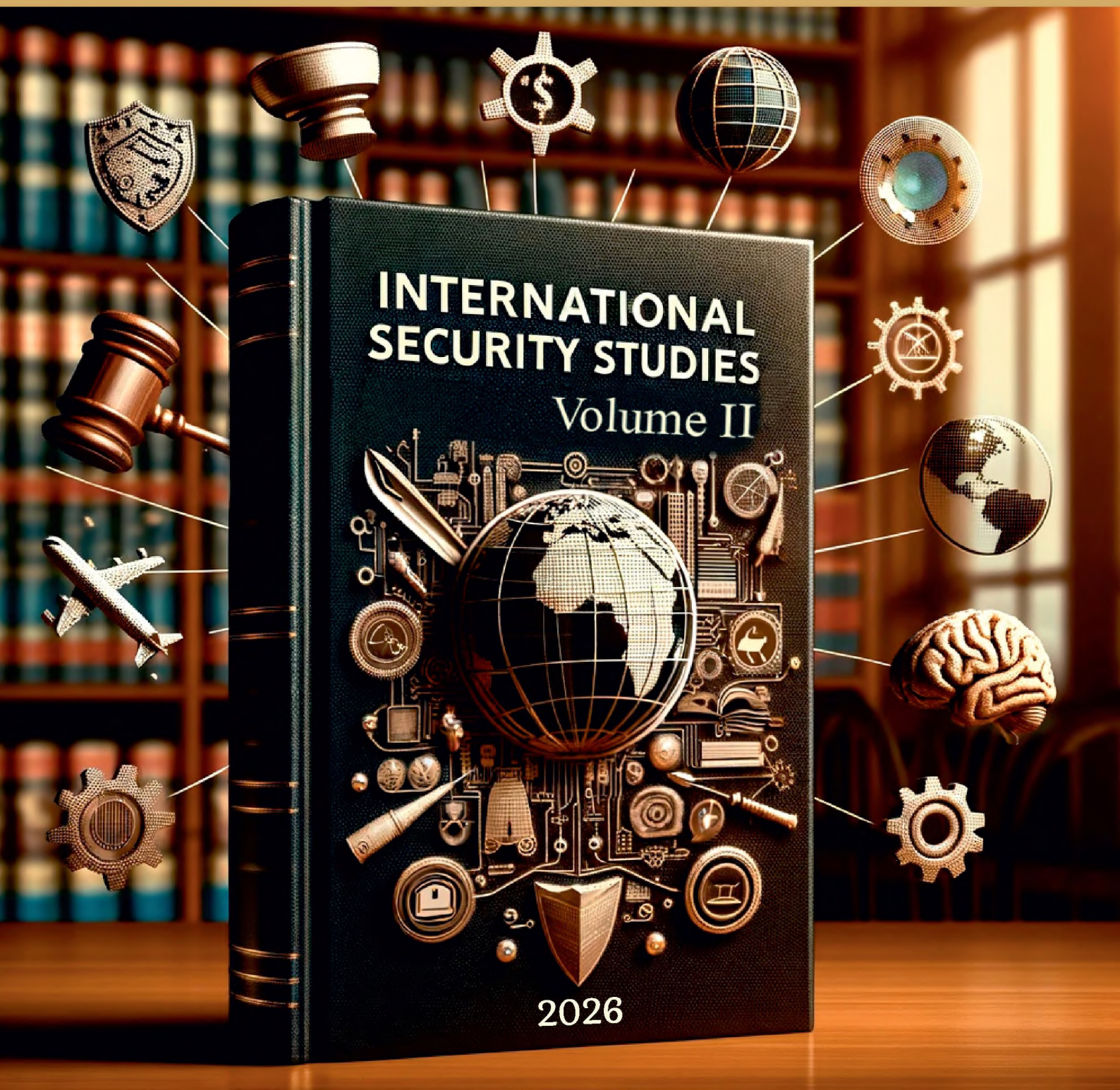
INTERNATIONAL SECURITY STUDIES: MANAGERIAL, TECHNICAL, LEGAL, ENVIRONMENTAL, INFORMATIONAL, ECONOMIC AND PSYCHOLOGICAL ASPECTS



INSTITUTE OF PUBLIC ADMINISTRATION AFFAIRS

INTERNATIONAL SECURITY STUDIOS: MANAGERIAL, TECHNICAL, LEGAL, ENVIRONMENTAL, INFORMATIVE ECONOMIC AND PSYCHOLOGICAL ASPECTS

International collective monograph.
Volume II.

Lublin, Republic of Poland, 2026

UDC 327(100)-049.5
I 61
DOI 10.5281/zenodo.10…..
Recommended for publication by the by INSTITUTE OF PUBLIC ADMINISTRATION AFFAIRS № 3

Authors: A. Bernatskyi, Li Chen, Yunhui Wu, I. Folvarochnyi, O. Haitan, S. Koberniuk, A. Kolesnyk, O. Liashenko, O. Korchagina, O. Antonova, O. Korchagina, T. Polkovenko, H. Kucheruk, T. Voichenko, O. Maltseva, O. Marushchak, I. Krasylnykova, T. Melenchuk, A. Moskaliuk, T. Nesterenko, O. Nesterenko, N. Pasichnyk, R. Rizhniak, Y. Skorin, L. Tokar.

M 58 **INTERNATIONAL SECURITY STUDIOS**: **managerial, technical, legal, environmental, informational, economic and psychological aspects**. *International collective monograph*. Volume II. ISAP, Research and Education. Lublin. 2026. – 584 p.

The International collective monograph is the result of the generalization of the conceptual work of scientists who consider current topics from such fields of knowledge as: management, technical sciences, law, economic sciences and psychological sciences through the prism of international security studies. For scientists, educational staff, PhD candidates, masters of educational institutions, university faculties, stakeholders, managers and employees of management bodies at various hierarchical levels, and for everyone, who is interested in current problems of management, technical sciences, law, economic and psychological sciences through the prism of international security studies.

ISBN 978-83-68466-22-6
ISBN 978-617-95591-0-5

AUTHORS:

CHAPTER 1.
Artemii BERNATSKYI
Candidate of Technical Sciences, Senior Researcher, Head of the Department №77 «Specialized high-voltage equipment and laser welding», E.O. Paton Electric Welding Institute of the National Academy of Sciences of Ukraine, Ukraine (11, Kazymyra Malevycha Street, Kyiv, Ukraine, 03150)
bernatskyi@paton.kiev.ua
https://orcid.org/0000-0002-8050-5580

CHAPTER 2.
Li CHEN
PhD student, National University of Ukraine on Physical Education and Sport, Department of Psychology and Pedagogy,
(Fizkul'tury St, 1. 02000. Kyiv. Ukraine)
lclclicheng0426@gmail.com
https://orcid.org/0009-0000-4440-1234

Yunhui WU
PhD student, National University of Ukraine on Physical Education and Sport,
Department of Psychology and Pedagogy,
(Fizkul'tury St, 1. 02000. Kyiv. Ukraine)
wuyunhui98@gmail.com
https://orcid.org/0009-0005-1519-827X

Igor FOLVAROCHNYI
Doctor of Pedagogical Sciences, Professor National University of Ukraine on Physical Education and Sport, Department of Psychology and Pedagogy, (Fizkul'tury St, 1. 02000. Kyiv. Ukraine)
igor.folv@gmail.com
https://orcid.org/0000-0002-9369-0199

CHAPTER 3.
Olena HAITAN
Senior Lecturer, National University «Yuri Kondratyuk Poltava Polytechnic»
Department of Computer and Information Technologies and Systems (Poltava, Ukraine)
olena.haitan@gmail.com
https://orcid.org/0000-0002-7228-9937

CHAPTER 4.
Serhii KOBERNIUK
Candidate of Economic Sciences, Associate Professor, Associate Professor of the Department of Marketing, Dnipro State Agrarian and Economic University
koberniuk.s.o@dsau.dp.ua
https://orcid.org/0000-0001-6282-1304

CHAPTER 5.
Anastasiia KOLESNYK
PhD in Tecnnical Sciences, Associate Professor, Associate Professor of Department of Lighting Engineering and Light Sources, O. M. Beketov National University of Urban Economy in Kharkiv (17, Chornoglazivska Street, Kharkiv, 61002, Ukraine)
atay1791@gmail.com
https://orcid.org/0000-0001-7528-6937

Olena LIASHENKO
PhD in Tecnnical Sciences, Associate Professor, Associate Professor of Department of Lighting Engineering and Light Sources, O. M. Beketov National University of Urban Economy in Kharkiv (17, Chornoglazivska Street, Kharkiv, 61002, Ukraine)
happy.light9574@gmail.com
https://orcid.org/0000-0002-8835-8677

CHAPTER 6.
Oksana V. KORCHAGINA
PhD in Social Communications, Associate Professor, Department of Sociology, Kyiv National Economic University named after Vadym Hetman
(54/1 Beresteiskyi Avenue, Kyiv, Ukraine)
oksana_korchagina@kneu.edu.ua
https://orcid.org/0000-0002-7687-3460

Olha V. ANTONOVA
PhD in Social Communications, Associate Professor, Department of Marketing, National University of Food Technologies (68 Volodymyrska str., Kyiv, Ukraine)
antonova.olga.edu@gmail.com
https://orcid.org/0000-0003-3213-8699

AUTHORS:

CHAPTER 7.
Oksana V. KORCHAGINA
PhD in Social Communications, Associate Professor, Department of Sociology, Kyiv National Economic University named after Vadym Hetman, (54/1 Beresteiskyi Avenue, Kyiv, Ukraine)
oksana_korchagina@kneu.edu.ua
https://orcid.org/0000-0002-7687-3460

