації питань, зокрема освітніх, діалогових і суб'єктивних [11, 25]. LLM-моделі демонструють здатність узагальнювати знання з широкого кола доменів і працювати в режимі few-shot або zero-shot, коли для нової задачі доступно небагато розмічених прикладів. Однак їх застосування у високоризикових завданнях модерації вимагає додаткових механізмів контролю, оскільки моделей важче інтерпретувати, а вартість інференсу значно вища, ніж у більш компактних рекурентних архітектур.

Узагальнюючи літературний огляд, можна виділити кілька еволюційних етапів: від лексикон-орієнтованих правил до статистичних моделей, від класичних ознак до розподілених представлень, від «плоских» класифікаторів до глибоких рекурентних і трансформерних архітектур. На кожному етапі покращувалась здатність моделей фіксувати контекстуальні залежності й латентні семантичні структури, однак питання інтерпретації, стійкості до шуму

та врахування етичних аспектів залишаються відкритими [5, 6, 9]. Це створює підґрунтя для подальших досліджень, у тому числі для розроблення методів класифікації суб’єктивних запитань на основі багаторівневих нейронних мереж, які поєднують високу якість прогнозування з відносною архітектурною простотою.

Запропонований у даній роботі підхід спирається саме на таку лінію розвитку. З одного боку, він успадковує від класичних методів використання попередньої токенізації, нормалізації та статистичних базових моделей (TF-IDF + XGBoost) як орієнтиру. З іншого – включає глибші рекурентні архітектури SimpleRNN та двонапрямлений GRU, які дозволяють моделювати послідовність слів у запитанні на різних рівнях абстракції. У сукупності це формує багаторівневу нейронну мережу, здатну одночасно враховувати локальні патерни суб’єктивності та глобальну структуру запитання.

1.3 Моделі глибинного навчання для класифікації запитань і характеристика корпусів даних

Глибинне навчання за останнє десятиліття стало домінуючою парадигмою в задачах оброблення природної мови, зокрема – у класифікації запитань у системах питань-відповідей [13–16]. На відміну від традиційних підходів, що покладаються на вручну сконструйовані ознаки, нейронні мережі автоматично навчаються багаторівневих представлень тексту, що дає змогу моделювати складні семантичні та синтаксичні залежності між словами [17–19]. Це особливо важливо для суб’єктивних та інсинсирних запитань, де поєднання лексичних, контекстуальних і прагматичних сигналів є ключовим для коректної класифікації [8, 9].

Базовою передумовою побудови глибинних моделей є використання розподілених векторних представлень слів (word embeddings). Такі моделі, як Word2Vec, GloVe та fastText, навчають вектори фіксованої розмірності, у яких

близькі за значенням слова розташовуються в суміжних областях латентного простору [17–19]. У задачі класифікації запитань ембедінги виступають нижнім рівнем багаторівневої мережі, забезпечуючи компактний та інформативний опис лексики, що надалі обробляється рекурентними, згортковими або трансформерними шарами.

Рекурентні нейронні мережі (RNN) є природним вибором для роботи з текстом, оскільки вони явно моделюють послідовність слів та їхній порядок у реченні [13]. Класичний варіант RNN (SimpleRNN) обчислює прихований стан як функцію поточного слова та попереднього стану, що дає змогу акумулювати інформацію вздовж усього запитання. Однак такі мережі страждають від проблеми зникання/вибуху градієнтів, що обмежує їхню здатність моделювати довготривалі залежності, характерні для складних дискурсивних структур [15].

Для подолання зазначених обмежень були запропоновані архітектури LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit), у яких вводяться механізми “воріт”, що контролюють потік інформації в часі [13, 15]. LSTM використовує комбінацію вхідних, вихідних та забувальних воріт для керування оновленням внутрішнього стану, тоді як GRU застосовує більш компактну схему з оновлювальними та редукційними воротами. Обидва підходи значно покращують здатність моделі фіксувати віддалені залежності, зокрема зв’язки між початковою рамкою запитання та його риторичним завершенням, що є критичним для виявлення інсинсирності [9, 13].

Подальшим удосконаленням рекурентних моделей стали двонапрямлені архітектури (BiRNN, BiLSTM, BiGRU), у яких послідовність обробляється одночасно зліва-направо та справа-наліво [14]. Це дозволяє мережі враховувати як попередній, так і наступний контекст для кожного слова, що особливо важливо у випадках, коли смисл маркера суб’єктивності залежить від подальших компонентів (наприклад, у конструкціях «чи не здається вам, що…»). Дослідження у сфері аналізу тональності та hate speech демонструють, що двонапрямлені GRU/LSTM стабільно перевершують односпрямовані аналоги за F1-мірою, особливо для «проблемних» класів [9, 23, 31].

Окремий клас моделей становлять згорткові нейронні мережі (CNN) для класифікації речень [16, 20]. Вони застосовують згорткові фільтри до послідовності ембедінгів, виділяючи локальні n-грамні патерни, які можуть відповідати типовим маркерам суб'єктивності чи агресії. Підхід CNN є обчислювально ефективним і добре підходить для коротких запитань, однак гірше відстежує довгострокові залежності й дискурсивні структури, порівняно з рекурентними або трансформерними архітектурами [13, 16].

У сучасних роботах моделі на основі рекурентних шарів часто доповнюються механізмом уваги (attention), який дозволяє «зважувати» внесок окремих слів у фінальне представлення запитання [8, 32]. У випадку суб'єктивних і інсинсирних запитань це дає змогу сфокусуватися на ключовій оціночній лексиці, узагальненнях або стереотипних позначеннях груп, ігноруючи другорядні елементи [23]. У роботах на корпусі Quora Insincere Questions було показано, що додавання уваги до BiLSTM/GRU-шарів суттєво підвищує F1-міру для «інсинсирного» класу порівняно з базовими рекурентними моделями [8, 32].

Новим етапом розвитку стали трансформерні мовні моделі, зокрема BERT, RoBERTa, DistilBERT, які використовують багатошарову самоувагу для моделювання залежностей у тексті [12]. Вони попередньо навчаються на великих неанотованих корпусах із використанням задач маскування токенів та прогнозування наступного речення, а потім донавчаються (fine-tuning) на конкретних задачах, серед яких – класифікація запитань, визначення наміру користувача та токсичності [11, 12]. Порівняльні дослідження показують, що навіть відносно компактні моделі, як-от DistilBERT, забезпечують значний приріст якості порівняно з «класичними» RNN- або CNN-архітектурами, особливо на великих корпусах [11].

Разом з тим трансформери мають низку недоліків у прикладних системах модерації контенту. По-перше, їхня обчислювальна вартість істотно перевищує вартість рекурентних моделей, що ускладнює розгортання в режимі реального часу для високонавантажених платформ [11, 12]. По-друге, через високу модельну складність ускладнюється інтерпретація рішень, що є проблемою в

контексті етики ШІ та вимог прозорості алгоритмів модерації [9]. Тому в ряді практичних сценаріїв доцільно використовувати гібридні або багаторівневі архітектури, які поєднують простіші моделі (наприклад, TF-IDF + XGBoost) із глибшими рекурентними мережами як «другу лінію захисту» [8].

Усі описані моделі потребують якісних корпусів даних для навчання та оцінювання. Для фактологічних запитань широко використовуються колекції SQuAD, TREC QA, MS MARCO, Natural Questions, у яких запитання поєднуються з відповідними фрагментами тексту [1, 26]. Натомість для дослідження суб'єктивності й інсинсирності ключову роль відіграють корпуси соціальних платформ – Twitter, Reddit, Quora – де запитання та коментарі містять виразну емоційно-оцінну складову [7–9, 21, 23]. Важливо, що схеми анотування таких корпусів відрізняються: одні концентруються на тональності, інші – на токсичності, треті – на щирості запитань, що ускладнює безпосереднє порівняння моделей.

Корпус Quora Insincere Questions Classification, який використовується в даному дослідженні, є одним із найбільших публічно доступних наборів для виявлення інсинсирних запитань [7, 8]. Він містить понад 1,3 млн англомовних запитань із бінарною розміткою (0 – щире, 1 – інсинсирне), при цьому частка позитивного класу становить близько 6 %, що зумовлює виражений класовий дисбаланс [7]. Запитання є відносно короткими (медіана – близько 11–13 слів), але демонструють значну лексичну різноманітність і охоплюють політичні, соціальні, релігійні та особистісні теми, в яких часто проявляються стереотипи та мова ворожнечі [8].

Важливим аспектом характеристики корпусу є якість та узгодженість анотації. У Quora Insincere питання розмічалися з урахуванням політик спільноти щодо неприйнятної поведінки, однак межа між «гострим, але щирим» запитанням і відверто провокативним формулюванням залишається розмитою [7, 8]. Це призводить до появи «сірої зони», де різні анотатори можуть по-різному інтерпретувати наміри автора, що зумовлює наявність шуму в мітках. Для

моделей глибинного навчання така ситуація є типовою і вимагає застосування робастних методів навчання та ретельного аналізу помилок [9, 23].

Окрім Quora, у літературі описано низку спеціалізованих корпусів для виявлення ненависті, образ та токсичності в коментарях, зокрема колекції Davidson et al., Waseem & Hovy, а також набори даних великих онлайн-платформ [21–24]. Хоча ці корпуси здебільшого містять твітів або короткі репліки, а не запитання, накопичений на них досвід побудови моделей hate speech безпосередньо переноситься на задачу інсинсирних запитань – зокрема в частині використання двонапрямлених рекурентних мереж та трансформерів [9, 23, 29–31]. Таким чином, Quora Insincere можна розглядати як перехідний формат між класичними задачами токсичності та більш загальним класифікуванням намірів у запитаннях.

Суттєвий вплив на якість класифікації має попередня обробка тексту й спосіб подання корпусу моделі. Для класичних підходів (TF-IDF + XGBoost) поширеним є токенізація, видалення стоп-слів, лематизація чи стемінг та перетворення запитань у розріджені вектори частот [6, 19]. У глибинних моделях іноді відмовляються від агресивного фільтрування, натомість покладаючись на ембедінги та рекурентні шари, здатні самостійно навчитися ігнорувати малозначущі токени [13, 16]. У випадку трансформерів стандартом є підхід субсловної токенізації (WordPiece, BPE), який дозволяє ефективно працювати з рідкісними словами і не потребує попереднього стемінгу [12].

Завдання класової нерівноваги в корпусах інсинсирних запитань вирішується різними стратегіями. На рівні даних застосовують надсемплінг рідкісного класу, генерацію штучних прикладів (SMOTE) або цілеспрямований даунсемплінг класу більшості [8, 9]. На рівні моделі використовують зважування втрат, фокус-втрату (focal loss) або адаптивний вибір порога бінаризації ймовірностей, оптимізуючи саме F1-міру для «інсинсирного» класу [9, 23]. Багато робіт демонструють, що правильне поєднання цих стратегій є не менш важливим, ніж вибір конкретної архітектури мережі [29–31].

З практичного погляду важливо також забезпечити узагальнюваність моделей за межі одного корпусу. Моделі, навчені на Quora, можуть демонструвати зниження якості на інших платформах через відмінності у стилі, довжині та тематиці запитань, а також через варіації в політиках модерації [2, 27, 28]. Для підвищення переносимості використовують багатодоменне навчання, domain adaptation та попереднє донавчання моделей на великих неанотованих корпусах із цільового домену [11, 12].

У цьому контексті багаторівневі нейронні архітектури для класифікації суб'єктивних запитань можуть розглядатися як компроміс між складністю трансформерів і простотою класичних методів. Нижній рівень моделі забезпечує базову фільтрацію та грубу оцінку суб'єктивності (наприклад, TF-IDF + XGBoost), тоді як верхні рекурентні шари (SimpleRNN, BiGRU) уточнюють рішення, враховуючи послідовну структуру запитання та складні контекстуальні патерни [8, 13]. Така ієрархія дозволяє зменшити витрати ресурсів і водночас досягти високої точності виявлення інсинсирних формулювань.

Підсумовуючи, моделі глибинного навчання для класифікації запитань пройшли шлях від простих RNN та CNN до багатошарових двонапрямлених рекурентних мереж і трансформерів, а корпуси даних – від невеликих наборів фактологічних запитань до багатомільйонних колекцій користувацького контенту, таких як Quora Insincere [1, 7, 8, 13, 16]. Попри значний прогрес, залишається низка викликів, пов'язаних із класовим дисбалансом, шумністю анотацій, переносимістю між доменами та етичними аспектами використання моделей [9, 21–24]. Ці обмеження визначають актуальність подальших досліджень, спрямованих на розроблення ефективних багаторівневих нейронних методів класифікації суб'єктивних запитань, які враховують як властивості сучасних корпусів даних, так і вимоги до відповідального використання ШІ.

