

## ДОДАТОК Б

### Порівняльні таблиці

Таблиця Б.1 – Загальний огляд основних існуючих підходів.

№	Данні для аналізу	Керованість навчання	Ознаки	Моделі
1	Текст статті, заголовок	кероване	TF-IDF вектори слів, емоційна забарвленість, стилістична оцінка, читабельність	SVM, Випадковий ліс, Логістична регресія
2	Текст статті, заголовок	кероване	Контекстні вектори, глибинні ознаки	Трансформери, RNN, CNN, LSTM
3	Текст статті, заголовок	частково кероване	Генерація штучних новин	LLM, GAN
4	Контекст ззовні	кероване	ГЗ з новинами, людьми, подіями, місцями тощо	GNN, GCN, GAT
5	Контекст ззовні	кероване	Контекстні вектори внутрішніх та зовнішніх даних	Моделі з механізмом глобальної та локальної уваги, LLM
6	Контекст ззовні	частково кероване	Генерація штучних ознак, ГЗ	GAN, GAT
7	Контекст ззовні	частково кероване	Попередньо треновані моделі, контекстні вектори нових даних	Трансформери (BERT, GPT)
8	Дані користувачів	кероване	Вектори коментарів, емоційна оцінка	Трансформери, RNN, CNN, LSTM
9	Контекст ззовні	кероване	Репутація та зв'язки між користувачами	GNN, GCN, GAT
10	Контекст ззовні	не кероване	Попередньо треновані моделі, контекстні вектори нових даних	Ймовірнісні моделі, алгоритми виявлення аномалій
11	Дані користувачів	частково кероване	Оцінка користувачів	CNN, трансформери, LSTM

## Продовження таблиці Б.1

№	Данні для аналізу	Керованість навчання	Ознаки	Моделі
12	Дані поширення мережею	кероване	Граф поширення, дерева	GNN, RNN, рекурсивні NN

## Продовження таблиці Б.1

№	Переваги	Недоліки
1	<ul style="list-style-type: none"> <li>– просто в реалізації;</li> <li>– не потрібно багато даних для навчання.</li> </ul>	<ul style="list-style-type: none"> <li>– відсутність розуміння контексту;</li> <li>– упередженість до стилю написання;</li> <li>– вразливість до підмінити понять.</li> </ul>
2	<ul style="list-style-type: none"> <li>– знаходять неочевидні закономірності</li> <li>– враховується контекст в межах речення</li> </ul>	<ul style="list-style-type: none"> <li>– вразливість до підмінити понять.</li> <li>– потреба в більших наборах даних</li> </ul>
3	<ul style="list-style-type: none"> <li>– розширюють набір даних</li> <li>– виявляють слабкі місця рішення</li> </ul>	<ul style="list-style-type: none"> <li>– може потребувати великі ресурсні витрати</li> </ul>
4	<ul style="list-style-type: none"> <li>– здатність дійсно перевірити інформацію</li> <li>– не упереджені до стилю</li> <li>– можливість пояснити результат</li> </ul>	<ul style="list-style-type: none"> <li>– потреба у великій кількості надійних даних</li> <li>– складність з визначенням у випадках нестачі контексту</li> <li>– займають багато місця</li> </ul>
5	<ul style="list-style-type: none"> <li>– здатність дійсно перевірити інформацію</li> <li>– не упереджені до стилю</li> </ul>	<ul style="list-style-type: none"> <li>– потреба у великій кількості надійних даних</li> <li>– складність з визначенням у випадках нестачі контексту</li> <li>– займають багато місця</li> </ul>
6	<ul style="list-style-type: none"> <li>– доповнення наборів даних</li> <li>– більша стійкість до шуму</li> </ul>	<ul style="list-style-type: none"> <li>– великі ресурсні витрати</li> </ul>
7	<ul style="list-style-type: none"> <li>– хороший результат при відсутності достатньої кількості початкових даних</li> </ul>	<ul style="list-style-type: none"> <li>– великі ресурсні витрати</li> </ul>

8	<ul style="list-style-type: none"> <li>– коментарі користувачів можуть містити пояснення</li> <li>– добре розпізнають маніпуляції</li> </ul>	<ul style="list-style-type: none"> <li>– відсутність перевірки інформації</li> <li>– потреба в великій кількості інформації</li> <li>– потреба часу</li> </ul>
---	--	--