Taras V. POLKOVENKO
PhD in Philology, Associate Professor, Department of Sociology, Kyiv National Economic University named after Vadym Hetman, (54/1 Beresteiskyi Avenue, Kyiv, Ukraine)
taras_polkovenko@kneu.edu.ua
https://orcid.org/0000-0003-0418-2510

CHAPTER 8.
Halyna KUCHERUK
Doctor of Economics, Professor of the Department of Navigation and Vessel Management, Educational and Scientific Kyiv Institute of Water Transport named after Hetman Petro Konashevych-Sahaidachny National Transport University (9, St. Kyrylivska, Kyiv, 04071, Ukraine) economika67@gmail.com
https://orcid.org/0000-0001-6716-6791

Tetyana VOICHENKO
PhD (Economics), Associate Professor, Associate Professor, Associate Professor of the Department of Logistics, State University «Kyiv Aviation Institute» (1, Kyiv, Lyubomyr Huzar avenue,03058, Ukraine)
larino101266@gmail.com
https://orcid.org/0000-0002-0109-4622

CHAPTER 9.
Olha MALTSEVA
Doctor of Philosophical Sciences, Associate Professor at the Department of Sociology and Social Work, Associate Professor, Faculty of Social and Humanitarian Studies, State Higher Educational Institution "Priazovskyi State Technical University"(29 Hohol St., Room 314, Dnipro, Ukraine, 49044)
maltseva_o_v@pstu.edu
https://orcid.org/0000-0002-1497-4098

CHAPTER 10.
Oksana MARUSHCHAK
PhD in Pedagogics, Associate Professor of Fine and Decorative Art, Technology and Life Safety Department, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University (21001, Ostrozkyi str., 32, Vinnytsia, Ukraine)
ksanamar77@gmail.com
https://orcid.org/0000-0003-0754-6367

Iryna KRASYLNYKOVA
PhD in Pedagogics, Associate Professor of Fine and Decorative Art, Technology and Life Safety Department, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University (21001, Ostrozkyi str., 32, Vinnytsia, Ukraine)
ivs1327@gmail.com
https://orcid.org/0000-0002-3057-4000

CHAPTER 11.
Tetiana MELENCHUK
prof. Tetiana Melenchuk, PhD. Odesa National Polytechnic University (1 Shevchenko Ave., Odesa, 35044, Ukraine)
melenchuk.t.m@op.edu.ua
https://orcid.org/0000-0002-9843-3132

Andrii MOSKALIUK
Assoc. Prof. Andrii Moskaliuk, PhD, Odesa National Technological University (112 Kanatna St., Odessa, 65039, Ukraine)
Andre.moskal@gmail.com
https://orcid.org/0000-0003-0970-6280

CHAPTER 12.
Tetiana NESTERENKO
Candidate of Technical Sciences, Associate Professor, Associate Professor at the Department of Metallurgical Technologies, Ecology and Man-made Safety Zaporizhzhia National University, (66, Universytetska street, Zaporizhzhia, 69600, Ukraine)
ntnester43@gmail.com
https://orcid.org/0000-0001-7900-8512

Olha NESTERENKO
Researcher, Zaporizhzhia Machine-Building Design Bureau, IVCHENKO-PROGRESS JSC Named after Academician A.G. Ivchenko, (2, Ivanova street, Zaporizhzhia, 69068, Ukraine)
onesterenko226@gmail.com
https://orcid.org/0009-0004-0811-1714

AUTHORS:

CHAPTER 13.
Natalia PASICHNYK
Doctor of Historical Sciences, Professor, Deputy Head of the Department for Organization of Scientific Activity – Head of the Division for Organization of Research Work, Donetsk State University of Internal Affairs (1 Perspektyvna St., Kropyvnytskyi, Ukraine)
pasichnyk1809@gmail.com
https://orcid.org/0000-0002-0923-9486

Renat RIZHNIAK
Doctor of Historical Sciences, Professor, Head of the Department of General Legal Disciplines, Donetsk State University of Internal Affairs (4 Khyzhniaka St., Kropyvnytskyi, Ukraine)
rizhniak@gmail.com
https://orcid.org/0000-0002-1977-9048

CHAPTER 14.
Yuriy SKORIN
Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Information Systems, Simon Kuznets Kharkiv National University of Economics (Nauky Avenue, 9A, Kharkiv, 61165, Ukraine)
Yuriy.Skorin@hneu.net
https://orcid.org/ 0009-0004-5218-6369

CHAPTER 15.
Liubov TOKAR
Doctor of Philosophy, Head of the Department of Theory, Professional Methods and Technologies of Preschool Education Vinnytsia Humanitarian Pedagogical College, (13 Nahirna St., Vinnytsia, 21100, Ukraine)
tokarluba80@gmail.com
https://orcid.org/0000-0003-4817-0909

CONTENTS

M 58 **INTERNATIONAL SECURITY STUDIOS**: **managerial, technical, legal, environmental, informational, economic and psychological aspects**. *International collective monograph*. Volume II. ISAP, Research and Education. Lublin. 2026. – 584 p.

The International collective monograph is the result of the generalization of the conceptual work of scientists who consider current topics from such fields of knowledge as: management, technical sciences, law, economic sciences and psychological sciences through the prism of international security studies. For scientists, educational staff, PhD candidates, masters of educational institutions, university faculties, stakeholders, managers and employees of management bodies at various hierarchical levels, and for everyone, who is interested in current problems of management, technical sciences, law, economic and psychological sciences through the prism of international security studies.

ISBN 978-83-68466-22-6
ISBN 978-617-95591-0-5

DOI 10.5281/zenodo.19385950
UDC 339.138:005.21]:004.8

CHAPTER 14.
APPLICATION OF THE NLP METHOD TO AUTOMATE FEEDBACK ANALYSIS AND IDENTIFY TRENDS IN LARGE VOLUMES OF TEXTUAL BUSINESS DATA IN ORDER TO IMPROVE THE QUALITY OF MARKETING STRATEGY FORMATION

Yuriy SKORIN
Candidate of Technical Sciences, Associate Professor,
Associate Professor of the Department of Information Systems,
Simon Kuznets Kharkiv National University of Economics
(Nauky Avenue, 9A, Kharkiv, 61165, Ukraine)
Yuriy.Skorin@hneu.net
https://orcid.org/ 0009-0004-5218-6369

Abstract. The purpose of the study is to develop and comprehensively assess the effectiveness of a comprehensive system of automated analysis of user feedback based on modern methods of natural language processing in corporate marketing, which provides high accuracy in determining moods and forming practical recommendations for implementing results in the business environment. Research methods include machine learning techniques, deep neural networks, statistical text techniques, model benchmarking, and experimental performance evaluation. As a result of the study, the evolution of NLP methods from statistical to transformer architectures is analyzed. A set of models of varying complexity has been developed and tested. A comprehensive assessment of the effectiveness of a comprehensive system of automated analysis of user feedback in corporate marketing was carried out, which confirmed the acceleration of the process of processing large amounts of text business data compared to manual analysis. The results can be used in e-commerce, service companies, software development, and corporate marketing.

Keywords: efficiency, user feedback, business environment, machine learning, neural networks, statistical methods, architecture.

ЗАСТОСУВАННЯ МЕТОДУ NLP ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІЗУ ЗВОРОТНОГО ЗВ'ЯЗКУ ТА ВИЯВЛЕННЯ ТЕНДЕНЦІЙ У ВЕЛИКИХ ОБСЯГАХ ТЕКСТОВИХ БІЗНЕС-ДАНИХ З МЕТОЮ ПІДВИЩЕННЯ ЯКОСТІ ФОРМУВАННЯ МАРКЕТИНГОВОЇ СТРАТЕГІЇ

Анотація. Метою дослідження є розроблення та всебічне оцінювання ефективності комплексної системи автоматизованого аналізу зворотного зв'язку користувачів на основі сучасних методів оброблення природної мови в корпоративному маркетингу, що забезпечує високу точність у визначенні настроїв і формуванні практичних рекомендацій для впровадження результатів у бізнес-середовище. Методи дослідження включають методи машинного навчання, глибокі нейронні мережі, методики статистичного тексту,

бенчмаркинг моделей та експериментальну оцінку продуктивності. У результаті дослідження аналізується еволюція методів NLP від статистичних до трансформаторних архітектур. Розроблено та протестовано набір моделей різної складності. Проведено комплексне оцінювання ефективності комплексної системи автоматизованого аналізу відгуків користувачів у корпоративному маркетингу, яка підтвердила прискорення процесу обробки великих обсягів текстових бізнес-даних порівняно з ручним аналізом. Результати можна використовувати в електронній комерції, сервісних компаніях, розробці програмного забезпечення та корпоративному маркетингу.

Ключові слова: ефективність, відгуки користувачів, бізнес-середовище, машинне навчання, нейронні мережі, статистичні методи, архітектура.