1.4 Постановка задачі

У сучасних системах питань-відповідей спостерігається стрімке зростання обсягів користувацького контенту, значну частку якого становлять саме суб’єктивні запитання. Такі запитання є не просто носіями інформації, а відображають переконання, емоції та соціальні настанови користувачів. Наявність великої кількості суб’єктивних і, зокрема, інсинсирних запитань безпосередньо впливає на якість дискусій, довіру до платформи та рівень інформаційної безпеки, що робить задачу їх автоматичного розпізнавання надзвичайно актуальною.

Особливої ваги набуває проблема інсинсирних запитань, які формулюються не для отримання чесної відповіді, а для провокування конфліктів, відтворення стереотипів і легітимації мови ворожнечі. Такі запитання стають тригером токсичних обговорень, поляризації спільнот, підриву репутації платформ і можуть мати реальні соціальні наслідки. В умовах зростання чутливості суспільства до питань дискримінації та онлайн-агресії розроблення надійних алгоритмів виявлення інсинсирних формулювань є важливою складовою побудови безпечного цифрового середовища.

Актуальність дослідження зумовлена також обмеженістю традиційних методів аналізу тексту, які базуються на поверхневих ознаках – n-грамах, лексиконах суб’єктивної лексики, статичних порогах. Такі підходи часто виявляються неефективними у випадках, коли суб’єктивність і інсинсирність реалізуються через складні риторичні конструкції, імплікатурні смисли, контекстуально залежну іронію чи завуальовану агресію. Це вимагає застосування більш потужних моделей, здатних опрацьовувати послідовну структуру запитання та виявляти латентні семантичні й прагматичні патерни.

Важливим аргументом на користь актуальності є наявність вираженого класового дисбалансу у реальних даних: частка інсинсирних запитань у великих корпусах, таких як Quora Insincere Questions Classification, становить лише кілька відсотків від загальної кількості. У таких умовах базові моделі часто «ігнорують»

рідкісний клас, забезпечуючи високі значення загальної точності, але низьку повноту саме для проблемних запитань. Це створює розрив між формальною якістю моделі та її реальним практичним ефектом, що підсилює потребу в спеціалізованих методах, орієнтованих на баланс показників для інсинсирного класу.

Не менш суттєвою є етична складова. Автоматизація модерації суб'єктивних запитань пов'язана з ризиком помилкової класифікації: надмірне «перехоплення» може обмежувати легітимну критику та вільне висловлення думок, тоді як недостатня чутливість моделі залишатиме непоміченими дійсно токсичні й небезпечні формулювання. Пошук збалансованих рішень, які враховують як ефективність виявлення інсинсирних запитань, так і вимоги до справедливості, прозорості та недискримінаційності, є однією з центральних задач сучасних досліджень у галузі відповідального ШІ.

Наукова й практична значущість теми посилюється необхідністю масштабованих рішень для високонавантажених платформ. Ручна модерація в умовах мільйонів запитань є економічно та організаційно нереалістичною, а прості евристики не забезпечують достатньої якості. Багаторівневі нейронні архітектури на основі ембедингів, рекурентних шарів SimpleRNN і двонапрямлених GRU демонструють здатність до більш глибокого аналізу тексту при збереженні відносно помірних обчислювальних витрат, що робить їх придатними для практичного впровадження.

Актуальність обраної тематики підкріплюється й тим, що значна частина попередніх робіт зосереджена або на загальному виявленні токсичності, або на класифікації тональності, а специфіка саме запитань як комунікативних одиниць часто лишається поза фокусом. Інсинсирне запитання поєднує властивості суб'єктивної оцінки, риторики та часто прихованої агресії, що вимагає окремих моделей і спеціалізованих корпусів. Дослідження, орієнтовані саме на класифікацію суб'єктивних і інсинсирних запитань, заповнюють цю прогалину.

З погляду розвитку теорії та практики глибинного навчання важливою є перевірка ефективності рекурентних архітектур на великих, реалістичних

корпусах, таких як Quora Insincere. Це дозволяє не лише порівняти різні моделі (TF-IDF + XGBoost, SimpleRNN, BiGRU) за стандартними метриками, а й виявити їхні сильні та слабкі сторони в умовах шумних міток, класового дисбалансу та високої лексичної варіативності. Отримані висновки мають значення для проєктування майбутніх систем модерації та діалогових агентів.

Для українського наукового й освітнього контексту обрана тема є актуальною ще й тому, що підходи до класифікації суб'єктивних та інсинсирних запитань можуть бути адаптовані до україномовних платформ, форумів, освітніх сервісів і чат-ботів. Хоча експерименти в роботі виконуються на англомовному корпусі, розроблені методичні засади, архітектурні рішення та підхід до оцінювання якості є переносними й створюють основу для подальших досліджень у багатомовному середовищі, де українська мова співіснує з іншими.

Узагальнюючи, актуальність дослідження полягає в поєднанні кількох взаємопов'язаних чинників: зростання ролі суб'єктивних та інсинсирних запитань у сучасних онлайн-комунікаціях, обмеженість традиційних методів їхнього автоматичного виявлення, потреба в масштабованих і етично відповідальних рішеннях для модерації контенту, а також можливість адаптації запропонованих підходів до різних мов і платформ. У цьому контексті розроблення методу класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі є своєчасним і науково обґрунтованим завданням.

Метою кваліфікаційної роботи є розроблення та дослідження методу класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі, який забезпечує підвищення якості виявлення інсинсирних запитань у великих корпусах даних у порівнянні з базовими підходами.

Для досягнення поставленої мети в роботі сформульовано та розв'язано такі основні завдання:

1. проаналізувати особливості суб'єктивних та інсинсирних запитань у сучасних системах питань-відповідей і визначити їхні лінгвістичні та прагматичні маркери;

2. здійснити огляд наукових підходів до виявлення суб’єктивності, інсинсирності й токсичності текстів, зокрема методів класичного машинного навчання та глибинних нейронних архітектур;

3. дослідити можливості моделей глибинного навчання для класифікації запитань і виконати характеристику релевантних корпусів даних, обравши репрезентативний датасет для експериментів;

4. спроєктувати базову модель класифікації інсинсирних запитань на основі TF-IDF-представлення тексту та алгоритму XGBoost і оцінити її якість;

5. розробити та реалізувати багаторівневу нейронну архітектуру, що поєднує шар ембедингів із рекурентними моделями SimpleRNN та двонапрямленою GRU, для послідовнісного моделювання тексту запитань;

6. здійснити експериментальне порівняння базової моделі та рекурентних архітектур за метриками accuracy та F1-score на валідаційній вибірці з урахуванням класового дисбалансу;

7. проаналізувати отримані результати, визначити переваги та обмеження запропонованого методу й сформулювати рекомендації щодо його практичного використання та подальшого розвитку.

Висновки до розділу 1

У першому розділі показано, що сучасні системи питань-відповідей та онлайнові платформи генерують значні обсяги користувацького контенту, в якому суб’єктивні та інсинсирні запитання становлять окрему, особливо ризикову категорію. Визначено, що такі запитання поєднують лінгвістичні (оцінна лексика, узагальнення, стереотипні позначення груп) та прагматичні маркери (навідні формулювання, риторичні питання, імплікатурні смисли), що робить їхню автоматичну ідентифікацію нетривіальною задачею. Обґрунтовано, що саме інсинсирні формулювання слугують тригерами токсичних дискусій, поляризації

спільнот і підриву довіри до платформи, а отже, вимагають спеціалізованих методів виявлення.

На основі огляду наукових підходів до виявлення суб'єктивності, інсинсирності та токсичності текстів показано еволюцію методів – від лексикон-орієнтованих правил та статистичних моделей на основі bag-of-words і n-грам до глибинних нейронних архітектур, що працюють із розподіленими векторними представленнями. Підкреслено, що хоча класичні алгоритми (SVM, логістична регресія, XGBoost) забезпечують прийнятну якість на невеликих корпусах, вони погано враховують послідовний контекст та риторичну структуру запитань, що обмежує їхню придатність для задачі інсинсирних запитань.

У розділі проаналізовано можливості моделей глибинного навчання – рекурентних мереж (SimpleRNN, LSTM, GRU, BiGRU), CNN-архітектур та трансформерів – для класифікації запитань, а також охарактеризовано релевантні корпуси даних, зокрема Quora Insincere Questions Classification. Показано, що великі реалістичні набори даних поєднують високу лексичну варіативність, виражений класовий дисбаланс і шумність анотацій, що висуває додаткові вимоги до стійкості моделей та методів навчання. Окремо акцентовано на етичних викликах – ризиках цензури, упередженості моделей та необхідності збалансованого поєднання показників точності й повноти для «проблемного» класу.

На основі проведеного аналізу сформульовано постановку задачі класифікації суб'єктивних запитань як задачі бінарної класифікації, орієнтованої насамперед на надійне виявлення інсинсирних формулювань. Обґрунтовано доцільність використання багаторівневої нейронної мережі, яка поєднує базову статистичну модель (TF-IDF + XGBoost) з рекурентними шарами SimpleRNN та двонапрямленою GRU, здатними моделювати послідовний контекст і тонкі риторичні патерни. Тим самим визначено наукову й практичну актуальність подальшої розробки та дослідження такого методу.

2 МЕТОД КЛАСИФІКАЦІЇ СУБ'ЄКТИВНИХ ЗАПИТАНЬ НА ОСНОВІ БАГАТОРІВНЕВОЇ НЕЙРОННОЇ МЕРЕЖІ

2.1 Формалізація розробленого методу

У межах даного дослідження задача класифікації суб'єктивних запитань формулюється як задача бінарної класифікації текстів. Нехай задано множину запитань

$$\mathcal{D} = \{(q_i, y_i)\}_{i=1}^{N}, \tag{2.1}$$

де q_i– текст i-го запитання природною мовою, а $y_i \in \{0,1\}$– бінарна мітка класу. Значення $y_i = 1$ інтерпретується як суб'єктивне (нещире, упереджене, провокативне) запитання, тоді як $y_i = 0$відповідає нейтральному або щирому запитанню. Метою є побудова класифікатора

$$f_\theta : q \mapsto \hat{y} \in \{0,1\}, \tag{2.2}$$

параметризованого вектором параметрів θ, який для нового запитання qзабезпечує коректне віднесення до одного з класів.

На першому етапі здійснюється попередня обробка та формальне представлення текстових даних. Кожне запитання q_iрозглядається як послідовність токенів (слів або підслівних одиниць):

$$q_i = (w_{i,1}, w_{i,2}, \dots, w_{i,L_i}), \tag{2.3}$$

де L_i– довжина запитання у словах. На основі всієї навчальної множини $\mathcal{D}$формується словник V, який містить $|V|$найчастіше вживаних токенів (обмеження за розміром словника запобігає надмірній розмірності моделі). Визначається відображення

$$\text{idx}: V \cup \{<\text{OOV}>\} \to \{1,2, \dots, |\ V\ |\}, \tag{2.4}$$

яке кожному слову ставить у відповідність унікальний індекс, а рідкісні або невідомі токени агрегуються у спеціальний символ <OOV>.

На другому етапі текстові запитання переводяться у числовий формат фіксованої довжини. Для кожного q_iбудується послідовність індексів

$$x_i = (\text{idx}(w_{i,1}), \dots, \text{idx}(w_{i,L_i})), \tag{2.5}$$

після чого виконується обрізання або доповнення послідовності до фіксованої довжини T:

$$\tilde{x}_i = (x_{i,1}, \dots, x_{i,T}), x_{i,t} = \begin{cases} \text{idx}(w_{i,t}), & t \le L_i, \\ 0, & t > L_i, \end{cases} \tag{2.6}$$

де індекс 0 зарезервований як паддінг-символ. Таким чином, кожне запитання у вхідному шарі нейронної мережі представлено як вектор $\tilde{x}_i \in \{0, \dots, |\ V\ |\}^T$сталої довжини.

Третій етап полягає у переході від індексів токенів до щільного векторного подання в просторі ознак. Для цього використовується embedding-шар, який реалізує відображення

$$E: \{0, \dots, |\ V\ |\} \to \mathbb{R}^d, \tag{2.7}$$

що параметризується матрицею ембеддингів $W^{(e)} \in \mathbb{R}^{(|V|+1)\times d}$. Для кожного індексу jвідповідний вектор $\text{e}_j = W^{(e)}_{j,:}$є d-вимірним представленням слова. Таким чином, дискретна послідовність $\tilde{x}_i$перетворюється на послідовність векторів

$$\mathrm{X}_i = (\mathrm{e}_{x_{i,1}}, \mathrm{e}_{x_{i,2}}, \dots, \mathrm{e}_{x_{i,T}}), \mathrm{e}_{x_{i,t}} \in \mathbb{R}^d, \tag{2.8}$$

яка надалі обробляється рекурентними шарами.