## Продовження таблиці Б.1

№	Переваги	Недоліки
9	<ul style="list-style-type: none"> <li>– добре розпізнають дезінформацію на ранніх етапах</li> </ul>	<ul style="list-style-type: none"> <li>– відсутність перевірки інформації</li> <li>– потреба в великій кількості інформації</li> <li>– необхідна інформація часто є приватною</li> </ul>
10	<ul style="list-style-type: none"> <li>– хороший результат при відсутності достатньої кількості початкових даних</li> </ul>	<ul style="list-style-type: none"> <li>– великі ресурсні витрати</li> </ul>
11	<ul style="list-style-type: none"> <li>– підходять в ситуаціях нестачі промаркованих даних</li> </ul>	<ul style="list-style-type: none"> <li>– Потребують часу та надійних зацікавлених користувачів</li> </ul>
12	<ul style="list-style-type: none"> <li>– не потребують додаткової інформації про контекст</li> </ul>	<ul style="list-style-type: none"> <li>– відсутність перевірки інформації</li> <li>– необхідна інформація важко добувається</li> <li>– потреба часу</li> </ul>

## Таблиця Б.2 – Огляд доступних наборів даних.

№	Назва	Тематика	Джерело
1	BuzzFeedNews	політика	Публікації новинних агентств у Facebook
2	LIAR	політика	PolitiFact
3	Politifact Fact Check	політика	PolitiFact
4	Twitter15 та Twitter16	різне	Twitter
5	PHEME	різне	Twitter
6	Fakeddit	різне	Reddit

7	FakeHealth	медицина	Twitter
8	CoAID	медицина (COVID-19)	Сайти, Twitter
9	FakeCovid	медицина (COVID-19)	Різні медіа ресурси
10	ReCOVery	медицина (COVID-19)	Різні медіа ресурси
11	MM-COVID	медицина (COVID-19)	Різні медіа ресурси
10	ReCOVery	медицина (COVID-19)	Різні медіа ресурси

## Продовження таблиці Б.2

№	Назва	Тематика	Джерело
11	MM-COVID	медицина (COVID-19)	Різні медіа ресурси
12	News Headlines Dataset For Sarcasm Detection	комедія	The Onion, HuffPost
13	Tweets with Sarcasm and Irony	комедія	Twitter

## Продовження таблиці Б.2

№	Ознаки						Роки	Мова	N класів	N записів
	текст	заголовок	автор	посилання	відповіді / реакція користувачів	метадані				
1	X	X	X	X		X	2016	англ.	4	1 627
2		X	X			X	по 2017	англ.	7	12 836
3		X	X	X		X	2008 – 2022	англ.	7	21 152

4		X			X	X	2015 – 2016	англ.	4	2 308
5	X	X	X	X	X	X	2010 – 2018	англ.	2	137 496
6		X	X		X	X	по 2019	англ.	2, 3, 6	51 025
7	X	X	X	X	X	X	2018 – 2020	англ.	0	1 619
8		X	X	X	X		2020	англ.	2	5 216

Продовження таблиці Б.2

№	Ознаки						Роки	Мова	N класів	N записів
	текст	заголовок	автор	посилання	відповіді / реакція користувачів	метадані				
9	X	X	X	X		X	2020	багато	11	7 623
10	X	X	X	X		X	2020	англ.	2	2 029
11		X	X	X		X	2020	6	2	11 173
12		X		X			2019	англ.	2	26 709
13	X						2020	англ.	4	75 911

**ДОДАТОК А**  
Відомість кваліфікаційної роботи

Позначення					Найменування			Дод. відомості	
					Текстові документи				
<b>1.</b>					Пояснювальна записка			76 с.	
					Інші документи				
<b>2.</b>					Презентаційні матеріали			плакатів	
.	.	Прізвище та ініціали.	Підп.	Дата					
Розробив	Харіна О.С.								
Перевірив	Головянко М.В.				Дослідження методів інтелектуального аналізу медіаконтенту для визначення ключових інформаційних маніпуляцій	Шифр групи	Код напр./спец.		
Н.контр.	Малєєва І.А.					СШМ-23-2	122		
Затв.	Золотухін О.В.					<b>ХНУРЕ</b> кафедра ІІІ			