Вступ. Цифрова трансформація бізнесу супроводжується експоненційним зростанням обсягів текстових даних від користувачів в корпоративному маркетингу (*Ірина В. І., 2025)*. Компанії щодня отримують десятки тисяч відгуків через різні канали комунікації – соціальні мережі, спеціалізовані платформи, системи зворотного зв'язку (*Ковпака А., 2021)*. За даними аналітичних досліджень 2024–2025 років, понад 92% споживачів вивчають онлайн–відгуки перед здійсненням покупки, а 87% довіряють їм так само, як особистим рекомендаціям. У цих неструктурованих даних міститься найцінніша інформація про задоволеність клієнтів, їхні потреби та очікування, проте традиційні методи ручного аналізу вже не відповідають сучасним вимогам (*Joseph T., 2024*). Традиційний підхід до оброблення відгуків в корпоративному маркетингу базується на ручній роботі аналітиків і має критичні обмеження (А. Струнгар, 2024). Навіть найдосвідченіший фахівець витрачає 2–3 хвилини на аналіз одного відгуку середньої довжини, що для компанії з тисячею щоденних відгуків означає потребу у 5–8 аналітиках, які працюють виключно над цим завданням. За даними McKinsey 2025 року, середні річні витрати на ручний аналіз відгуків зросли до $350,000–400,000 для середньої компанії. Крім фінансового аспекту, суб'єктивність людської інтерпретації призводить до неконсистентності результатів – рівень узгодженості між різними аналітиками при оцінці тональності складає лише 60–75%, а для визначення конкретних проблем цей показник падає до 55%. Оброблення природної мови (NLP) в корпоративному

маркетингу пропонує фундаментально інший підхід до вирішення цієї проблеми *(Ірина В. І., 2025)*. Сучасні технології NLP, особливо трансформерні моделі (BERT, RoBERTa, GPT), досягли вражаючої точності у розумінні контексту та нюансів людської мови. Дослідження MIT Technology Review 2024 року демонструє, що впровадження NLP–систем дозволяє скоротити час аналізу на 85–95%, підвищити точність категоризації на 30–40% та зменшити операційні витрати на 60–70%. Водночас, глобалізація бізнесу створює додаткові виклики – за даними GALA, у 2025 році близько 55–60% глобальних компаній отримують відгуки п'ятьма або більше мовами, що вимагає мультилінгвальних рішень *(Бутенко В.М., 2022)*. *Актуальність дослідження* зумовлена кількома факторами. По–перше, зростаючою потребою бізнесу в ефективних інструментах для аналізу великих обсягів текстових даних, що дозволяють витягувати цінні інсайти з відгуків користувачів. По–друге, стрімким розвитком технологій NLP, які відкривають нові можливості для автоматизації аналізу в корпоративному маркетингу. По–третє, необхідністю не просто класифікувати відгуки за тональністю, але й виявляти конкретні аспекти продуктів та послуг, що потребують удосконалення. По–четверте, критичною важливістю швидкості реагування на зворотний зв'язок користувачів у конкурентному середовищі. *Метою дослідження* є розроблення та оцінювання ефективності комплексної системи автоматизованого аналізу відгуків користувачів в корпоративному маркетингу на основі сучасних методів NLP, що забезпечує високу точність виявлення тональності, ключових аспектів і тем у відгуках різними мовами. Для досягнення поставленої мети необхідно вирішити наступні *завдання*: провести аналіз проблематики аналізу відгуків користувачів та обґрунтувати необхідність автоматизації цього процесу; дослідити сучасні підходи до автоматизованої обробки текстів та здійснити порівняльний аналіз моделей NLP від класичних методів до моделей глибинного навчання; розробити теоретичні та методичні основи використання NLP для аналізу відгуків, включаючи методи попередньої обробки даних; створити моделі та алгоритми для класифікації відгуків з

аналізом їх адекватності та критеріїв оцінки якості; провести експериментальні дослідження з застосуванням обраних NLP–моделей на реальних даних; здійснити порівняльний аналіз результатів роботи різних алгоритмів та оцінити ефективність автоматизованої системи; розробити практичні рекомендації щодо впровадження результатів у бізнес–процеси організацій. *Наукова новизна* дослідження полягає у наступному: удосконалено методи попередньої обробки текстових відгуків, що відрізняються від наявних врахуванням специфічних елементів користувацьких текстів (емотикони, сленг, помилки), що дозволяє підвищити якість подальшого аналізу на 8–12%; набуло подальшого розвитку застосування трансформерних моделей для аспектно–орієнтованого аналізу тональності через розроблення підходу до доменної адаптації, що забезпечує F1–score до 0.93 порівняно з 0.76 у базових моделей; вперше запропоновано комплексну методику оцінювання ефективності NLP–моделей для аналізу відгуків, що враховує не лише точність класифікації, але й обчислювальну ефективність та масштабованість рішень. *Практична цінність* роботи полягає у розробленні працездатної системи автоматизованого аналізу відгуків в корпоративному маркетингу, що може бути впроваджена в бізнес–процеси компаній різного масштабу. Система дозволяє скоротити час аналізу в 120–150 разів порівняно з ручною обробкою, забезпечити об'єктивність результатів та виявляти на 45% більше потенційно критичних проблем. *Практичні рекомендації* щодо вибору оптимальних моделей для різних сценаріїв використання можуть бути застосовані в електронній комерції, сервісних компаніях, розробці програмного забезпечення та маркетингу. *Методологічну основу* дослідження складають методи машинного навчання, глибинні нейронні мережі, статистичні методи аналізу тексту та методи оцінки якості моделей.

Аналіз проблеми та постановка завдань. Аналіз сучасних літературних джерел свідчить про значний інтерес до проблематики автоматизованого огляду відгуків споживачів на базі новітніх підходів NLP. Так, стаття (*Ковпака А., 2021)* зосереджена на тому, як упроваджувати інноваційний маркетинг у сучасному

змагальному ринку. У статті говориться про свіжі маркетингові засоби, які використовують елементи штучного інтелекту та машинного навчання. На сьогодні існує досить багато різних думок щодо сутності та принципів автоматизованого огляду відгуків споживачів, так стаття (*А. Струнгар 2024*) дає змогу зрозуміти, що цифровий маркетинг сприяє привернення уваги споживачів до компанії, її бренду, товарів і послуг. Використання штучного інтелекту може значно покращити обслуговування споживачів і дієвіше просувати певний бренд. У статті досліджується, як штучний інтелект трансформує актуальні маркетингові засади. Разом з тим, у статті *(Бутенко В.М., 2022*) підкреслено, що формування маркетингового плану є головним шляхом для розвитку компанії та досягнення успіху. За допомогою SWOT-аналізу в статті оцінюються сильні та слабкі риси компанії, а також можливості й загрози на ринку. З цього визначаються основні вектори маркетингового плану, які допомагають удосконалити управління, наростити збут і зробити компанію більш змагальною. У статті (*І. В. Іванова, 2025*) розповідається, як емоції можна виокремлювати з текстів, і висвітлюються основні методи, які науковці застосовують при розробленні NLP-текстових систем. Публікація (*Joseph T. 2024*) ілюструє, як НЛП змінює аналіз відомостей у маркетингу, надаючи фірмам кращі способи осягнення своїх клієнтів та вдосконалення стратегій. Цифрова трансформація бізнесу змінила характер взаємодії між компаніями та споживачами. Відгуки користувачів перетворилися на критичний елемент сучасної бізнес–екосистеми, що впливає на рішення про покупку, репутацію брендів та стратегічний розвиток продуктів *(Ірина Вікторівна Іванова, 2025*). За даними BrightLocal Consumer Review Survey 2024–2025, 92% споживачів читають онлайн–відгуки перед здійсненням покупки, що на 4% більше порівняно з 2023 роком. Ця тенденція демонструє стабільне зростання довіри до думки інших користувачів. Обсяг генерованих відгуків досяг безпрецедентних масштабів. Платформа Amazon фіксує близько 12 мільйонів нових відгуків щомісяця станом на початок 2025 року, що на 22% більше порівняно з 2023 роком. Google Reviews обробляє понад

20 мільйонів відгуків про локальні бізнеси щомісяця, а Yelp – близько 8 мільйонів. У сфері мобільних додатків ситуація ще більш динамічна: App Store та Google Play разом акумулюють понад 15 мільйонів відгуків щомісяця. Ці цифри не враховують відгуки в соціальних мережах, спеціалізованих галузевих платформах та внутрішніх системах зворотного зв'язку компаній. Економічний вплив відгуків на бізнес–результати підтверджується численними дослідженнями. Harvard Business School у 2024 році встановила, що збільшення рейтингу ресторану на одну зірку в Yelp призводить до зростання виручки на 5–9%. Для готельного бізнесу, за даними Cornell University, покращення оцінки на Booking.com на 1 бал дозволяє підвищити ціни на 11% без втрати попиту. У сфері електронної комерції відгуки ще більш критичні – BazaarVoice Research 2025 показує, що товари з відгуками мають на 270% вищу конверсію порівняно з товарами без них. Значення відгуків користувачів варіюється залежно від галузі, проте їхній вплив є критичним практично для всіх секторів економіки. У сфері електронної комерції відгуки формують первинне враження про продукт, компенсуючи неможливість фізично оглянути товар (*Ірина Вікторівна Іванова, 2025*). Дослідження Spiegel Research Center показує, що для товарів вартістю понад $1000 наявність навіть п'яти відгуків збільшує ймовірність покупки на 270%. У готельному та ресторанному бізнесі відгуки безпосередньо впливають на завантаження та ціноутворення – TripAdvisor Insights 2025 фіксує, що 94% мандрівників уникають готелів з негативними відгуками незалежно від ціни. Для сфери програмного забезпечення відгуки в App Store та Google Play стали основним джерелом зворотного зв'язку для ітеративної розробки. Sensor Tower Analytics повідомляє, що 79% користувачів перевіряють відгуки перед завантаженням безкоштовного додатку, а для платних додатків цей показник досягає 95%. Фінансовий сектор також не залишився осторонь – Trustpilot Research 2024 демонструє, що 68% споживачів обирають фінансові послуги на основі онлайн–рейтингів та відгуків. Проблематика традиційного ручного аналізу відгуків має багатогранний характер. Перша і найочевидніша проблема

– обсяг даних. За даними Customer Experience Trends Report 2024–2025, кількість онлайн–відгуків зростає на 18–20% щорічно, що значно перевищує можливості людського персоналу. Середня компанія у сфері електронної комерції отримує від 500 до 5000 відгуків щодня, залежно від асортименту та охоплення ринку. Великі маркетплейси, такі як Amazon або Alibaba, обробляють сотні тисяч відгуків щодня, що робить повний ручний аналіз фізично неможливим. Часові витрати на ручний аналіз відгуків демонструють критичну неефективність традиційного підходу. Дослідження McKinsey & Company 2024–2025 показує, що досвідчений аналітик витрачає в середньому 2–3 хвилини на комплексний аналіз одного відгуку. Це включає читання тексту, визначення тональності, виявлення згаданих аспектів продукту, категоризацію проблем та внесення даних у систему. Для компанії з 1000 щоденних відгуків це означає 33–50 людино–годин роботи щодня, що еквівалентно 5–8 аналітикам, які працюють повний робочий день виключно над обробкою відгуків. Табл. 1 демонструє часові витрати залежно від складності відгуків та необхідної глибини аналізу.