Четвертий етап – моделювання послідовного контексту за допомогою багаторівневої рекурентної нейронної мережі. У найпростішому випадку використовується одношаровий SimpleRNN, де прихований стан $\mathrm{h}_t \in \mathbb{R}^k$на кроці tобчислюється за правилом

$$\mathrm{h}_t = \phi(W^{(h)}\mathrm{e}_{x_{i,t}} + U^{(h)}\mathrm{h}_{t-1} + \mathrm{b}^{(h)}), \tag{2.9}$$

де $W^{(h)}, U^{(h)}$– матриці ваг, $\mathrm{b}^{(h)}$– вектор зміщень, $\phi(\cdot)$– нелінійна активаційна функція (наприклад, tanh). Початковий стан h_0задається нульовим вектором. Після обробки всієї послідовності отримуємо фінальний прихований стан h_T, який використовується як узагальнене векторне представлення запитання.

Для підвищення здатності моделі враховувати контекст як зліва направо, так і справа наліво застосовується двонапрямлена GRU (Bidirectional GRU). У цьому випадку визначаються дві рекурентні гілки: пряма, що обробляє послідовність $(\mathrm{e}_{x_{i,1}}, \dots, \mathrm{e}_{x_{i,T}})$, та зворотна, що обробляє її у зворотному порядку. Для кожного кроку tобчислюються прямий стан $\overrightarrow{\mathrm{h}}_t$та зворотний стан $\overleftarrow{\mathrm{h}}_t$, після чого формують конкатенований вектор

$$\mathrm{h}_t^* = \left[\overrightarrow{\mathrm{h}}_t; \overleftarrow{\mathrm{h}}_t\right] \in \mathbb{R}^{2k}. \tag{2.10}$$

Як вектор документного подання використовується фінальний агрегований стан h_T^*, який містить інформацію про всю послідовність у двох напрямах. Для GRU-осередків перехідні рівняння включають гейти оновлення та скидання, що дозволяє краще контролювати довгострокові залежності та зменшувати проблему зникання градієнтів.

Наступний рівень нейронної мережі – шар регуляризації Dropout. На етапі навчання для вектора h(SimpleRNN або BiGRU представлення) формується маска випадкових нулів $\mathrm{m} \sim \mathrm{Bernoulli}(p)$, і на вхід подається модифікований вектор $\tilde{\mathrm{h}} = \mathrm{m} \odot \mathrm{h}$, де $\odot$позначає поелементне множення. Це зменшує ризик перенавчання, примушуючи мережу не покладатися на окремі домінуючі компоненти представлення і підвищує узагальнювальну здатність моделі при класифікації нових запитань.

Фінальний рівень багаторівневої нейронної мережі – вихідний щільний (Dense) шар із сигмоїдною активацією. Для обраного векторного представлення $\mathrm{z} \in \mathbb{R}^m$(SimpleRNN: h_T, BiGRU: h_T^*) обчислюється скаляр

$$s = \mathrm{w}^\top \mathrm{z} + b, \tag{2.11}$$

де $\mathrm{w} \in \mathbb{R}^m$, $b \in \mathbb{R}$– параметри вихідного шару. Ймовірність належності запитання до суб'єктивного класу визначається як

$$\hat{p}(\, y = 1 \mid q\,) = \sigma(s) = \frac{1}{1+e^{-s}}. \tag{2.12}$$

Рішення класифікатора приймається за правилом порогового відсікання:

$$\hat{y} = \begin{cases} 1, & \text{якщо } \hat{p}(\, y = 1 \mid q\,) \geq \tau, \\ 0, & \text{інакше,} \end{cases} \tag{2.13}$$

де $\tau \in (0,1)$– поріг, який за замовчуванням може дорівнювати 0.5 або підбиратися за F1-мірою на валідаційній вибірці.

Навчання параметрів $\theta = \{W^{(e)}, W^{(h)}, U^{(h)}, \mathrm{b}^{(h)}, \mathrm{w}, b, \dots\}$здійснюється методом стохастичного градієнтного спуску з оптимізатором Adam. В якості функції втрат використовується бінарна крос-ентропія:

$$\mathcal{L}(\theta) = -\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} [y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)], \qquad (2.14)$$

де $\hat{p}_i = \hat{p}(y = 1 \mid q_i)$. На кожній ітерації параметри оновлюються у напрямку мінімізації $\mathcal{L}(\theta)$з використанням міні-батчів фіксованого розміру, що забезпечує ефективне навчання на великому корпусі запитань.

Для оцінювання якості методу на валідаційній множині застосовуються як загальна точність (accuracy), так і більш чутливі до дисбалансу метрики – F1-міра для суб'єктивного класу та макро-середній F1. Нехай $\text{TP}, \text{FP}, \text{FN}$– кількість істинно-позитивних, хибно-позитивних та хибно-негативних класифікацій для класу 1. Тоді precision, recall та F1-міра визначаються як

$$\text{Prec} = \frac{\text{TP}}{\text{TP+FP}}, \text{Rec} = \frac{\text{TP}}{\text{TP+FN}}, \text{F1} = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec+Rec}}. \qquad (2.15)$$

Максимізація F1-міри саме для суб'єктивного класу є критичною, оскільки цей клас суттєво менш чисельний, але має більшу практичну важливість.

З метою демонстрації переваг запропонованого методу на основі багаторівневої нейронної мережі здійснюється порівняння з базовим підходом TF-IDF + XGBoost. У базовій моделі кожне запитання представляється у вигляді високорозмірного вектора статистичних ознак (TF-IDF для уніграм і біграм), після чого застосовується градієнтний бустинг дерев рішень. Така модель не враховує послідовний порядок слів і локальні контекстні залежності, що обмежує її здатність виділяти тонкі маркери суб'єктивності. Натомість запропонована нейромережева архітектура за рахунок ембеддингів та рекурентних шарів явно моделює послідовність та контекст, що теоретично обґрунтовує її вищу ефективність у задачі класифікації суб'єктивних запитань.

У підсумку, метод класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі може бути формально описаний як композиція відображень

$$f_\theta = \psi \circ g_{\text{RNN}} \circ E \circ \phi, \tag{2.16}$$

де ϕ– процедура токенізації та приведення до фіксованої довжини, E– embedding-відображення, g_{RNN}– рекурентна (SimpleRNN або BiGRU) трансформація послідовності у вектор ознак, ψ– вихідний класифікаційний шар із сигмоїдною активацією. Навчання цієї композиції на розміченому корпусі забезпечує автоматичне виділення релевантних ознак суб'єктивності та побудову ефективного класифікатора запитань, який перевершує класичні векторні моделі за F1-мірою для цільового класу.

2.2 Модель класифікації суб'єктивних запитань на основі архітектури SimpleRNN

У першому варіанті запропонованого методу послідовний контекст запитання моделюється за допомогою одношарового рекурентного шару типу SimpleRNN. Дана архітектура розглядається як базова багаторівнева нейромережева модель, яка поєднує embedding-подання слів, рекурентну обробку послідовності та фінальний щільний класифікаційний шар. Таким чином, модель реалізує глибоку композицію відображень «текст → послідовність векторів → прихований стан → ймовірність суб'єктивності».

Формально, вхідне запитання q_iпісля токенізації та перетворення до фіксованої довжини Tзадається як послідовність індексів $\tilde{x}_i \in \{0, \dots, |V|\}^T$. Embedding-шар реалізує перехід від індексів токенів до щільного векторного простору розмірності dза допомогою матриці $W^{(e)} \in \mathbb{R}^{(|V|+1)\times d}$, у результаті чого отримуємо послідовність векторів

$$\mathrm{X}_i = (\mathrm{e}_{x_{i,1}}, \dots, \mathrm{e}_{x_{i,T}}), \mathrm{e}_{x_{i,t}} \in \mathbb{R}^d. \tag{2.17}$$

Цей рівень можна трактувати як навчувану операцію побудови ознак, що є спільною для всіх подальших запитань.

На наступному рівні застосовується одношаровий SimpleRNN з розмірністю прихованого стану k. Для кожного такту tрекурентний осередок приймає на вхід поточний вектор $\mathrm{e}_{x_{i,t}}$та попередній прихований стан h_{t-1}, формуючи новий стан h_t. Базові рівняння SimpleRNN мають вигляд

$$\mathrm{h}_t = \phi(W^{(h)}\mathrm{e}_{x_{i,t}} + U^{(h)}\mathrm{h}_{t-1} + \mathrm{b}^{(h)}), \tag{2.18}$$

де $W^{(h)} \in \mathbb{R}^{k\times d}$, $U^{(h)} \in \mathbb{R}^{k\times k}$– матриці ваг відповідно для входу та попереднього стану, $\mathrm{b}^{(h)} \in \mathbb{R}^{k}$– вектор зміщень, а $\phi(\cdot)$– нелінійна активаційна функція (зазвичай $\tanh$ або ReLU). Початковий стан h_0задається нульовим вектором, що відповідає відсутності попереднього контексту.

Після проходження всієї послідовності $(\mathrm{e}_{x_{i,1}}, \dots, \mathrm{e}_{x_{i,T}})$SimpleRNN-мережа формує фінальний прихований стан h_T, який інтерпретується як компактне, фіксованого розміру представлення всього запитання. Інтуїтивно h_Tакумулює інформацію про лексичні та контекстні патерни, що вказують на суб’єктивність або нещирість формулювання, включаючи наявність оцінних суджень, риторичних конструкцій, узагальнень та упереджень.

Для зменшення ризику перенавчання поверх вектора h_Tзастосовується Dropout-рівень з імовірністю занулення $p \in (0,1)$. На етапі навчання формується маска $\mathrm{m} \sim \mathrm{Bernoulli}(1-p)$, а на вхід наступного шару подається модифіковане представлення $\tilde{\mathrm{h}}_T = \mathrm{m} \odot \mathrm{h}_T$. Така регуляризація примушує мережу не покладатися на обмежений набір компонент прихованого стану й підвищує здатність моделі узагальнювати закономірності на нові, невидимі раніше запитання.

Вихідний шар моделі реалізує логістичну регресію над представленням $\tilde{\mathrm{h}}_T$. Обчислюється лінійна комбінація

$$s_i = \mathrm{w}^\top \tilde{\mathrm{h}}_T + b, \tag{2.19}$$

де $\mathrm{w} \in \mathbb{R}^k$, $b \in \mathbb{R}$– параметри вихідного шару, після чого до неї застосовується сигмоїдна активація:

$$\hat{p}_i = \sigma(s_i) = \frac{1}{1+e^{-s_i}}. \tag{2.20}$$

Отримане значення $\hat{p}_i \in (0,1)$інтерпретується як оцінка ймовірності того, що запитання q_iє суб'єктивним (клас 1). Після встановлення порогу τприймається дискретне рішення $\hat{y}_i \in \{0,1\}$.

Навчання параметрів шару embedding, матриць SimpleRNN і вихідного шару здійснюється спільно, методом мінімізації функції бінарної крос-ентропії

$$\mathcal{L}(\theta) = -\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} [y_i \log\, \hat{p}_i + (1-y_i)\log\,(1-\hat{p}_i)] \tag{2.21}$$

за допомогою оптимізатора Adam. На практиці використовується міні-батчова схема оновлення параметрів, що забезпечує стабільну збіжність навіть на великих корпусах запитань. Ключовою перевагою є те, що градієнти поширюються через усі рівні мережі, дозволяючи одночасно уточнювати і словникові представлення, і структуру рекурентного шару.

З теоретичної точки зору SimpleRNN є найбільш базовою формою рекурентної нейронної мережі, для якої доведено універсальну апроксимувальну здатність щодо послідовних функцій. Проте через відсутність механізмів контролю потоку градієнта (як у LSTM чи GRU) ця архітектура схильна до проблеми зникання або вибуху градієнтів на довгих послідовностях. У контексті класифікації запитань, де середня довжина тексту є відносно невеликою, цей недолік не є критичним, а натомість простота моделі дозволяє використовувати її як добре керований та інтерпретований базовий варіант багаторівневої нейромережевої архітектури.

Особливість задачі виявлення суб’єктивності полягає у тому, що ключові маркери часто розташовані в обмеженому контекстному вікні (наприклад, використання загальних оціночних прикметників, звинувачувальних конструкцій, риторичних запитань), тому SimpleRNN здатна ефективно «зчитувати» ці патерни без необхідності тримати в пам’яті дуже довгі залежності. Водночас архітектура враховує порядок слів, на відміну від класичного TF-IDF-подання, де інформація про послідовність втрачається.

У запропонованому методі модель на основі SimpleRNN виконує подвійну роль. По-перше, вона виступає як референсна глибинна архітектура, що демонструє, який приріст якості дає перехід від статичного векторного подання (TF-IDF) до послідовної обробки тексту. По-друге, вона є концептуальною основою для порівняння з більш складною архітектурою BiGRU, що дозволяє чітко продемонструвати, як двонапрямленість і гейтовані механізми впливають на здатність моделі розрізняти суб’єктивні запитання.