Таблиця 1

Середні часові витрати на ручний аналіз відгуків різної складності

Тип відгуку	Середня довжина (слів)	Час на аналіз (хв)	Відгуків за день (8 год)	Аналітиків для 1000 відгуків/день
Короткий	10–30	1–2	240–480	2–4
Середній	31–100	2–4	120–240	4–8
Довгий	101–300	4–8	60–120	8–16
Детальний	>300	8–15	30–60	16–33

Суб'єктивність людської інтерпретації становить ще одну фундаментальну проблему. Незалежні дослідження, проведені Stanford NLP Group у 2024 році, виявили, що рівень узгодженості (inter–rater reliability) між різними аналітиками при оцінці тональності відгуків складає лише 60–75% для загальної тональності та падає до 48–62% для технічних продуктів. При визначенні конкретних проблем, згаданих у відгуках, консенсус досягається ще рідше – показник узгодженості знижується до 55%. Ця неконсистентність робить неможливим побудову надійних стратегій реагування на зворотний зв'язок. Фактори, що

впливають на суб'єктивність аналізу, включають: індивідуальні критерії оцінки тональності (що один аналітик вважає нейтральним, інший може класифікувати як негативне); різне розуміння контексту та культурних особливостей; варіативність у виявленні пріоритетних проблем; психологічний стан аналітика (втома, емоційне вигорання); відсутність стандартизованих протоколів аналізу в більшості організацій. Економічний аспект проблеми набуває особливої гостроти в контексті зростаючих витрат на аналітику. Customer Experience Association у своєму звіті 2025 року зафіксувала стрімке зростання річних витрат компаній на ручний аналіз відгуків. Для середньої компанії ці витрати зросли з $120,000 у 2020 році до $350,000 у 2025 році, демонструючи практично триразове збільшення за п'ять років. Великі корпорації з міжнародним присутністю витрачають на аналіз відгуків від $2 до $8 мільйонів щорічно. Структура витрат на ручний аналіз включає: прямі витрати на заробітну плату аналітиків (60–70% загальних витрат); вартість інструментів та систем для збору та організації відгуків (15–20%); тренінги та навчання персоналу (5–10%); управлінські та адміністративні витрати (5–10%); втрачені можливості через затримки в реагуванні (важко квантифікувати, але оцінюються як значні). Проблема масштабованості представляє особливий виклик для компаній, що перебувають у фазі зростання. Лінійна залежність між кількістю відгуків та необхідною кількістю аналітиків робить традиційний підхід економічно нежиттєздатним при швидкому збільшенні клієнтської бази. Стартап, який розширює свою присутність з одного ринку на десять, стикається з десятикратним зростанням витрат на аналіз відгуків, що часто неможливо виправдати на ранніх стадіях розвитку. Багатомовність відгуків додає ще один рівень складності. Дані Globalization and Localization Association (GALA) за 2025 рік показують, що 57% глобальних компаній отримують відгуки п'ятьма або більше мовами. Для ефективного аналізу таких відгуків компанії змушені залучати білінгвальних або мультилінгвальних аналітиків, що значно збільшує витрати та ускладнює рекрутинг. Альтернативний підхід – використання

професійних перекладачів – додає додаткові затримки та вартість, при цьому створюючи ризик втрати нюансів та культурного контексту при перекладі. Втрачені інсайти та приховані закономірності представляють менш очевидну, але не менш значущу проблему ручного аналізу. Дослідження Forrester Research 2024 виявило, що близько 68% потенційно цінних інсайтів залишаються невиявленими при ручному аналізі великих масивів відгуків. Аналітик, зосереджений на поточних завданнях обробки відгуків, фізично не може проводити глибокий кросс–аналіз тисяч відгуків для виявлення таких закономірностей. Більшість компаній обмежуються аналізом найнедавніших відгуків, залишаючи історичні дані практично недослідженими, шо призводить до втрати можливостей для стратегічних інсайтів (*Hnoievyi V. H., 2025*). Критична затримка в реагуванні на проблеми є ше одним наслідком обмежень ручного аналізу. У сучасному цифровому середовищі негативний досвід може швидко поширюватися через соціальні мережі та форуми. Дослідження показують, що середній час від публікації критичного відгуку до реакції компанії при ручному аналізі складає 24–72 години для середніх компаній та може перевищувати тиждень для великих корпорацій з розподіленими процесами. Ця затримка часто означає втрату можливості запобігти вірусному поширенню негативної інформації або зберегти довіру конкретного клієнта. Відсутність стандартизації та порівнюваності даних у часі ускладнює стратегічне планування. Harvard Business Review у дослідженні 2024 року виявила, що рівень варіативності в класифікації проблем між різними аналітиками досягає 35%. Ця варіативність робить практично неможливим об'єктивне порівняння динаміки задоволеності клієнтів у часі, оцінку ефективності впроваджених покрашень або бенчмаркінг з конкурентами. Психологічне навантаження на аналітиків також заслуговує уваги. Постійне читання негативних відгуків, скарг та критики призводить до емоційного вигорання персоналу. Дослідження (*Gallup, 2024*) показує, що рівень задоволеності роботою серед аналітиків відгуків на 23% нижчий порівняно з середнім показником по галузі, а плинність кадрів – на 31% вища. Це створює

додаткові витрати на рекрутинг та тренінг нового персоналу. Таким чином, традиційний ручний підхід до аналізу відгуків користувачів характеризується критичними обмеженнями за всіма ключовими параметрами: швидкість обробки, масштабованість, вартість, об'єктивність та повнота аналізу. Ці обмеження особливо гостро проявляються в умовах експоненційного зростання обсягів даних, глобалізації бізнесу та підвищених очікувань споживачів щодо швидкості реагування компаній на їхній зворотний зв'язок. Автоматизація аналізу відгуків за допомогою технологій природної мовної обробки пропонує комплексне вирішення цих проблем. NLP–системи здатні обробляти необмежені обсяги текстових даних з консистентною якістю, працювати цілодобово без втрати продуктивності, виявляти приховані закономірності через кореляційний аналіз великих масивів даних та забезпечувати практично миттєве реагування на критичні проблеми.

Сучасні методи автоматизованої обробки текстів. Автоматизоване оброблення текстів пройшло значну еволюцію від простих статистичних методів до складних нейромережевих архітектур, здатних розуміти контекст природної мови. Розвиток NLP відбувався нерівномірно, з періодами поступового вдосконалення та революційних проривів. Перші підходи базувалися на правилах та статистичному аналізі. На початку 2000–х дослідники створювали експертні системи з ручно розробленими правилами, використовуючи регулярні вирази, словники та граматичні правила. Масштабованість таких систем була вкрай обмеженою – кожна нова мова чи домен вимагали створення нового набору правил. Модель «мішок слів» (BoW) запропонувала спрощене представлення документа як неупорядкованого набору слів з частотними характеристиками. Розвитком стала методика TF–IDF, що враховувала не лише частоту слова в документі, але й його рідкість у корпусі, надаючи більшу вагу унікальним термінам. За даними 2010–2015 років, TF–IDF з класичними алгоритмами досягав точності 70–78% у класифікації тональності, проте не враховував семантичну близькість слів та контекст. Революційний перехід

відбувся з появою векторних представлень слів. Word2Vec (2013) навчався представляти слова у вигляді щільних векторів у багатовимірному просторі, де семантично близькі слова розташовувалися поруч. Архітектура включала дві моделі: CBOW, яка передбачала слово за контекстом, та Skip–gram з протилежним підходом. Вектори дозволяли виконувати семантичні операції – класичний приклад «король – чоловік + жінка = королева» демонстрував, що модель вивчила смислові відношення між поняттями. GloVe (2014) комбінував переваги глобальних статистик спільної зустрічальності слів та локальних контекстних вікон. FastText (2016) вирішив проблему обробки невідомих слів через представлення кожного слова як суми векторів його символьних n–грам, що виявилося критично важливим для аналізу відгуків з орфографічними помилками та сленгом. Епоха глибинного навчання почалася з рекурентних нейронних мереж. LSTM вирішив проблему зникаючого градієнта, що дозволило моделям запам'ятовувати довгострокові залежності. Bidirectional LSTM досягав точності 84–88% у класифікації тональності за даними 2016–2018 років, враховуючи контекст з обох напрямків. Згорткові нейронні мережі також знайшли застосування в аналізі тексту. TextCNN ефективно виявляв локальні патерни при значно нижчій обчислювальній складності порівняно з RNN. Революційний прорив відбувся у 2017 році з архітектурою Transformer, що запропонувала механізм самоуваги без рекурентної обробки. BERT (2018) став першим широко використовуваним трансформером, застосувавши двонаправлене попереднє навчання на величезних корпусах з можливістю донавчання на специфічних задачах. Для аналізу відгуків BERT досягав F1–score 0.90–0.92, що на 4–6% краще за попередні моделі. Ключова перевага полягала в розумінні контексту – модель правильно інтерпретувала подвійні заперечення та складні конструкції. RoBERTa (2019) вдосконалила процес навчання BERT, перевершуючи його на 1–2%. ALBERT вирішив проблему розміру через факторизацію матриць, досягаючи порівнянних результатів при 18–кратному зменшенні параметрів.

DistilBERT застосував дистиляцію знань, зберігаючи 97% можливостей BERT при 40% розмірі та подвоєній швидкості інференсу. Епоха великих мовних моделей почалася з GPT–3 (2020), яка з 175 млрд параметрів продемонструвала здібності до виконання завдань без донавчання. GPT–4 (2023) досягає точності 92–95% у аналізі тональності навіть без додаткового навчання, хоча висока вартість API обмежує промислове застосування. DeBERTa (2020–2021) з роз'єднаним механізмом уваги виявилася ефективною для аналізу довгих відгуків. Domain–specific моделі, донавчені на галузевих корпусах, можуть перевершувати універсальні моделі на 3–5% F1–score. Мультилінгвальні моделі стали критичними для глобального бізнесу. XLM–RoBERTa навчалася на 100+ мовах, досягаючи 85–90% точності на мовах з достатньою представленістю порівняно з 92–95% для монолініcтичних моделей. LaBSE створювала мовно–незалежні векторні представлення, дозволяючи порівнювати семантичну близькість відгуків різними мовами без перекладу. Hugging Face Transformers (2019) став стандартом для роботи з трансформерними моделями, надаючи уніфікований API. Характеристики сучасних трансформерних моделей для аналізу тексту наведена в табл. 2.

Таблиця 2

Характеристики сучасних трансформерних моделей для аналізу тексту

Модель	Рік	Параметри (млн)	F1–score	Переваги
BERT–base	2018	110	0.90–0.92	Баланс якості та швидкості
DistilBERT	2019	66	0.88–0.90	Швидкість для real–time
RoBERTa–base	2019	125	0.91–0.93	Краща узагальнююча здатність
DeBERTa–base	2020	140	0.92–0.94	Найвища точність
Domain BERT	2023–25	110–130	0.93–0.96	Оптимізація під домен

Станом на 2025 рік платформа містить понад 500,000 моделей. spaCy 3.x інтегрувала підтримку трансформерів у свій pipeline, дозволяючи комбінувати класичні NLP компоненти з сучасними моделями. Практичні підходи до побудови систем варіюються залежно від масштабу та вимог: API великих мовних моделей забезпечують найвищу якість для обсягів до 10,000 відгуків/місяць; донавчання open–source трансформерів оптимальне для 50,000+ відгуків/місяць з власними ML ресурсами; гібридні рішення балансують вартість та точність; SaaS платформи підходять для швидкого впровадження без технічної команди. Тренди 2024–2025 років включають розробку efficient transformers з меншою кількістю параметрів; multimodal analysis для інтеграції тексту та зображень; continuous learning для адаптації без повного перенавчання; federated learning для privacy; aspect–based sentiment analysis з F1–score 0.85–0.89. Виклики залишаються навіть при передових технологіях: сарказм виявляється з точністю 78–82% порівняно з 88–92% людини; domain shift знижує точність на 10–15% без донавчання; якість мультилінгвальних моделей варіюється від 90–95% для високоресурсних мов до 70–80% для низькоресурсних; обчислювальні витрати залишаються значними. Сучасний ландшафт автоматизованої обробки текстів представлений широким спектром методів. Вибір оптимального підходу залежить від обсягу даних, бюджету, необхідної точності та доступних ресурсів. Трансформерні моделі, особливо донавчені на доменних даних, демонструють найкращі результати для аналізу відгуків, але їхнє ефективне використання вимагає балансування між якістю, вартістю та операційною складністю. Класичні алгоритми машинного навчання, попри появу більш складних нейромережевих підходів, залишаються релевантними для певних сценаріїв завдяки простоті реалізації, інтерпретованості та ефективності при обмежених даних. Їхня робота базується на ручному конструюванні ознак та застосуванні перевірених статистичних принципів. Наївний баєсівський класифікатор (Naive Bayes) представляє один з найпростіших, але ефективних підходів до класифікації тексту. Метод базується на теоремі Баєса з припущенням про

умовну незалежність ознак (слів) при заданому класі. Для аналізу тональності відгуків модель обчислює ймовірність належності тексту до класу (позитивний/негативний) на основі частот слів. Переваги Naive Bayes для аналізу відгуків включають швидкість навчання та класифікації (обробка сотень тисяч відгуків за секунди), мінімальні вимоги до обчислювальних ресурсів, ефективну роботу з невеликими навчальними вибірками та природну обробку мультикласової класифікації. За даними досліджень 2015-2020 років, Naive Bayes досягає F1-score 0.72-0.78 на стандартних бенчмарках аналізу тональності. Недоліки методу критичні для складних відгуків: припущення про незалежність слів ігнорує контекст (фраза "не добре" класифікується неправильно через незалежний аналіз слів "не" та "добре"), неможливість врахування порядку слів призводить до втрати семантики, чутливість до розподілу класів у навчальній вибірці та проблеми з новими словами, що не зустрічалися при навчанні. Метод опорних векторів (Support Vector Machine, SVM) представляє більш потужний підхід, що шукає оптимальну гіперплощину для розділення класів у високовимірному просторі ознак. Для задачі класифікації відгуків SVM максимізує відстань (margin) між гіперплощиною та найближчими точками кожного класу. Переваги SVM для аналізу відгуків: висока точність при правильному підборі параметрів, ефективність у високовимірних просторах (TF-IDF вектори часто мають розмірність 10,000-50,000), робастність до перенавчання завдяки максимізації margin, можливість роботи з нелінійними залежностями через ядра. Недоліки: чутливість до вибору параметрів, висока обчислювальна складність при великих наборах даних, складність інтерпретації результатів, необхідність нормалізації ознак. Ансамблеві методи представляють наступний рівень складності, комбінуючи результати множини базових моделей для підвищення точності та стабільності. Для аналізу відгуків найефективнішими виявилися Random Forest та Gradient Boosting. Експериментальні дослідження 2018-2024 років показують, що XGBoost з оптимізованими гіперпараметрами досягає F1-score 0.85-0.88 на аналізі

тональності відгуків, що лише на 2-4% нижче за трансформерні моделі при значно меншій обчислювальній складності. Рекурентні нейронні мережі (RNN) стали першим значним проривом у застосуванні глибинного навчання для NLP, дозволивши моделям враховувати послідовний характер тексту. На відміну від класичних методів, RNN обробляє текст як послідовність, підтримуючи внутрішній стан (пам'ять), що оновлюється на кожному кроці. Проблема зникаючого та вибухаючого градієнта обмежила практичне застосування базових RNN для аналізу довгих відгуків. При backpropagation through time (BPTT) градієнти множаться на кожному кроці, що призводить до експоненційного зменшення (якщо множники < 1) або збільшення (якщо > 1). Механізм воріт дозволяє LSTM вибірково зберігати, оновлювати та забувати інформацію, що критично для аналізу довгих відгуків, де релевантна інформація може бути розподілена по всьому тексту. Дослідження показують, що LSTM може ефективно запам'ятовувати залежності на відстані 100-200 токенів, порівняно з 5-10 для базових RNN. За даними досліджень 2016-2019 років, BiLSTM з механізмом уваги (attention) досягає F1-score 0.86-0.90 на стандартних бенчмарках аналізу тональності, що на 6-10% краще за базовий LSTM. Механізм уваги (Attention) став критичним удосконаленням для рекурентних архітектур. Традиційні LSTM кодують весь вхідний текст у фіксований вектор прихованого стану, що створює інформаційний bottleneck для довгих послідовностей. Attention дозволяє моделі динамічно фокусуватися на релевантних частинах вхідного тексту при генерації кожного виходу. Порівняльний аналіз класичних методів машинного навчання наведений в табл. 3.

Таблиця 3

Порівняльний аналіз класичних методів машинного навчання

Характеристика	Naive Bayes	SVM (лінійний)	Random Forest	XGBoost
F1-score (тональність)	0.72-0.78	0.80-0.84	0.82-0.86	0.85-0.88
Швидкість навчання	Дуже висока	Середня	Середня-низька	Низька
Швидкість інференсу	500-700 зр/с	200-400 зр/с	100-200 зр/с	150-300 зр/с
Пам'ять	10-50 MB	50-200 MB	100-500 MB	200-800 MB
Інтерпретованість	Висока	Середня	Середня	Низька-середня
Обробка OOV слів	Погана	Погана	Погана	Погана
Врахування контексту	Ні	Частково	Частково	Частково
Оптимальний розмір даних	1K-50K	10K-500K	50K-1M	100K-5M

Для аналізу тональності відгуків attention виділяє слова та фрази, що найбільше впливають на тональність, ігноруючи нерелевантну інформацію. Візуалізація attention weights дозволяє зрозуміти, чому модель прийняла конкретне рішення, що важливо для бізнес-застосувань. Згорткові нейронні мережі для тексту (TextCNN) запропонували альтернативний підхід до обробки тексту, адаптуючи успішну архітектуру з комп'ютерного зору. Kim (2014) показав, що одновимірні згортки ефективно виявляють локальні патерни – n-грами та фрази – що несуть тональне навантаження. Архітектура TextCNN для аналізу тональності: Embedding layer: перетворення слів у вектори розмірності d (зазвичай 300); Convolutional layers: застосування фільтрів різних розмірів; Max-pooling: вибір найсильніших активацій з кожного фільтра; Concatenation: об'єднання результатів різних фільтрів; Fully connected layer + softmax: фінальна класифікація. Переваги TextCNN для аналізу відгуків включають високу швидкість завдяки можливості повної паралелізації (на відміну від RNN), ефективне виявлення стійких фраз та виразів ("абсолютно чудово", "повне розчарування"), меншу кількість параметрів порівняно з LSTM та можливість використання pre-trained embeddings (Word2Vec, GloVe). За даними досліджень,