З позицій теорії глибинного навчання, описана SimpleRNN-конфігурація реалізує багаторівневу модель: рівень embedding, рекурентний рівень і щільний класифікаційний рівень із нелінійністю. Композиція цих перетворень дозволяє автоматично виділяти багатовимірні латентні ознаки, релевантні саме до задачі суб’єктивності, без ручного конструювання індикаторів чи шаблонів, що особливо важливо для гібридних і завуальованих форм нещирості або упередженості.

2.3 Модель класифікації суб’єктивних запитань на основі двонапрямленої GRU (BiGRU)

Друга реалізація запропонованого методу використовує двонапрямлену архітектуру на основі gated recurrent unit (GRU). На відміну від SimpleRNN, GRU-осередки містять гейтовані механізми, які керують запам’ятовуванням та «стиранням» інформації, що дозволяє суттєво полегшити проблему зникання

градієнтів та ефективніше моделювати довші контекстні залежності. Двонапрямлене розташування шарів (Bidirectional) додатково дає змогу враховувати інформацію як з початку, так і з кінця запитання.

Подібно до попередньої архітектури, вхідний текст q_iпроходить через етапи токенізації, індексації та паддінгу до сталої довжини T, а потім відображається embedding-шаром у послідовність векторів $\mathrm{X}_i = (\mathrm{e}_{x_{i,1}}, \dots, \mathrm{e}_{x_{i,T}})$. На цьому рівні обидві моделі (SimpleRNN та BiGRU) мають однакову структуру й відрізняються саме типом рекурентного шару.

В основі GRU лежать два гейти: гейт оновлення z_tта гейт скидання r_t. Для кожного такту tобчислення виконується за формулами:

$$\begin{aligned} \mathrm{z}_t &= \sigma(W^{(z)}\mathrm{e}_{x_{i,t}} + U^{(z)}\mathrm{h}_{t-1} + \mathrm{b}^{(z)}), \\ \mathrm{r}_t &= \sigma(W^{(r)}\mathrm{e}_{x_{i,t}} + U^{(r)}\mathrm{h}_{t-1} + \mathrm{b}^{(r)}), \\ \tilde{\mathrm{h}}_t &= \tanh\,(W^{(h)}\mathrm{e}_{x_{i,t}} + U^{(h)}(\mathrm{r}_t \odot \mathrm{h}_{t-1}) + \mathrm{b}^{(h)}), \\ \mathrm{h}_t &= (1 - \mathrm{z}_t) \odot \mathrm{h}_{t-1} + \mathrm{z}_t \odot \tilde{\mathrm{h}}_t, \end{aligned} \tag{2.22}$$

де $\sigma(\cdot)$– сигмоїдна функція, $W^{(\cdot)}$, $U^{(\cdot)}$– матриці ваг, $\mathrm{b}^{(\cdot)}$– вектори зміщень, $\odot$– поелементне множення. Така структура дає мережі можливість згортати історію прихованого стану у компактне представлення, вибірково оновлюючи або зберігаючи інформацію про попередні токени.

Для двонапрямленої конфігурації BiGRU використовуються два незалежні GRU-шари: прямий, що обробляє послідовність $(\mathrm{e}_{x_{i,1}}, \dots, \mathrm{e}_{x_{i,T}})$у звичайному порядку, та зворотний, який проходить послідовність у зворотному порядку $(\mathrm{e}_{x_{i,T}}, \dots, \mathrm{e}_{x_{i,1}})$. На кожному кроці утворюються два вектори прихованих станів $\overrightarrow{\mathrm{h}}_t$і $\overleftarrow{\mathrm{h}}_t$, які конкатенуються:

$$\mathrm{h}_t^* = [\overrightarrow{\mathrm{h}}_t; \overleftarrow{\mathrm{h}}_t] \in \mathbb{R}^{2k}. \tag{2.23}$$

Як документне представлення використовується фінальний агрегований стан h_T^*, який одночасно містить інформацію про контекст «зліва направо» та «справа наліво».

З точки зору задачі класифікації суб'єктивних запитань це означає, що модель BiGRU здатна враховувати не лише те, як певний токен впливає на наступні, але й те, як майбутні лексеми змінюють інтерпретацію попередніх. Наприклад, оцінне слово або узагальнення, розташоване ближче до кінця запитання, може ретроспективно змінювати тон попередньої нейтральної частини. Двонапрямлене проходження дозволяє захоплювати подібні феномени більш повно, ніж односпрямована SimpleRNN.

Як і у випадку SimpleRNN, поверх вектора h_T^*застосовується Dropout-рівень, що зануляє випадкову підмножину компонент під час навчання. Далі послідовність обробки аналогічна: лінійне перетворення h_T^*слабко параметризованим щільним шаром, сигмоїдна активація та перетворення ймовірності у бінарне рішення за порогом τ. При цьому завдяки удвічі більшій розмірності векторного подання (конкатенація прямого та зворотного станів) BiGRU має більшу виразну здатність порівняно з одношаровим SimpleRNN за тієї ж кількості нейронів у кожній гілці.

Навчання параметрів BiGRU здійснюється аналогічно, методом мінімізації бінарної крос-ентропії за допомогою оптимізатора Adam. При цьому гейтовані механізми GRU, на відміну від «чистої» SimpleRNN, дозволяють стабільніше поширювати градієнти через велику кількість часових кроків, що робить модель менш вразливою до зникання градієнтів. Це особливо важливо при обробці тих запитань, де інформація, важлива для класифікації, розподілена по всій довжині тексту, а не зосереджена в одному фрагменті.

З теоретичної точки зору можна розглядати BiGRU як більш потужне узагальнення SimpleRNN, яке наближається за властивостями до LSTM, але має меншу кількість параметрів завдяки спрощеній гейт-структурі. Використання двонапрямленого варіанта робить архітектуру особливо привабливою для задач,

де важливі як ліво-, так і правоконтекстні залежності, тобто фактично для більшості завдань аналізу природномовних текстів.

У контексті даної роботи BiGRU виступає як покращена багаторівнева нейронна архітектура, покликана дослідити, чи забезпечує додаткова виразна здатність та врахування двонапрямленого контексту істотне покращення якості класифікації суб'єктивних запитань порівняно з базовою SimpleRNN. У поєднанні з embedding-шаром і регуляризацією Dropout BiGRU формує глибоку модель із кількох послідовних рівнів нелінійних перетворень, яка автоматично витягує лінгвістичні ознаки, релевантні до виявлення упереджених, агресивних або маніпулятивних формулювань.

Важливим аспектом є те, що, попри більшу складність, BiGRU залишається відносно компактною за кількістю параметрів і придатною для навчання на великих корпусах без надмірної потреби в обчислювальних ресурсах. Це робить її практично застосовною в умовах, коли класифікація суб'єктивних запитань має виконуватися в режимі наближеного реального часу (наприклад, модерація контенту в онлайн-платформах запитань-відповідей).

Таким чином, модель на основі двонапрямленої GRU логічно доповнює базову SimpleRNN-архітектуру в рамках запропонованого методу класифікації суб'єктивних запитань на основі багаторівневої нейронної мережі. Вона реалізує більш гнучкий механізм моделювання контексту, краще пристосований до виявлення складних патернів суб'єктивності, і виступає важливим елементом порівняльного аналізу ефективності різних типів рекурентних архітектур у цій задачі.

Висновки до розділу 2

У другому розділі формально описано задачу класифікації суб'єктивних запитань як бінарну задачу машинного навчання, де кожне запитання розглядається як послідовність токенів із подальшим перетворенням у фіксовану

за довжиною числову репрезентацію. Запропоновано послідовність етапів: токенізація й побудова словника обмеженого розміру, кодування слів індексами, паддінг/обрізання послідовностей, перехід до щільних векторів за допомогою embedding-шару, рекурентне моделювання послідовного контексту та фінальна бінарна класифікація на основі сигмоїдного вихідного шару. Навчання здійснюється шляхом мінімізації бінарної крос-ентропії з використанням оптимізатора Adam, а якість оцінюється за метриками accuracy та F1-міри з фокусом на «інсинсирному» класі.

Як перший варіант багаторівневої архітектури розглянуто модель на основі SimpleRNN, де вхідні ембедінги послідовно обробляються одношаровим рекурентним осередком із фіксованою розмірністю прихованого стану. Показано, що така конфігурація забезпечує компактне векторне представлення запитання, яке акумулює інформацію про його лексичні й контекстуальні патерни, та у поєднанні з Dropout-регуляризацією дозволяє зменшити перенавчання. SimpleRNN-модель розглядається як базова глибинна архітектура, яка вже враховує послідовний порядок слів, але має обмежену здатність моделювати довгострокові залежності через проблему зникання градієнтів.

Другий варіант реалізації методу ґрунтується на двонапрямленій архітектурі GRU (BiGRU), яка дозволяє одночасно враховувати як попередній, так і наступний контекст для кожного слова запитання. Використання гейтингових механізмів GRU дає змогу краще контролювати потік інформації в часі й зменшити втрати важливих сигналів при зростанні довжини послідовності. Бінапрямлене оброблення послідовності забезпечує більш повне захоплення риторичної структури й завуальованих маркерів інсинсирності, а поєднання двох напрямів у конкатенованому векторі ознак підвищує дискримінативну здатність моделі порівняно з односпрямованими RNN.

Узагальнюючи, у розділі запропоновано цілісний метод класифікації суб’єктивних запитань, який реалізується як композиція процедур попередньої обробки тексту, навчуваного переходу до векторних представлень та багаторівневого рекурентного моделювання. Базова модель TF-IDF + XGBoost

використовується як порівняльний орієнтир, тоді як архітектури SimpleRNN і BiGRU формують основні рівні глибинної нейронної мережі, відповідальні за захоплення послідовних і контекстуальних залежностей. Такий підхід створює передумови для досягнення вищих значень F1-міри для інсинсирного класу та подальшого експериментального підтвердження переваг запропонованого методу над класичними векторними підходами.

3 РЕАЛІЗАЦІЯ МЕТОДУ КЛАСИФІКАЦІЇ СУБ'ЄКТИВНИХ ЗАПИТАНЬ НА ОСНОВІ БАГАТОРІВНЕВОЇ НЕЙРОННОЇ МЕРЕЖІ

3.1 Підготовка корпусу запитань та первинний аналіз даних

Початковим етапом реалізації методу класифікації суб'єктивних запитань є формування репрезентативного корпусу текстових даних. У роботі використано відкритий набір Quora Insincere Questions Classification, який містить 1 306 122 навчальних прикладів і 375 806 прикладів тестового набору. Кожен запис включає ідентифікатор запитання qid, текст запитання англійською мовою (question_text) та бінарну цільову змінну target, що визначає, чи є запитання «несумлінним» (інсинсирним, провокативним, риторичним) або нейтральним/інформативним.

З погляду задачі класифікації суб'єктивних запитань такий корпус є природним полігоном для моделювання, оскільки інсинсирні запитання містять виражений компонент суб'єктивної оцінки або упередженого ставлення, тоді як решта запитань можна трактувати як умовно об'єктивні. Отже, змінна target інтерпретується як індикатор суб'єктивності висловлювання, що узгоджується з загальною постановкою задачі побудови методу виявлення суб'єктивних запитань.

Розподіл цільових класів у навчальній вибірці показано на гістограмі (рисунок 3.1). Видно, що корпус має суттєвий дисбаланс: частка класу 0 (сумлінні/об'єктивні запитання) становить 1 225 312 прикладів, або 93,81 %, тоді як клас 1 (суб'єктивні/інсинсирні запитання) охоплює лише 80 810 прикладів, тобто 6,19 %. Такий дисбаланс є типовим для реальних задач модерації контенту й безпосередньо впливає на вибір метрик та стратегію навчання моделі.

Наявний перекіс у бік «норми» означає, що наївний класифікатор, який завжди прогнозує клас 0, уже досягав би точності понад 93 %. Тому проста метрика accuracy не є інформативною; для оцінювання якості багаторівневої нейронної мережі доцільно використовувати F1-міру для міноритарного класу, а

також повний класифікаційний звіт з precision та recall. У реалізації методу ці міркування враховано під час інтерпретації результатів як для базового підходу TF-IDF+XGBoost, так і для нейронних моделей SimpleRNN та BiGRU.

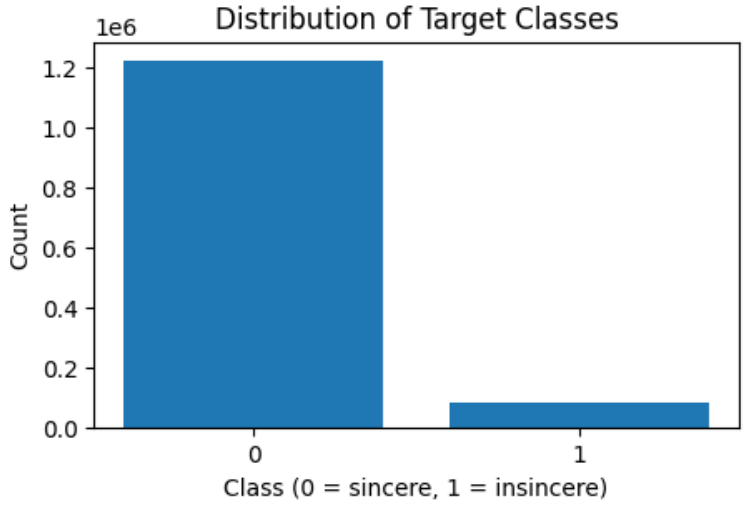

Рисунок 3.1 – Розподіл цільових класів у навчальній вибірці корпусу Quora (0 – сумлінні запитання, 1 – інсинсирні/суб'єктивні запитання).

Додатково проведено аналіз довжин запитань у словах, результат якого наведено у вигляді коробкової діаграми (рисунок 3.2). Для кожного запитання обчислювалася кількість токенів, отриманих при розбитті тексту за пробілами. Статистичний опис показує, що медіана довжини дорівнює 11 словам, міжквартильний інтервал лежить у межах від 8 до 15 слів, а середнє значення становить приблизно 12,8 слова. Лише одиничні запитання сягають довжини понад 60–70 токенів, а максимум становить 134 слова.

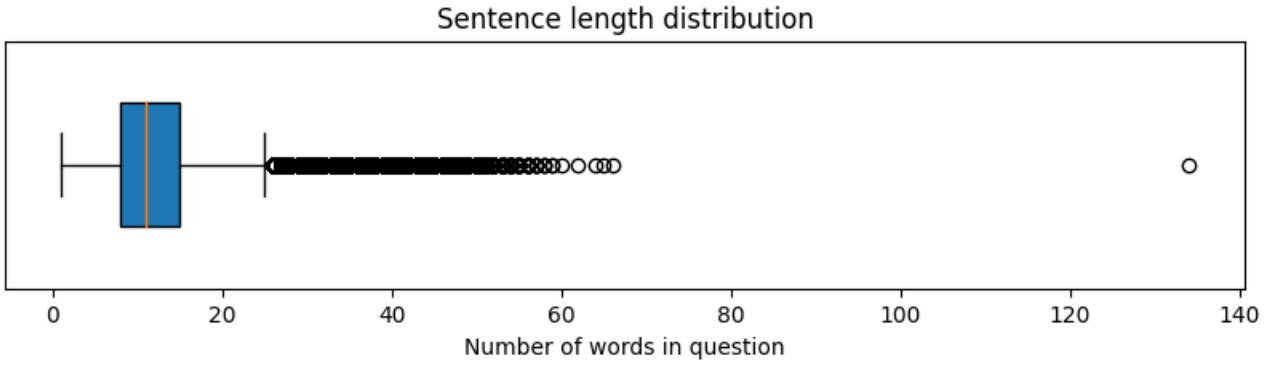

Рисунок 3.2 – Розподіл довжин запитань за кількістю слів у корпусі Quora (коробкова діаграма).

Отриманий розподіл довжин запитань безпосередньо вплинув на вибір гіперпараметра максимальної довжини послідовності max_len для рекурентних моделей. Значення max_len = 50 дає змогу повністю покривати переважну більшість запитань (понад 95 % корпусу) без зайвої надмірності, водночас скорочуючи дуже довгі запитання шляхом обрізання «хвоста». Таким чином досягається компроміс між збереженням контексту та ефективністю обчислень.

Обрізання довших послідовностей до 50 токенів інтерпретується як припущення, що ключові маркери суб'єктивності та інсинсирності зосереджені у початковій частині запитання (формулювання проблеми, оціночні епітети, упереджені формулювання). Такий підхід є виправданим з погляду лінгвістичного аналізу, оскільки саме початок запитання визначає його прагматичний намір.

Після первинного аналізу корпус було поділено на тренувальну та валідаційну підмножини в пропорції 80 % до 20 % зі стратифікацією за цільовим класом. Отримано 1 044 897 навчальних і 261 225 валідаційних прикладів, причому в обох підмножинах збережено вихідний відсотковий розподіл класів (93,8 % до 6,2 %). Така стратифікація забезпечує коректне порівняння моделей і запобігає зміщенню оцінок якості через випадкові флуктуації частки рідкісного класу.

Важливим аспектом реалізації є масштаб корпусу: понад один мільйон навчальних прикладів дозволяє тренувати глибокі моделі без ризику критичного перенавчання вже на перших епохах, однак підвищує вимоги до обчислювальних ресурсів. У роботі всі експерименти проводяться в середовищі TensorFlow 2.18.0 з використанням GPU-прискорення, що дає змогу підтримувати прийнятний час навчання навіть для двонапрямлених рекурентних мереж.

На завершення етапу підготовки даних організовано єдиний конвеєр попередньої обробки, який включає токенізацію, побудову словника розмірністю 50 000 найчастотніших токенів, заміну рідкісних слів спеціальним маркером

<OOV> та заповнення послідовностей до фіксованої довжини max_len. Цей конвеєр застосовується однаково до тренувальної, валідаційної й тестової вибірок, що забезпечує узгодженість репрезентації текстів на всіх етапах багаторівневого методу класифікації суб'єктивних запитань.

Сформований таким чином корпус з чітко задокументованими статистичними характеристиками (див. рисунки 3.1–3.2) слугує основою для побудови базового та нейронних класифікаторів, що детально розглядаються у наступних підрозділах.

3.2 Реалізація базового підходу TF-IDF + XGBoost і одношарового рекурентного рівня SimpleRNN

Перед переходом до багаторівневих рекурентних архітектур доцільно побудувати еталонний (baseline) класифікатор, який використовує класичні векторні представлення тексту. У реалізації методу як базову модель застосовано комбінацію TF-IDF-векторизатора та градієнного бустингу XGBoost. Такий підхід ігнорує послідовний порядок слів, проте добре захоплює частотні та n-грамні патерни, що дозволяє оцінити, яку додаткову вигоду дає врахування послідовності в рекурентних мережах.

Конвеєр базової моделі складається з двох основних кроків. Спочатку для кожного запитання виконується токенізація, видалення англомовних стоп-слів та стемінг за алгоритмом Snowball. На другому кроці формується матриця ознак розмірністю 50 000, яка містить уніграми та біграми з TF-IDF-зважуванням. Отримані вектори подаються на вхід класифікатора XGBoost з налаштуванням scale_pos_weight, що компенсує дисбаланс класів шляхом підвищення ваги помилок у міноритарному класі.

Результати базової моделі на валідаційній вибірці демонструють точність 0,885 та F1-міру для класу «суб'єктивних» запитань 0,439. Класифікатор досягає дуже високої точності для класу 0 (precision 0,98, recall 0,90), але працює менш

стабільно для класу 1: хоча recall сягає 0,73, precision залишається на рівні 0,31, що означає значну кількість хибно позитивних спрацьовувань. Ці обмеження слугують мотивацією для використання моделей, здатних враховувати порядок слів і контекст.

Наступним рівнем запропонованого багаторівневого методу є одношарова рекурентна нейронна мережа SimpleRNN з вбудованим шаром векторних уявлень слів. На вході моделі використовується шар Embedding розмірності 50 000 × 50, який перетворює індекси токенів у щільні 50-вимірні вектори. Таким чином, кожне запитання представлено матрицею розміру 50 × 50, де перша координата відповідає позиції слова у запитанні, друга — його семантичній проєкції в embedding-простір.

Над послідовністю векторів працює шар SimpleRNN з 64 прихованими станами. На відміну від статичних моделей, цей шар ітеративно обробляє токени зліва направо, акумулюючи інформацію про вже прочитаний контекст у прихованому векторі стану. Після обробки останнього токена формується фінальний стан мережі, який подається на повнозв’язний шар з одним нейроном та сигмоїдальною активацією, що повертає оцінку ймовірності приналежності запитання до класу 1. Для зниження перенавчання між рекурентним і вихідним шаром вставлено шар Dropout з коефіцієнтом 0,3.

Навчання мережі здійснюється методом стохастичного градієнтного спуску з оптимізатором Adam, функцією втрат binary_crossentropy, розміром батчу 256 та числом епох 10. Така конфігурація обрана як компроміс між часом навчання та можливістю простежити динаміку збіжності моделі. Повна історія навчання SimpleRNN представлена на рисунку 3.3.

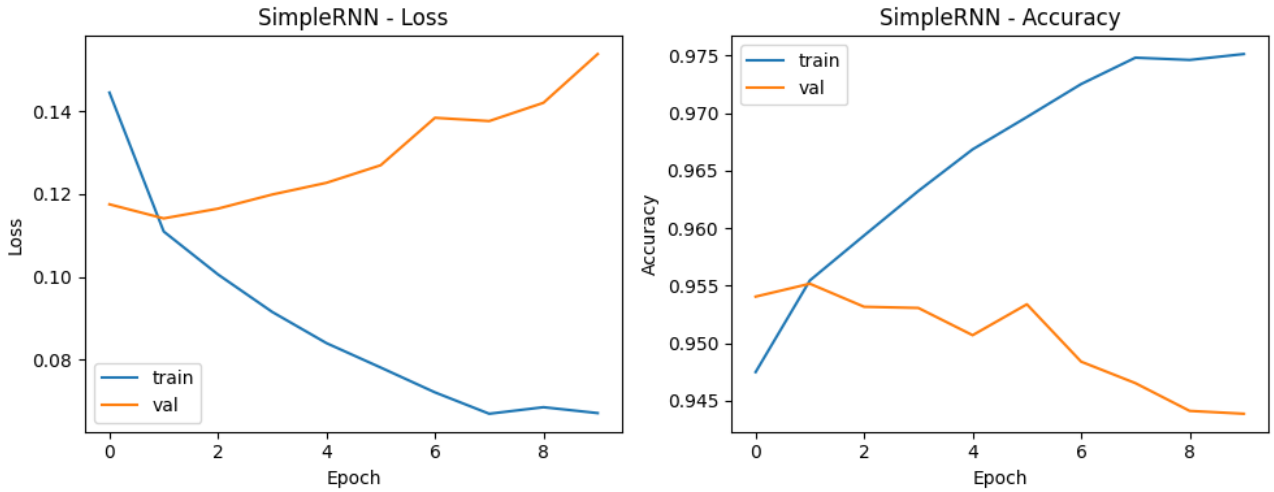

Рисунок 3.3 – Динаміка значень функції втрат (a) та точності (b) для моделі SimpleRNN на тренувальній та валідаційній вибірках.

Крива тренувальної втрати на лівому графіку рисунка 3.3 демонструє монотонне спадання від 0,18 до 0,06, що свідчить про стабільне засвоєння патернів корпусу мережею. Натомість валідаційна втрата досягає мінімуму в районі другої–третьої епохи (≈0,114–0,116), після чого починає зростати, що є ознакою поступового перенавчання. Це підтверджує необхідність використання механізмів ранньої зупинки або збільшення регуляризації, якщо метою є максимальна узагальнювальна здатність.

Графік точності (див. рисунок 3.3, праворуч) показує, що навчальна точність швидко зростає до 0,97–0,98, тоді як валідаційна точність стабілізується на рівні ≈0,955 на перших епохах і злегка знижується до ≈0,944 до кінця навчання. Така різниця між навчальною й валідаційною точністю також вказує на певний ступінь перенавчання, проте в межах, прийнятних для великих корпоративних наборів даних.

Після завершення навчання модель SimpleRNN досягає на валідаційній вибірці точності 0,944 та F1-міри 0,575 для класу 1. У порівнянні з базовою моделлю TF-IDF+XGBoost F1-міра збільшується більш ніж на 0,13, що свідчить про відчутне покращення балансу між precision і recall для суб’єктивних запитань. Зокрема, SimpleRNN забезпечує достатньо високу точність (precision

≈0,54) та recall ≈0,61, тоді як базовий підхід хоч і виявляє більшу частку суб'єктивних запитань, але генерує значно більше хибно позитивних сигналів.

Отримані результати підтверджують, що навіть відносно проста рекурентна архітектура, яка враховує послідовний порядок слів, суттєво покращує якість виявлення суб'єктивних запитань порівняно з класичними методами на основі «мішка слів». У подальшому рівні методу послідовність обробляється більш потужною двонапрямленою GRU-мережею, яка покликана використати контекст як у прямому, так і в зворотному напрямках.