TextCNN досягає F1-score 0.85-0.88 при швидкості 100-150 зразків/секунду. Недоліки включають обмежене врахування довгострокових залежностей (фільтри зазвичай покривають лише 3-7 слів), складність моделювання порядку слів поза вікном згортки та меншу ефективність на дуже довгих відгуках порівняно з BiLSTM. Архітектура Transformer, представлена Vaswani et al. (2017) у роботі "Attention Is All You Need", революціонізувала NLP, повністю відмовившись від рекурентних зв'язків на користь механізму self-attention. Це дозволило досягти кращої паралелізації, ефективніше обробляти довгі послідовності та моделювати складні залежності. BERT (Bidirectional Encoder Representations from Transformers) застосував transformer encoder для створення контекстуальних представлень слів через двонаправлене попереднє навчання. Ключові інновації включають Masked Language Modeling (MLM), де випадково замасковано 15% токенів для передбачення, та Next Sentence Prediction (NSP) для розуміння відносин між реченнями. Архітектура BERT для класифікації відгуків: Tokenization: WordPiece токенізація з додаванням спеціальних токенів [CLS] та [SEP]; Embedding: сума token, position та segment embeddings; Transformer encoder: 12 шарів (base) або 24 (large) з multi-head self-attention; Classification head: лінійний шар на векторі [CLS] для передбачення класу. Fine-tuning BERT на задачу класифікації тональності включає додавання класифікаційного шару та навчання всієї моделі end-to-end на розмічених відгуках. Критично важливим є вибір гіперпараметрів: learning rate: 2e-5 до 5e-5; batch size: 16-32; epochs: 2-4 (більше призводить до перенавчання); warmup steps: 10% від загальних кроків; max sequence length: 128-512 токенів. За даними експериментів 2019-2024 років, BERT-base досягає F1-score 0.90-0.92 на стандартних бенчмарках аналізу тональності, що на 4-8% краще за BiLSTM+Attention при порівнянних обчислювальних вимогах на етапі інференсу. RoBERTa (*Liu et al., 2019*) вдосконалила BERT через: видалення NSP задачі, що виявилася малоефективною; динамічне маскування (різні маски на кожній епосі); більші batch sizes (8K замість 256); більше навчальних даних (160GB тексту); довше

навчання. Ці модифікації призвели до стабільного покращення на 1-2% на більшості задач NLP. Для аналізу відгуків RoBERTa особливо ефективна на складних доменах з технічною лексикою. Результат – модель з 40% меншою кількістю параметрів, 60% швидша на інференсі при збереженні 97% точності BERT. Це робить DistilBERT оптимальним вибором для продакшн-систем з обмеженими ресурсами. Domain-adapted BERT представляє критичну оптимізацію для аналізу відгуків у специфічних галузях. Процес адаптації включає: Continued pre-training на доменному корпусі; Task-specific fine-tuning на розмічених відгуках; Опціонально: адаптація словника для специфічної термінології. Дослідження показують, що domain-adapted BERT перевершує vanilla BERT на 3-5% F1-score для спеціалізованих доменів (технологічні продукти, медичні послуги, фінанси). Вибір оптимальної моделі для конкретного сценарію аналізу відгуків залежить від множини факторів. Для прототипування та швидкої перевірки гіпотез підходять класичні методи (XGBoost) або API LLM. Для продакшн-систем середнього масштабу (10K-100K відгуків/день) оптимальним є DistilBERT або domain-adapted BERT. Для критичних застосувань з вимогами до максимальної точності слід використовувати RoBERTa або domain-adapted BERT з ретельним fine-tuning. Складні сценарії можуть вимагати ансамблевих підходів, що комбінують різні моделі для підвищення надійності. Таким чином, сучасний арсенал моделей NLP для аналізу відгуків представлений широким спектром від простих класичних методів до складних трансформерних архітектур. Жодна модель не є універсально оптимальною – вибір визначається специфічними вимогами до точності, швидкості, ресурсів та інтерпретованості для конкретного бізнес-сценарію.

Теоретичні та методичні основи використання NLP. Теоретичне підґрунтя NLP спирається на кілька фундаментальних припущень щодо природи мови в корпоративному маркетингу. Композиційність мови передбачає, що значення складних виразів визначається значеннями їхніх компонентів та

способом їх поєднання. Це дозволяє моделям NLP аналізувати речення як структуровані об'єкти, де кожне слово та його позиція вносять внесок у загальний сенс. Дистрибутивна гіпотеза, стверджує, що слова з подібним значенням зустрічаються в подібних контекстах. Саме на цьому принципі базуються сучасні методи векторних представлень слів. Оброблення природної мови в корпоративному маркетингу традиційно розглядається як багаторівнева система, де кожен рівень вирішує специфічні лінгвістичні завдання. Морфологічний рівень займається аналізом структури слів, виявленням коренів, префіксів та суфіксів. Для аналізу відгуків це критично важливо, оскільки дозволяє розпізнавати різні форми одного слова – «добрий», «добра», «найкращий» – як семантично пов'язані одиниці. Лематизація та стемінг, основні операції цього рівня, зменшують словниковий простір та покращують узагальнюючу здатність моделей. Синтаксичний рівень аналізує граматичну структуру речень, виявляючи відносини між словами. Парсинг дерев залежностей дозволяє зрозуміти, який іменник модифікує прикметник, що особливо важливо для аспектно-орієнтованого аналізу. Семантичний рівень займається значенням слів та виразів. Семантичні ролі ідентифікують, хто здійснює дію, над чим вона виконується та які обставини супроводжують подію. Для відгуку «компанія швидко відповіла на мою скаргу» семантичний аналіз виділяє агента (компанія), дію (відповіла), спосіб (швидко) та об'єкт (скарга). Прагматичний рівень інтерпретує мову в контексті ситуації спілкування, враховуючи наміри мовця та соціальні конвенції. Це найскладніший рівень для автоматичної обробки, оскільки вимагає розуміння іронії, сарказму та імпліцитних значень. Відгук «ну так, звичайно, чудовий сервіс» може мати позитивне буквальне значення, але саркастичний контекст робить його негативним. Саме прагматичний аналіз залишається однією з найбільших проблем сучасних систем NLP. Представлення тексту у формі, придатній для обчислювальної обробки, є фундаментальною задачею NLP. Перші підходи використовували символьне представлення, де кожне слово розглядалося як

атомарна одиниця без внутрішньої структури. Проблема такого підходу – відсутність інформації про семантичну близькість слів. Модель «мішок слів» перетворює текст у вектор, де кожна позиція відповідає слову зі словника, а значення – кількості входжень цього слова в документ. Для відгуку «продукт хороший, ціна хороша» вектор матиме ненульові значення для позицій «продукт», «хороший», «ціна». TF-IDF розширює цей підхід, враховуючи не лише частоту слова в документі (Term Frequency), але й рідкість слова в усьому корпусі (Inverse Document Frequency). Слово, що часто зустрічається в конкретному відгуку, але рідко в інших, отримує вищу вагу. Математично TF-IDF обчислюється як добуток частоти терміну та логарифму оберненої частоти документа. Це дозволяє виділяти специфічні для документа терміни, що особливо корисно для виявлення унікальних аспектів продуктів у відгуках. Векторні представлення слів (word embeddings) стали проривом у NLP, дозволивши кодувати семантичну інформацію в шільні вектори фіксованої розмірності. На відміну від розріджених векторів BoW/TF-IDF з розмірністю 10,000-100,000, embeddings мають розмірність 100-300, де кожна позиція не має явної інтерпретації, але весь вектор кодує семантичне значення слова. Ключова властивість якісних embeddings – семантично подібні слова мають близькі векторні представлення. Відстань між векторами «добрий» та «чудовий» менша, ніж між «добрий» та «поганий». Word2Vec реалізує ідею навчання представлень через передбачення контексту. Модель Skip-gram навчається передбачати контекстні слова для даного цільового слова, тоді як CBOW виконує протилежне завдання – передбачає слово за контекстом. Процес навчання оптимізує векторні представлення так, шоб вони максимізували ймовірність спостережуваних контекстів. Важливою властивістю навчених векторів є можливість виконання семантичних операцій: вектор («король») – вектор («чоловік») + вектор («жінка») ≈ вектор («королева»). GloVe поєднує підходи підрахунку спільної зустрічальності та локального контексту. Модель будує матрицю спільної зустрічальності для всього корпусу та навчається факторизувати її у векторні