3.3 Реалізація двонапрямленого рівня GRU (BiGRU) та модулів прогнозування

Третій рівень запропонованого методу класифікації суб'єктивних запитань реалізує більш складну рекурентну архітектуру — двонапрямлену мережу на основі Gated Recurrent Unit (BiGRU). На відміну від SimpleRNN, яка обробляє послідовність лише зліва направо, BiGRU одночасно проходить запитання у прямому та зворотному напрямках, що дозволяє враховувати як «ліворучний», так і «праворучний» контекст для кожної позиції. Це особливо важливо для коротких запитань, де важливі маркери суб'єктивності можуть розміщуватися як на початку, так і наприкінці речення.

Архітектура BiGRU зберігає шар Embedding з параметрами, ідентичними до SimpleRNN (словник 50 000 слів, розмірність векторів 50), тому порівняння моделей відбувається за однакових вхідних умов. Поверх embeddings розташовано двонапрямлений шар GRU з 64 прихованими одиницями у кожному напрямку, внаслідок чого під час конкатенації формується 128-вимірне представлення запитання. Далі, як і в попередній моделі, використовується шар Dropout (0,3) та вихідний повнозв'язний шар з сигмоїдальною активацією для бінарної класифікації.

Навчання BiGRU-мережі здійснюється з такими ж гіперпараметрами, як і для SimpleRNN (10 епох, розмір батчу 256, оптимізатор Adam), що забезпечує коректну порівнюваність результатів. Однак через більш високу обчислювальну складність двонапрямленого шару час однієї епохи навчання значно зростає (≈500 с проти ≈112 с для SimpleRNN). Це демонструє типовий компроміс між точністю моделі та витратами обчислювальних ресурсів.

Динаміку навчання BiGRU показано на рисунку 3.4. На графіку втрат (ліворуч) тренувальна крива монотонно спадає від 0,14 до 0,028, що свідчить про сильну здатність моделі до апроксимації навчальних даних. Валідаційна втрата досягає мінімуму приблизно на другій–третій епосі (0,11–0,113), після чого демонструє чітке зростання, перевищуючи значення 0,24 наприкінці навчання. Така поведінка є більш вираженим проявом перенавчання порівняно з SimpleRNN і свідчить про необхідність застосування ранньої зупинки або зменшення числа епох у практичних налаштуваннях.

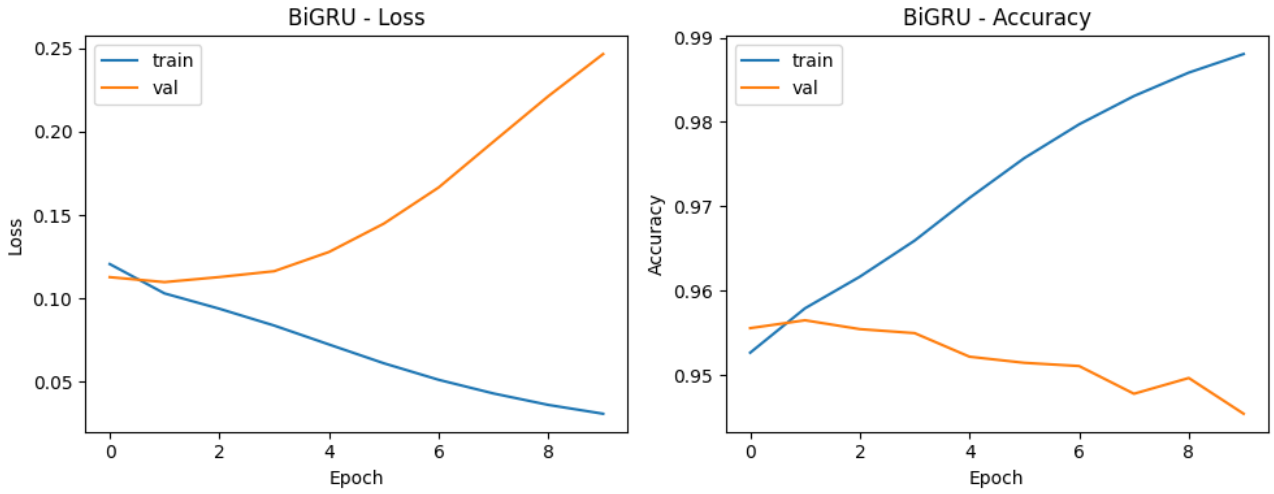

Рисунок 3.4 – Динаміка значень функції втрат (a) та точності (b) для двонапрямленої моделі BiGRU на тренувальній та валідаційній вибірках.

Графік точності (див. рисунок 3.4, праворуч) показує, що навчальна точність BiGRU швидко наближається до 0,99, тоді як валідаційна точність стабілізується в діапазоні 0,955–0,960 на перших епохах, а потім дещо знижується до ≈0,945. На відміну від SimpleRNN, де розрив між тренувальною

та валідаційною точністю є менш вираженим, у BiGRU спостерігається суттєвіший «generalization gap», що підтверджує більшу модельну потужність і відповідно вищу схильність до перенавчання.

Підсумкова якість BiGRU на валідаційній вибірці характеризується точністю 0,945 і F1-мірою 0,560 для класу 1. Таким чином, за точністю BiGRU незначно переважає SimpleRNN (0,945 проти 0,944), однак поступається йому за F1-мірою (0,560 проти 0,575). Класифікаційний звіт показує майже симетричні значення precision і recall для міноритарного класу (обидва ≈0,56), що означає більш збалансоване співвідношення хибно позитивних та хибно негативних рішень, але водночас деяке зниження загальної здатності моделі виділяти суб’єктивні запитання порівняно з SimpleRNN.

На етапі прогнозування на повному тестовому наборі BiGRU використовується як основна модель. Після токенізації та паддингу тестових запитань мережа генерує оцінки ймовірностей належності до класу 1 для кожного з 375 806 прикладів. Подальше бінарне рішення приймається за порогом 0,5. Розподіл отриманих бінарних прогнозів подано на рисунку 3.5. Видно, що частка інсинсирних запитань у тестових прогнозах становить 6,21 %, тоді як вихідний розподіл у навчальних даних дорівнював 6,19 %.

Висока подібність розподілів класів між тренувальною вибіркою та тестовими прогнозами (див. рисунок 3.5) інтерпретується як непряма ознака коректної калібровки моделі: BiGRU не демонструє систематичного зміщення в бік будь-якого класу й приблизно відтворює статистику корпусу. Це важливо для подальшого використання моделі в реальних системах модерації, де надмірний перекіс у бік «підозрілих» запитань може призвести до зростання кількості хибних спрацювань, а надто «обережна» модель пропускатиме потенційно проблемний контент.

Рисунок 3.5 – Порівняння розподілу класів у навчальній вибірці та серед тестових прогнозів моделі BiGRU.

Детальнішу картину поведінки моделі дає гістограма розподілу прогнозованих імовірностей для класу 1 (рис. 3.6). Більшість запитань отримує імовірності, близькі до нуля, що відповідає інтуїції про домінування нейтральних запитань. Водночас виділяється компактний кластер запитань з імовірностями, близькими до одиниці, які модель класифікує як чітко суб'єктивні або провокативні. Невелика кількість запитань має проміжні значення імовірностей, що може вказувати на лінгвістично неоднозначні або контекстуально складні випадки.

Таке «бімодальне» розташування імовірностей (рис. 3.6) свідчить про високу впевненість моделі в більшості рішень, але одночасно вказує на потенційну потребу в калібруванні (наприклад, за допомогою Platt scaling або ізотонічної регресії), якщо модель використовуватиметься для прийняття ризик-чутливих рішень. У контексті класифікації суб'єктивних запитань ця властивість

є скоріше перевагою, оскільки дозволяє явно виділити групу «найбільш суб'єктивних» запитань за високими значеннями вихідної ймовірності.

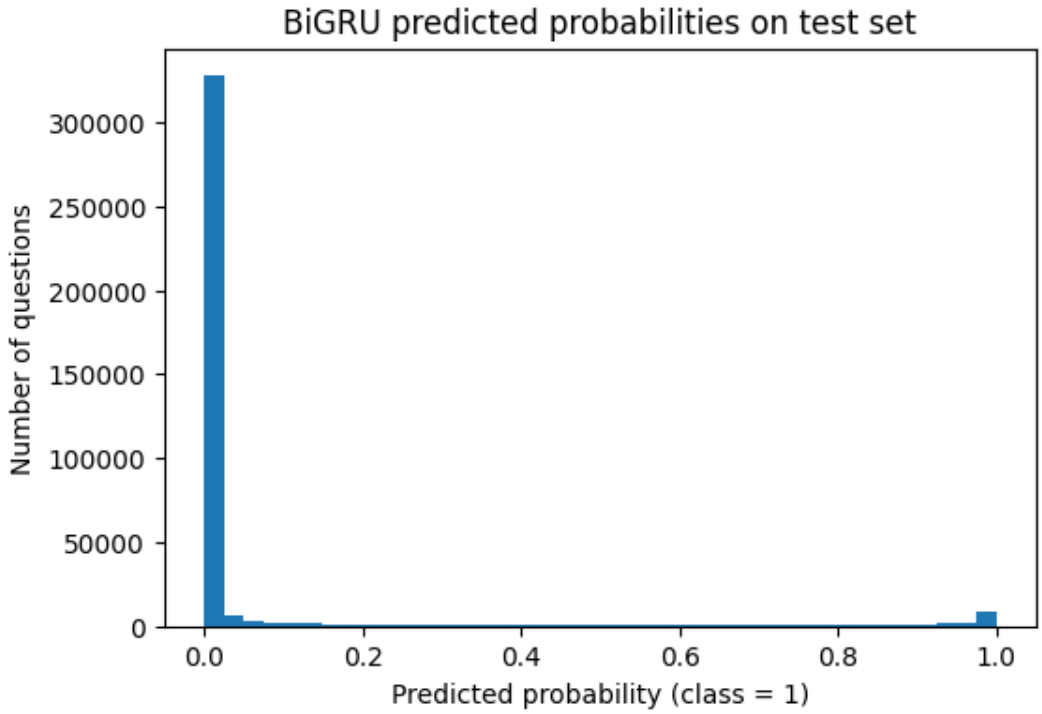

Рисунок 3.6 – Розподіл передбачених імовірностей належності запитань до класу 1 (інсинсирні/суб'єктивні) для тестового набору за результатами моделі BiGRU.

Загалом, результати реалізації двонапрямленого рівня GRU показують, що збільшення модельної складності не гарантує автоматичного зростання F1-міри для міноритарного класу, але забезпечує більш збалансований профіль помилок і коректне відтворення розподілу класів на тестових даних. У поєднанні з базовим підходом TF-IDF+XGBoost та одношаровою моделлю SimpleRNN BiGRU формує завершений багаторівневий метод класифікації суб'єктивних запитань, який поєднує переваги класичних векторних ознак, послідовних рекурентних представлень та двонапрямленого контексту.

Висновки до розділу 3

У третьому розділі здійснено практичну реалізацію запропонованого методу класифікації суб’єктивних запитань на основі багаторівневої нейронної мережі з використанням корпусу Quora Insincere Questions. На першому етапі продемонстровано побудову базової моделі TF-IDF + XGBoost, яка забезпечила точність близько 0.885 та F1-міру для інсинсирного класу на рівні ≈0.44. Ці результати підтверджують, що навіть потужні ансамблеві методи на основі розріджених векторних представлень мають обмежену здатність ураховувати послідовний контекст та риторичні особливості формулювань запитань, особливо за умов вираженого класового дисбалансу.

Наступним кроком розглянуто глибинну модель на основі архітектури Embedding + SimpleRNN. Навчання моделі на токенізованих та падованих послідовностях продемонструвало суттєве зростання якості: досягнуто точності ≈0.944 та F1-міри ≈0.575 для «проблемного» класу. Аналіз кривих навчання (залежностей loss та accuracy від епох) показав, що SimpleRNN ефективно засвоює патерни суб’єктивності та інсинсирності, однак із ростом кількості епох спостерігаються ознаки перенавчання – зниження валідаційної точності та зростання валідаційних втрат за стабільно високої тренувальної точності. Це вказує на необхідність використання додаткових регуляризаційних механізмів (рання зупинка, оптимальний підбір кількості епох).

Двонапрямлена модель на основі GRU (Bidirectional GRU) продемонструвала близькі до SimpleRNN, але дещо кращі глобальні показники точності (≈0.945) при F1-мірі ≈0.56 для інсинсирного класу. З одного боку, бінапрямлене рекурентне моделювання дозволило краще врахувати контекст з обох напрямків, що позитивно позначилося на класифікації складніших формулювань. З іншого боку, аналіз кривих навчання для BiGRU виявив більш виражене зростання валідаційних втрат на пізніх епохах, що свідчить про ще більшу схильність до перенавчання порівняно з SimpleRNN. Отже, додаткове налаштування гіперпараметрів (розмір прихованого шару, рівень Dropout,

кількість епох) є критично важливим для повної реалізації потенціалу цієї архітектури.