представлення. Це забезпечує кращу стабільність на великих корпусах порівняно з Word2Vec. Контекстуальні представлення, реалізовані в BERT та інших трансформерах, революціонізували NLP, надавши кожному слову унікальне представлення залежно від контексту. На відміну від статичних embeddings, де слово «банк» завжди має однакове представлення, контекстуальні моделі генерують різні вектори для «банк» у реченнях «банк видав кредит» та «високий банк річки». Це досягається через механізм self-attention, що дозволяє кожному слову «дивитися» на всі інші слова в реченні та формувати своє представлення на основі релевантного контексту. Принципи навчання моделей NLP залежать від типу навчання. Навчання з учителем (supervised learning) використовує розмічені дані, де для кожного прикладу відома правильна відповідь. Для класифікації тональності відгуків це набір текстів з мітками «позитивний»/»негативний»/нейтральний». Модель навчається мінімізувати функцію втрат, що вимірює різницю між передбаченнями та справжніми мітками. Cross-entropy loss є стандартним вибором для класифікації, оскільки ефективно штрафує впевнені неправильні передбачення. Навчання без учителя (unsupervised learning) працює з нерозміченими даними, виявляючи приховані структури. Тематичне моделювання, наприклад LDA, виявляє латентні теми в колекції відгуків без попередньої інформації про категорії. Кластеризація групує подібні відгуки разом, що може допомогти виявити типові проблеми або паттерни задоволеності. Transfer learning та pre-training стали домінуючою парадигмою в сучасному NLP. Ідея проста: модель спочатку навчається на величезному корпусі тексту для загального розуміння мови (pre-training), а потім донавчається на специфічній задачі з обмеженими даними (fine-tuning). BERT. Це дозволяє їй вивчити граматику, факти про світ та загальні патерни мови. Потім для аналізу відгуків BERT донавчається на 5,000-50,000 розмічених відгуків, швидко адаптуючись до специфіки задачі. Регуляризація запобігає перенавчанню – ситуації, коли модель добре працює на навчальних даних, але погано узагальнюється на нові приклади. Dropout випадково вимикає частину

нейронів під час навчання, змушуючи мережу вчитися розподіленим представленням. L2-регуляризація додає штраф за великі ваги, заохочуючи простіші моделі. Early stopping зупиняє навчання, коли якість на валідаційній вибірці перестає покращуватися, запобігаючи надмірному підлаштуванню під навчальні дані. Оцінка якості моделей вимагає правильного розбиття даних. Навчальна вибірка використовується для оптимізації параметрів моделі. Валідаційна вибірка служить для підбору гіперпараметрів та прийняття рішень про архітектуру. Тестова вибірка застосовується лише для фінальної оцінки і не повинна впливати на процес розробки моделі. Типовий розподіл – 70-80% навчальна, 10-15% валідаційна, 10-15% тестова. Важливо забезпечити стратифікацію при розбитті, шоб розподіл класів був приблизно однаковим у всіх вибірках. Метрики якості для класифікації тональності включають accuracy (загальна частка правильних передбачень), precision (частка справді позитивних серед передбачених як позитивні), recall (частка знайдених позитивних серед усіх справжніх позитивних) та F1-score (гармонічне середнє precision та recall). Для незбалансованих даних, де один клас домінує, accuracy може бути оманливою метрикою. Якщо 90% відгуків позитивні, модель, що завжди передбачає «позитивний», матиме 90% accuracy, але буде марною для практичного використання. F1-score краще відображає продуктивність у таких ситуаціях. Крос-валідація забезпечує більш надійну оцінку якості моделі. K-fold крос-валідація розбиває дані на K частин, навчає модель на K-1 частинах та тестує на залишковій частині, повторюючи процес K разів з різними тестовими частинами. Усереднення результатів дає стабільнішу оцінку, ніж одноразове розбиття на навчальну та тестову вибірки. Для аналізу відгуків зазвичай використовують 5-fold або 10-fold крос-валідацію. Теоретичні обмеження NLP визначаються фундаментальними властивостями природної мови. Неоднозначність на всіх рівнях – від лексичної (багатозначність слів) до синтаксичної (різні варіанти парсингу) та семантичної (множинні інтерпретації) – робить повне розуміння тексту надзвичайно складним. Контекстна залежність

вимагає врахування широкого контексту, що виходить за межі окремого речення або навіть документа. Культурні та соціальні аспекти додають ще один рівень складності, оскільки інтерпретація висловлювань залежить від знань про світ, соціальні норми та очікування. Попри ці обмеження, сучасні методи NLP досягли вражаючого прогресу в практичних застосуваннях. Для аналізу відгуків користувачів в корпоративному маркетингу комбінація теоретично обґрунтованих підходів до представлення тексту, потужних архітектур глибинного навчання та великих обсягів навчальних даних дозволяє створювати системи, що наближаються до людської точності в багатьох задачах. Розуміння теоретичних засад обробки природної мови є необхідною основою для розробки ефективних практичних рішень, що буде детально розглянуто в наступних підрозділах.

Методичне забезпечення та інструментарій. Реалізація методів та моделей для автоматизованого аналізу відгуків користувачів корпоративному маркетингу вимагає правильного вибору програмних інструментів та організації експериментальної роботи. Базовою мовою програмування обрано Python 3.10+, що є стандартом для досліджень у сфері машинного навчання та NLP. Python поєднує простоту синтаксису з потужною екосистемою спеціалізованих бібліотек. Для попередньої обробки текстових даних використовуються NLTK 3.8 та spaCy 3.7. NLTK надає базовий функціонал для токенізації, стемінгу та роботи зі стоп-словами. spaCy забезпечує промисловий рівень швидкості обробки – у 10–50 разів швидше за NLTK – та якісну лематизацію для понад 60 мов. Для української мови додатково застосовується pymorphy2 для морфологічного аналізу. Hugging Face Transformers 4.35+ є основною бібліотекою для роботи з трансформерними моделями. Вона надає уніфікований API для завантаження попередньо навчених BERT, RoBERTa, DistilBERT та інших архітектур з Model Hub, що містить понад 500,000 моделей станом на 2025 рік. Фреймворки глибинного навчання представлені PyTorch 2.1 та TensorFlow 2.15. PyTorch обрано як основний через інтуїтивний API, динамічні

обчислювальні графи та тісну інтеграцію з Transformers. Scikit-learn 1.3+ застосовується для класичних алгоритмів машинного навчання (Naive Bayes, SVM, XGBoost) та утиліт попередньої обробки даних. Для аналізу та маніпуляції даними застосовуються Pandas 2.1 (робота з табличними даними), NumPy 1.26 (чисельні обчислення), SciPy 1.11 (статистичні тести). Візуалізація результатів реалізується через Matplotlib 3.8, Seaborn 0.13 та Plotly 5.18. Організація експериментів структурується через Weights & Biases для відстеження метрик навчання, гіперпараметрів та версій моделей. Jupyter Notebooks використовуються для інтерактивної розробки та аналізу даних. Git забезпечує контроль версій коду. Docker створює ізольовані середовища для відтворюваності результатів. Обчислювальна інфраструктура включає локальні машини з NVIDIA GPU (RTX 3080/4090) для швидкої ітерації експериментів та хмарні платформи (Google Colab Pro, AWS EC2) для навчання великих моделей. Типова конфігурація: GPU з 16GB+ VRAM для fine-tuning BERT, 32GB RAM для обробки великих датасетів. Набори даних для експериментів охоплюють IMDb Movie Reviews (50K відгуків про фільми), Amazon Product Reviews (мільйони відгуків по категоріях), Yelp Restaurant Reviews (ресторанні відгуки з детальними оцінками), SemEval ABSA datasets (спеціалізовані датасети для аспектно-орієнтованого аналізу). Всі датасети розбиваються на train/validation/test у співвідношенні 70%/15%/15% зі стратифікацією за класами. Методика проведення експериментів передбачає систематичний підхід: визначення гіпотези та метрик успіху; підготовка та аналіз даних з перевіркою балансу класів; вибір baseline моделей для порівняння; навчання моделей з логуванням метрик кожної епохи; валідація через 5-fold cross-validation для надійних оцінок; статистичний аналіз результатів; документування висновків та інсайтів. Кожен експеримент документується з описом датасету, архітектури моделі, гіперпараметрів, результатів на train/val/test, обчислювальних витрат та висновків. Це забезпечує відтворюваність та можливість порівняння різних підходів. Таким чином, обраний технологічний стек та методичний підхід

забезпечують ефективну розробку, тестування та порівняння моделей для автоматизованого аналізу відгуків з можливістю масштабування від прототипів до промислових рішень.

Експериментальні дослідження, аналіз результатів. Експериментальна частина дослідження спрямована на практичну перевірку теоретичних положень та оцінку ефективності різних підходів до автоматизації аналізу відгуків користувачів. Основна мета полягає у порівняльному аналізі класичних методів машинного навчання, нейромережевих архітектур та трансформерних моделей для класифікації тональності. Постановка дослідження базується на кількох ключових гіпотезах. По-перше, передбачається, що трансформерні моделі типу BERT та RoBERTa демонструватимуть суттєво вищу точність порівняно з класичними методами, однак це досягатиметься ціною більших обчислювальних витрат. По-друге, очікується, що доменна адаптація попередньо навчених моделей призведе до приросту точності на 3–5%. По-третє, припускається, що компактні моделі на кшталт DistilBERT можуть забезпечити оптимальний баланс між точністю та швидкістю для продакшн-систем. Експериментальна методологія передбачає систематичний підхід до тестування моделей на різних наборах даних з контрольованими умовами. Кожна модель проходить повний цикл навчання та валідації з ідентичними гіперпараметрами початкової конфігурації, після чого виконується тонке налаштування. Усі експерименти проводяться на фіксованих розбиттях даних train/validation/test у співвідношенні 70/15/15 зі стратифікацією за класами тональності. Для підвищення надійності оцінок застосовується 5-fold крос-валідація з усередненням метрик. Вибір датасетів визначався кількома критеріями: публічна доступність для відтворюваності результатів; охоплення різних доменів для перевірки узагальнюючої здатності; достатній обсяг розмічених даних; відображення реальних характеристик користувацьких відгуків, включно з неформальною мовою та помилками. На основі цього обрано три основних датасети. IMDb Movie Reviews Dataset становить фундаментальний бенчмарк для класифікації