Додатковий аналіз розподілу передбачених міток на тестовому наборі показав, що модель BiGRU відтворює емпіричний розподіл класів, дуже близький до реального розподілу в тренувальних даних (≈93.8 % щирих проти ≈6.2 % інсинсирних запитань). Це свідчить про коректну роботу моделі в умовах незбалансованого корпусу, без явного «зриву» у бік одного з класів. Сукупність проведених експериментів дозволяє зробити висновок, що багаторівневий підхід із використанням ембедінгів та рекурентних архітектур (SimpleRNN, BiGRU) істотно перевершує базову TF-IDF + XGBoost-модель щодо виявлення інсинсирних запитань і може розглядатися як перспективний напрям для подальшого вдосконалення систем автоматичної модерації контенту.

ВИСНОВКИ

У результаті виконання кваліфікаційної роботи отримано такі основні висновки:

1. Проведено комплексний аналіз суб’єктивних та інсинсирних запитань у сучасних системах питань-відповідей. Показано, що такі запитання поєднують лінгвістичні (оцінна лексика, узагальнення, стереотипні позначення соціальних груп) та прагматичні маркери (навідні формулювання, риторичні запитання, імплікатурні смисли), що ускладнює їх автоматичну ідентифікацію. Обґрунтовано, що саме інсинсирні запитання є важливим фактором формування токсичних дискусій, поляризації спільнот та підриву довіри до онлайн-платформ.

2. Систематизовано наукові підходи до виявлення суб’єктивності, інсинсирності й токсичності текстів. Показано еволюцію методів від лексикон-орієнтованих правил і статистичних моделей (bag-of-words, n-грами, SVM, логістична регресія, XGBoost) до глибинних нейронних архітектур, що працюють із розподіленими векторними представленнями слів і контексту. Встановлено, що класичні методи є інтерпретованими та обчислювально відносно дешевими, однак недостатньо повно враховують послідовну структуру запитань і складні риторичні патерни, властиві інсинсирним формулюванням.

3. Проаналізовано можливості моделей глибинного навчання для класифікації запитань та охарактеризовано релевантні корпуси даних. Показано, що рекурентні мережі (SimpleRNN, LSTM, GRU, BiGRU), згорткові архітектури та трансформери здатні моделювати як локальні, так і глобальні залежності в тексті. Обґрунтовано вибір датасету Quora Insincere Questions Classification як репрезентативного корпусу для дослідження інсинсирних запитань, оскільки він містить понад 1,3 млн запитань із вираженим класовим дисбалансом і широким тематичним охопленням.

4. Спроєктовано та реалізовано базову модель класифікації інсинсирних запитань на основі TF-IDF і алгоритму XGBoost. Текст запитань було перетворено у високорозмірні TF-IDF-вектори (уніграм і біграм), після чого

побудовано ансамблевий класифікатор XGBoost. Експериментально показано, що така модель досягає точності близько 0.885 на валідаційній вибірці, при цьому F1-міра для інсинсирного класу становить близько 0.44. Базова модель є інтерпретованим орієнтиром якості, але демонструє обмежену здатність до виявлення складних контекстуальних та риторичних маркерів інсинсирності.

5. Розроблено багаторівневу нейронну архітектуру класифікації суб'єктивних запитань на основі шарів ембедингів та рекурентних моделей SimpleRNN і двонапрямленої GRU. На нижньому рівні реалізовано перехід від індексів токенів до щільних векторних представлень у просторі ембедингів, що забезпечило компактне подання лексики. Верхній рівень формує послідовнісне подання запитання за допомогою SimpleRNN або Bidirectional GRU, що дозволяє моделі враховувати порядок слів, довгострокові залежності та контекст у двох напрямах. Додатково застосовано шар Dropout для регуляризації та вихідний щільний шар із сигмоїдною активацією для отримання ймовірнісної оцінки класу.

6. Здійснено експериментальне порівняння базової моделі TF-IDF + XGBoost із рекурентними архітектурами SimpleRNN та BiGRU. Показано, що SimpleRNN досягає точності близько 0.944 і F1-міри для інсинсирного класу на рівні ≈0.575, тоді як двонапрямлена GRU забезпечує точність близько 0.945 і F1-міру ≈0.560. Таким чином, обидві рекурентні моделі суттєво перевершують базовий підхід за F1-мірою для рідкісного інсинсирного класу (приріст орієнтовно на 0.12–0.14), зберігаючи при цьому високу загальну точність класифікації.

7. Проведено аналіз впливу архітектури моделі та обчислювальних витрат на практичну придатність методу. Встановлено, що модель BiGRU дещо покращує загальну точність порівняно з SimpleRNN, однак потребує значно більших обчислювальних ресурсів і часу навчання. SimpleRNN, навпаки, забезпечує найвищу F1-міру для інсинсирного класу за помірного часу тренування, що робить її привабливою як «робочу» модель у багаторівневій

архітектурі, особливо для систем із обмеженими ресурсами або жорсткими вимогами до латентності.

8. Сформульовано практичні рекомендації та напрямки подальшого розвитку запропонованого методу. Запропоновано розглядати багаторівневу архітектуру як гнучкий каркас, у якому базова статистична модель (TF-IDF + XGBoost) може використовуватися для швидкого попереднього фільтрування, а рекурентні мережі SimpleRNN/BiGRU – як точніший рівень для детекції проблемних запитань. Окреслено перспективи розширення моделі за рахунок механізмів уваги, адаптації до багатомовних корпусів, інтеграції позатекстових ознак (метаданих користувача, реакцій спільноти) та перенесення отриманих результатів на україномовні платформи й освітні сервіси.

У сукупності отримані результати підтверджують, що поставлена у вступі мета – розроблення та дослідження методу класифікації суб’єктивних запитань на основі багаторівневої нейронної мережі з використанням SimpleRNN та BiGRU – досягнута, а всі сформульовані завдання кваліфікаційної роботи реалізовано та логічно узгоджено між собою.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Mishra, A., & Jain, S. (2016). A survey on question answering systems with classification. *Journal of King Saud University – Computer and Information Sciences, 28*(3), 345–361. ScienceDirect

2. Chen, L., Li, Y., Tay, Y., & Kan, M.-Y. (2012). Understanding user intent in community question answering. In *Proceedings of WWW 2012 Companion* (pp. 823–828). archives.iw3c2.org

3. Zhou, T. C., Luan, H., & Sun, M. (2012). A data-driven approach to question subjectivity for asker-intent interpretation. In *Proceedings of AAAI* (pp. 1370–1376). ojs.aaai.org+1

4. Palshikar, G. K., & Apte, M. (2016). Learning to identify subjective sentences. In *Proceedings of the WASSA Workshop* (pp. 211–220). aclanthology.org

5. Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool.

6. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

7. Quora. (2019). Quora Insincere Questions Classification (Kaggle competition description). Kaggle. Medium+1

8. Chakraborty, S., Kundu, S., & Das, D. (2022). Quora insincere questions classification using attention based model. *International Journal of Advanced Computer Science and Applications, 13*(4), 495–505. researchonline.ljmu.ac.uk

9. Malik, J. S., Ahmad, T., & Iqbal, F. (2024). Deep learning for hate speech detection: A comparative study. *Online Social Networks and Media, 38*, 100243. smusg.elsevierpure.com

10. Gudumotu, C. E., & colleagues. (2023). A survey on deep learning models to detect hate speech on social media. In *Recent Advances in Computational Intelligence* (pp. 23–45). Springer. SpringerLink

11. Al Faraby, S., Rahman, M., & Rahman, M. S. (2024). Analysis of large language models for educational question classification. *Patterns, 5*(6), 100971. ScienceDirect

12. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186).

13. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of EMNLP 2014* (pp. 1724–1734).

14. Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing, 45*(11), 2673–2681.

15. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

16. Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of EMNLP 2014* (pp. 1746–1751).

17. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

18. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the ACL, 5*, 135–146.

19. Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In *Proceedings of EACL 2017* (pp. 427–431).

20. Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems 28* (pp. 649–657).

21. Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of ICWSM 2017* (pp. 512–515).

22. Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). Deep learning for hate speech detection in tweets. In *Proceedings of WWW 2017 Companion* (pp. 759–760).

23. Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys, 51*(4), 85.

24. Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of NAACL Workshop on Abusive Language* (pp. 88–93).

25. Sood, S., Antin, J., & Churchill, E. (2012). Profanity use in online communities. In *Proceedings of CHI 2012* (pp. 1481–1490).

26. Biyani, P., Bhatia, S., Caragea, C., & Mitra, P. (2014). Using subjective logic to model user trust in online communities. In *Proceedings related to StackOverflow question quality* (pp. 1–10).

27. Harper, F. M., Moy, D., & Konstan, J. A. (2009). Facts or friends? Distinguishing informational and conversational questions in social Q&A sites. In *Proceedings of CHI 2009* (pp. 759–768).

28. Shah, C., & Pomerantz, J. (2010). Evaluating and analyzing the quality of answers in community QA sites. In *Proceedings of SIGIR 2010* (pp. 543–550).

29. Malik, M., & colleagues. (2024). Hate speech detection in social media using deep learning. *International Journal of Applied Engineering and Management, 4*(11), 101–110.

30. Lee, K., Lee, S., & Choi, J. (2024). Deep learning for hate speech detection: A personality-aware approach. In *Proceedings of WWW 2024* (pp. 201–211).

31. Чіп С., Лось М., Козак А. (2025). Ієрархічний енкодер для атрибуції, генерації сегментів і суб'єктивності. У Матеріали IX Міжнародної студентської наукової конференції «Глобалізація наукових знань: міжнародна співпраця та інтеграція галузей наук», м. Черкаси, Україна.

32. Штинда В., Чіп С., Козак А. (2025). Огляд сучасних моделей у задачах класифікації зображень і текстів. У Збірник тез доповідей студентської науково-

практичної конференції «Інтелектуальні інформаційні технології в прикладних дослідженнях» (IITAR-2025) (с.103–105)

33. Комар М.П., Саченко А.О., Васильків Н.М., Загородня Д.І. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп’ютерні науки» спеціальності 122 «Комп’ютерні науки» за другим (магістерським) рівнем вищої освіти. – Тернопіль: ЗУНУ, 2024. – 32 с.

ДОДАТОК А

Апробація отриманих результатів

Рисунок А.1 – Фрагмент змісту журналу

Штинда Віктор
студентка групи КНм-11
shtynda_v@gmail.com
Чіп Святослав
студент групи КНм-11
chip_svyatik@gmail.com
Козак Аліна
студентка групи КНм-11
alinka_kozak@gmail.com
Західноукраїнський національний університет
Тернопіль, Україна

ОГЛЯД СУЧАСНИХ МОДЕЛЕЙ У ЗАДАЧАХ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ І ТЕКСТІВ

Сучасні моделі класифікації зображень і текстів демонструють стрімкий прогрес завдяки глибоким нейронним мережам, трансформерам та мульти-модальним підходам. Використання великих датасетів із зображень і текстів, таких як LAION або ImageNet, забезпечує точність до 95–98% у стандартизованих задачах. Значна увага приділяється zero-shot та few-shot навчанню, що дає змогу класифікувати нові класи без попередньої адаптації [1–15].

У дослідженні [1] було протестовано ResNet101 та EfficientNet для виявлення повеней із соцмереж. Досягнута точність становила 93.5% на датасеті із понад 10 тис. зображень. Також використано LSTM-RNN для текстового аналізу з точністю 91%.

Модель RemoteSAM [2] використовує 270K пар зображень і масок та підтримує zero-shot сегментацію. На Earth Observation Bench вона досягла mIoU = 78.4% і точність класифікації у 85.6% без fine-tuning.

Patho-R1 [3] пропонує мультиагентну модель для діагностики патологій, що покращує класифікацію медичних зображень на 12% порівняно з CLIP. Застосовано reinforcement learning на 36K зображеннях і описах, досягнуто F1 = 0.91.

У роботі [4] для розпізнавання емоцій застосовано Attention + LoRA, де комбінуються текст і зображення. Модель досягла точності 94.2% на датасеті MELD, перевищивши BERT на 3.5%.

У сфері хірургії [5] було проаналізовано performance моделей CLIP, BioViL-T, та GIT. Найвища точність класифікації інструментів — 89.7% при explainability score = 4.1/5 за шкалою SHAP.

MLLM-моделі [6], адаптовані до zero-shot класифікації, досягли точності 84% при розпізнаванні дій людини на зображенні. Було протестовано 8 типів prompt-методик на датасеті з 30K пар.

Рисунок А.2 – Сторінка статті

Для аналізу Airbnb стилів [7] було застосовано CNN+NLP pipeline, який досяг precision = 90.3% при класифікації дизайну приміщень на основі зображення та опису.

Уніфікована модель UniMoCo [8] дозволяє доповнювати відсутні модальності в embeddings. F1-score покращено на 6.1% при розпізнаванні QA-даних без повного вхідного набору.

Модель [9] для довгих новинних статей об'єднує вхідні layout, текст і зображення. MAP становив 87% при виділенні структурних елементів (заголовки, абзаци, підписи).

Контрастивна модель [10] для фейкових новин використовувала пари текст-зображення. На FakeNewsNet точність досягла 92.6% із приростом +7% до базових трансформерів.

Крос-атенційна модель BERT-ResNet [11] у мультимодальному аналізі емоцій досягла 95.1% точності на Twitter+VisualSentiment датасеті, обійшовши попередні моделі на ~4%.

CNN-огляд [12] моделей для класифікації хвороб рослин показав, що MobileNetV3 досягає точності 98.2% на PlantVillage, при цьому середній час інференсу — 22 мс.

VT2Music [13] — унікальна мультимодальна модель генерації музики на основі зображення і тексту. BLEU-4 = 32.5 і MOS оцінка слухачів = 4.2/5 на тестах з 150 аудіозаписами.

OCR+NLP pipeline [14] дозволив обробити тексти на низькоресурсних мовах з точністю 88.4% у задачах резюмування та 85.2% у перекладі, що на 12% вище ніж стандартні seq2seq.

Модель [15] класифікації гіперспектральних зображень із knowledge transfer досягла 91.3% точності при інкрементальному навчанні, що скорочує потребу у повному retraining.

Огляд наведених досліджень демонструє безпрецедентний прогрес у сфері класифікації зображень і текстів завдяки поєднанню глибокого навчання, мультимодальних підходів і контрастивного навчання. Сучасні моделі забезпечують високі показники точності — понад 90% у більшості задач, а також дозволяють досягати значних результатів у zero-shot та few-shot сценаріях. Інтеграція текстових і візуальних даних, застосування attention-механізмів та уніфікованих embeddings моделей створює основу для універсальних рішень у різних галузях — від медицини до агропромисловості. Такий технологічний прорив свідчить про те, що мульти-модальні класифікатори поступово трансформуються з вузькоспеціалізованих інструментів у ключові елементи майбутніх інтелектуальних систем.

Список використаних джерел

1. Arapostathis, S. G. (2025). *A Mash-Up of Social Media Datasets for Extracting Hydrological Information*. Preprints.

Рисунок А.3 – Сторінка статті

https://www.preprints.org/frontend/manuscript/d0ca393c0a99cefabcb0fcc21b8eb9c2/download_pub
2. Yao, L., Liu, F., & Chen, D. (2025). *RemoteSAM: Towards Segment Anything for Earth Observation*. https://www.researchgate.net/publication/391856415
3. Zhang, W., Zhang, P., Guo, J., Cheng, T., & Chen, J. (2025). *Patho-R1: A Multimodal Pathology Expert Reasoner*. https://arxiv.org/abs/2505.11404
4. Gorai, J., & Shaw, D. K. (2025). *Multimodal Emotion Detection*. https://doi.org/10.1007/s10579-025-09841-4
5. Cheng, J., Zhao, X., & Lin, S. (2025). *Benchmarking vision-language models for surgery*. https://arxiv.org/abs/2505.10764
6. Rabadessa Alcaide, O. (2025). *Multimodal LLMs for Zero-Shot Classification*. https://upcommons.upc.edu/handle/2117/430265
7. Woo, H., & Yoo, S. (2025). *Visual servicescape analytics in Airbnb*. https://doi.org/10.1108/jsm-09-2024-0481
8. Qin, J., Pu, Y., He, Z., Pan, D. Z., & Yu, B. (2025). *UniMoCo: Unified Modality Completion*. https://arxiv.org/abs/2505.11815
9. Almutairi, A., & Ali, A. A. (2025). *Article Element Detection Using Multimodal Transformers*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5264257
10. Chen, W., Cai, F., & Guo, Y. (2025). *Contrastive Learning for Fake News Detection*. https://doi.org/10.1007/s40747-025-01919-4
11. Gold, R. G. (2025). *BERT-ResNet Fusion for Sentiment Classification*. https://search.proquest.com/openview/9c87a30c54d3fead4363a8e18e286dbf
12. Priyadarshini, G. P., & Zahoor-Ul-Huq, S. (2025). *CNN Models for Plant Disease Detection*. https://www.atlantis-press.com/article/126011378.pdf
13. Zheng, J., Cao, M., & Zhang, C. (2025). *VT2Music: Text-Visual Music Generation*. https://ieeexplore.ieee.org/document/11014098/
14. Madhavi, H., Cherian, J., & Khamkar, Y. (2025). *OCR-Driven Low-Resource Processing*. https://arxiv.org/abs/2505.11177
15. Wang, K., Li, Z., & Guo, J. (2025). *Hyperspectral Image Incremental Classification*. https://ieeexplore.ieee.org/document/11002484/

Рисунок А.4 – Сторінка статті

Рисунок А.5 – Фрагмент змісту журналу

Чіп Святослав, здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Лось Марія, здобувачка вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Козак Аліна, здобувачка вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

ІЄРАРХІЧНИЙ ЕНКОДЕР ДЛЯ АТРИБУЦІЇ, ГЕНЕРАЦІЇ СЕГМЕНТІВ І СУБ'ЄКТИВНОСТІ

Запропонований метод розглядає три задачі як узгоджений багатозадачний процес обробки природної мови над спільним текстовим корпусом: авторська атрибуція (багатокласова класифікація), генерація відсутніх/нових сегментів тексту (умовне доповнення), та класифікація суб'єктивних запитань (бінарна/багатокласова класифікація). Усі задачі поділяють єдину трансформерну основу та словникову розмітку, а також використовують узгоджені протоколи попередньої обробки даних. Стратегія багатозадачного навчання дозволяє спільному енкодеру вчитися стабільнішим і більш універсальним ознакам стилю, змісту та прагматики, що підвищує узагальнювальну здатність кожної окремої підзадачі [1–4, 11, 16].

Корпус формується з документів, речень та (для запитань) коротких висловлювань. Для авторської атрибуції до кожного фрагмента додається мітка автора; для задачі генерації підготовлюються масковані або пропущені спани, а також умови (ключові слова/підказки/контекстні речення); для класифікації суб'єктивності питання/речення маркуються як «суб'єктивне/об'єктивне» або за градаціями суб'єктивності. Текст нормалізується, токенізується WordPiece/власним байт-парним токенізатором, зберігаються розриви речень і абзаців як структурні індекси для ієрархічної агрегації [1, 5, 6, 12]. Для балансування класів використовуються стратифіковані спліти, ваги класів, а також слабке розширення даних (перефразування, синонімізація, EDA-операції) з контролем смислової інваріантності [17].

Ядром системи є багатомовний/доменно-адаптований BERT (або його аналог), який кодує токени у контекстні ембеддинги. Для збереження багаторівневих сигналів (лексика—синтаксис—дискурс) поверх токенових ознак застосовується ієрархічна агрегація: (а) токен→речення (self-attention pooling), (б) речення→документ (hierarchical attention), (в) опційно — дискурсивні графові зв'язки для довгих документів [1, 7, 8, 13]. Таке шарування корисне і для стилометрії автора (макро-ритм, пунктуаційні патерни), і для виявлення суб'єктивності (лексико-прагматичні маркери), і для генерації (узгодженість сегментів) [9, 10].

Підхід енкодер для атрибуції, генерації сегментів і суб'єктивності:

1. Авторська атрибуція. На виході енкодера використовується стильометричний проєктор: багатоголовий self-attention для вибіркового

Рисунок А.6 – Сторінка статті

фокусування на маркерах стилю + MLP із softmax над множиною авторів. За потреби додається контрастивний допоміжний лосс з «якорями» автора (author prototypes) для зменшення міжкласного перетину в латентному просторі [2, 9].

2. Генерація сегментів на основі BERT. Для інфілінгу спанів застосовується span-masking (SpanBERT-подібна стратегія) або «BERT-енкодер + легкий авторегресивний декодер» (BERT2BERT) для умовної генерації сегментів за контекстом. Навчальна мета — поєднання маскованого мовного моделювання (MLM) зі словопослідовним крос-ентропійним лоссом для декодера; під час інференсу — beam search/Top-p семплінг із обмеженнями стилю/довжини [3, 14, 15].

3. Класифікація суб'єктивних запитань. Використовується багаторівнева (ієрархічна) голова: токенові вектори агрегуються в представлення висловлювання; далі — шар ко-уваги між запитанням і його контекстом (наприклад, попередній діалог/джерельний документ) і MLP-класифікатор. За потреби застосовується фокальна втрата для класів із рідкою суб'єктивністю [7, 8, 18].

Система навчається спільно з комбінованою втратою:

$$\mathcal{L} = \lambda_{\text{auth}}\mathcal{L}_{\text{CE}}^{\text{author}} + \lambda_{\text{gen}}(\mathcal{L}_{\text{MLM}} + \mathcal{L}_{\text{AR}}) + \lambda_{\text{subj}}\mathcal{L}_{\text{CE}}^{\text{subj}},$$

де ваги λавтоматично адаптуються за гетероскедастичною невизначеністю або методом динамічних градієнтів, щоб уникати домінування однієї підзадачі [11]. Оптимізація — AdamW з лінійним зменшенням швидкості навчання та warmup; регуляризація — dropout, label smoothing, Stochastic Weight Averaging [1, 4, 11].

Для корпусів зі специфічним стилем (наукові тексти, публіцистика, соцмережі) застосовується подальше донавчання енкодера на немаркованих даних (MLM) та адаптерні шари (parameter-efficient tuning). Для задачі генерації додаються контейнти «стилю автора» як умовні теги; для атрибуції — стилометричні фічі (довжини речень, частота пунктуації) в пізньому ф'южені; для суб'єктивності — лексикони емоцій/модальностей як додаткові сигнали, інтегровані через фьюжн-проєктор [2, 9, 12, 19].

Авторська атрибуція: macro-F1, balanced accuracy, ECE-калібрування; стійкість перевіряється cross-domain/leave-author-out сплітами [2, 9]. Генерація: автоматичні метрики (BLEU/ROUGE/BERTScore), а також людська оцінка узгодженості, стилю та фактичності; для інфілінгу — exact span match та перплексія [14, 15]. Суб'єктивність: macro-F1/ROC-AUC, Robustness-тести на перефразування, зовнішні валідуючі набори [7, 8, 18]. Для всієї системи проводиться абляційний аналіз впливу багатозадачності та ієрархічної агрегації.

Запроваджується перевірка фактичності для згенерованих сегментів (retrieval-augmented валідація або NLI-фільтрація), контроль авторських прав, політики приватності, детектування зсувів (bias) й аудит помилок. Для критичних сценаріїв застосовується конфіденційне логування та human-in-the-loop рецензування [10, 16, 20].

Список використаних джерел:

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT.
2. Stamatatos, E. (2009). A survey of modern authorship attribution methods. Journal of the American Society for Information Science and Technology, 60(3), 538–556.
3. Joshi, M., Chen, D., Liu, Y., Weld, D., Zettlemoyer, L., & Levy, O. (2020). SpanBERT: Improving pre-training by representing and predicting spans. TACL, 8, 64–77.

Рисунок А.7 – Сторінка статті

4. Vaswani, A., et al. (2017). Attention Is All You Need. NeurIPS.
5. Wu, Y., et al. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144.
6. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. ICLR (workshop).
7. Yang, Z., et al. (2016). Hierarchical Attention Networks for Document Classification. NAACL-HLT.
8. Liu, P., Qiu, X., & Huang, X. (2016). Recurrent Neural Network for Text Classification with Multi-Task Learning. IJCAI.
9. [9] Kestemont, M., et al. (2019). Overview of the Cross-domain Authorship Attribution Task at PAN 2019. CLEF.
10. Gehrmann, S., Strobelt, H., & Rush, A. (2019). GLTR: Statistical Detection and Visualization of Generated Text. ACL demo.
11. Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-Task Learning Using Uncertainty to Weigh Losses. CVPR.
12. Gururangan, S., et al. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. ACL.
13. Lin, Z., et al. (2017). A Structured Self-Attentive Sentence Embedding. ICLR.
14. Rothe, S., Narayan, S., & Severyn, A. (2020). Leveraging Pre-trained Checkpoints for Sequence Generation Tasks. TACL, 8, 264–280.
15. Holtzman, A., et al. (2020). The Curious Case of Neural Text Degeneration (Top-k/Top-p). ICLR.
16. Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. arXiv:1706.05098.
17. Wei, J., & Zou, K. (2019). EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. EMNLP (short).
18. Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization. ACL.
19. Houlsby, N., et al. (2019). Parameter-Efficient Transfer Learning for NLP. ICML.
20. Nie, Y., et al. (2020). Adversarial NLI: A New Benchmark for Natural Language Understanding. ACL.

Рисунок А.8 – Сторінка статті