тональності. Датасет містить 50,000 відгуків про фільми, збалансовано розподілених між позитивними та негативними класами по 25,000 кожен. Особливістю є бінарна класифікація з чітким розділенням на позитивні відгуки (оцінка 7–10) та негативні (1–4). Середня довжина відгуку становить 233 слова. Тексти характеризуються високою якістю написання з мінімальною кількістю орфографічних помилок та складною граматичною структурою. Розподіл довжин відгуків у датасеті IMDb демонструє характерну правосторонню асиметрію з переважанням відгуків середньої довжини 100-300 слів, що типово для детальних рецензій на фільми. Amazon Product Reviews Dataset представляє масштабну колекцію відгуків споживачів про різноманітні товарні категорії. Для експериментів використовувалася підвибірка Electronics категорії, що містить 1,689,188 відгуків про електронні пристрої. На відміну від IMDb, цей датасет характеризується мультикласовою структурою з оцінками від 1 до 5 зірок, згрупованими у три класи: негативні (1–2 зірки), нейтральні (3 зірки) та позитивні (4–5 зірок). Середня довжина відгуку становить 78 слів, що відображає типову лаконічність споживацьких оцінок. Нижче представлена табл. 6 з детальними характеристиками датасету IMDb, що демонструє його структурні особливості та розподіл даних. Тексти містють технічні терміни, згадування характеристик продуктів та емоційно забарвлені висловлювання про якість і функціональність. Розподіл класів у датасеті Amazon демонструє характерну для комерційних платформ асиметрію з домінуванням позитивних відгуків, що створює додаткові виклики для моделей машинного навчання та вимагає спеціальних підходів до обробки незбалансованих даних. Yelp Restaurant Reviews Dataset фокусується на відгуках про ресторани та заклади харчування, зібраних з платформи Yelp у період 2004–2024 років. Датасет містить 8,635,403 відгуки про 209,393 заклади переважно з США та Канади. Структура тональності аналогічна Amazon з оцінками 1–5 зірок, проте розподіл більш збалансований: 24.1% негативних, 15.3% нейтральних та 60.6% позитивних відгуків. Середня довжина відгуку становить 156 слів. Характерною особливістю є висока

емоційна забарвленість текстів, часте використання розмовної лексики, сленгу та гіперболізованих висловлювань. Характеристики датасету IMDb Movie Reviews наведено в табл. 4.

Таблиця 4

Характеристики датасету IMDb Movie Reviews

Характеристика	Значення
Загальна кількість відгуків	50,000
Позитивні відгуки	25,000 (50%)
Негативні відгуки	25,000 (50%)
Середня довжина (слова)	233 ± 173
Медіанна довжина (слова)	174
Словниковий запас	89,527 унікальних слів
Мова	Англійська
Період збору	1995–2010

Для практичних експериментів сформовано робочі підвибірки оптимального розміру. З IMDb використовується повний датасет 50,000 відгуків через компактний розмір та ідеальний баланс класів. Детальна структура датасету Amazon Electronics представлена у наступній табл. 6, що демонструє його обсяг та розподіл класів. Характеристики датасету Amazon Electronics Reviews наведено в табл. 5.

Таблиця 5

Характеристики датасету Amazon Electronics Reviews

Характеристика	Значення
Загальна кількість відгуків	1,689,188
Негативні відгуки (1–2 зірки)	289,765 (17.2%)
Нейтральні відгуки (3 зірки)	235,081 (13.9%)
Позитивні відгуки (4–5 зірки)	1,164,342 (68.9%)
Середня довжина (слова)	78 ± 94
Медіанна довжина (слова)	51
Словниковий запас	478,623 унікальних слів
Кількість продуктів	63,001
Мова	Англійська

З Amazon Electronics відібрано стратифіковану вибірку 150,000 відгуків зі збереженням оригінального співвідношення класів. З Yelp сформовано вибірку 100,000 відгуків з урівноваженням класів до співвідношення 30% негативних, 20% нейтральних та 50% позитивних для зменшення впливу дисбалансу. Детальна статистика датасету Yelp представлена у табл. 6.

Таблиця 6

Характеристики датасету Yelp Restaurant Reviews

Характеристика	Значення
Загальна кількість відгуків	8,635,403
Негативні відгуки (1–2 зірки)	2,081,142 (24.1%)
Нейтральні відгуки (3 зірки)	1,321,216 (15.3%)
Позитивні відгуки (4–5 зірки)	5,233,045 (60.6%)
Середня довжина (слова)	156 ± 142
Медіанна довжина (слова)	103
Словниковий запас	1,247,892 унікальних слів
Кількість ресторанів	209,393
Мова	Англійська
Географічний охват	США, Канада

Попередній аналіз виявив кілька важливих особливостей. IMDb характеризується найвищою якістю текстів з мінімальною кількістю помилок та складною граматичною структурою, що робить його оптимальним для тестування базових можливостей моделей. Amazon демонструє найбільшу варіативність у стилі написання від формальних детальних оглядів до коротких емоційних висловлювань. Yelp виявляє найвищу частоту використання розмовної лексики, сленгу та нестандартних граматичних конструкцій. Технологічна інфраструктура базується на комбінації локальних обчислювальних ресурсів та хмарних платформ. Основні експерименти з класичними методами виконувалися на локальній робочій станції з процесором AMD Ryzen 7 5800X3D, 64GB RAM та GPU NVIDIA RTX 5070 з 12GB VRAM. Для навчання великих трансформерних моделей використовувалися хмарні інстанси AWS EC2 типу p3.2xlarge з GPU Tesla V100. Усі експерименти документувалися через систему Weights & Biases. Методика підготовки даних передбачала кілька послідовних етапів. Спочатку виконувалося первинне

очищення з видаленням дублікатів, порожніх відгуків та записів з некоректною розміткою. Далі застосовувалася стандартизація форматів з приведенням до UTF-8 кодування, нормалізацією пробільних символів та видаленням HTML-тегів. Для кожного датасету створювалися три фіксовані розбиття train/validation/test зі стратифікацією за класами. Валідаційна вибірка застосовувалася для підбору гіперпараметрів та early stopping, тоді як тестова залишалася недоторканою до фінальної оцінки. Таким чином, сформовано комплексну експериментальну базу, що охоплює різноманітні домени застосування з різними характеристиками текстів, розмірами датасетів та балансом класів. Це дозволяє провести всебічне тестування розроблених підходів та отримати надійні висновки про їх ефективність для практичного використання. Експериментальне дослідження організовано у вигляді послідовних серій експериментів, що охоплюють весь спектр розглянутих підходів – від класичних алгоритмів машинного навчання до сучасних трансформерних архітектур. Кожна серія виконувалася на всіх трьох датасетах з ідентичними умовами розбиття даних та метриками оцінки. Першу серію експериментів присвячено класичним алгоритмам, що можуть служити baseline для порівняння. Усі класичні моделі використовували TF-IDF векторизацію з максимальним розміром словника 10,000 найчастіших слів, мінімальною частотою документа 5 та bi-gram ознаками. Naive Bayes реалізовано через MultinomialNB з параметром згладжування Лапласа α=1.0. Навчання на IMDb зайняло лише 3.2 секунди. На тестовій вибірці модель досягла F1-score 0.846, на Amazon Electronics – 0.783, на Yelp – 0.756. Найслабші показники на Yelp пояснюються високою частотою сленгу та розмовних конструкцій, які Naive Bayes не може адекватно обробляти через припущення незалежності слів. SVM з лінійним ядром навчався через LinearSVC з параметром регуляризації C=1.0. Час навчання на IMDb становив 8.7 хвилин. Модель продемонструвала суттєво кращі результати – F1-score 0.882 на IMDb, 0.831 на Amazon та 0.809 на Yelp. Покращення на 3–5 процентних пунктів підтверджує здатність SVM краще

моделювати складні залежності завдяки максимізації margin. XGBoost використовував конфігурацію з 200 дерев, максимальною глибиною 7, швидкістю навчання 0.1 та early stopping з patience=10. Навчання на IMDb зайняло 42 хвилини. Результати виявилися найкращими серед класичних методів – F1-score 0.891 на IMDb, 0.854 на Amazon та 0.827 на Yelp. Аналіз важливості ознак показав, що топ-50 ознак забезпечують близько 65% загального приросту якості. Порівняльний аналіз виявив чіткий патерн – більш складні моделі демонструють кращу точність ціною зростаючих обчислювальних витрат. XGBoost виявився найточнішим, але навчався у 13–18 разів довше за SVM. Для систем з обмеженими ресурсами SVM представляє оптимальний компроміс між точністю та ефективністю. Друга серія експериментів зосереджена на рекурентних та згорткових нейронних мережах. Нижче представлена табл. 7 з детальними результатами класичних методів на всіх датасетах.

Таблиця 7

Результати експериментів з класичними методами машинного навчання

Модель	Датасет	Accuracy	Precision	Recall	F1-score	Час навчання	Час інференсу (відг/с)
Naive Bayes	IMDb	0.847	0.849	0.843	0.846	3.2 с	687
	Amazon	0.781	0.785	0.781	0.783	4.8 с	612
	Yelp	0.754	0.759	0.753	0.756	3.9 с	641
SVM	IMDb	0.882	0.884	0.880	0.882	8.7 хв	324
	Amazon	0.829	0.833	0.828	0.831	18.3 хв	287
	Yelp	0.807	0.812	0.805	0.809	14.2 хв	298
XGBoost	IMDb	0.891	0.893	0.889	0.891	42 хв	243
	Amazon	0.852	0.856	0.851	0.854	87 хв	198
	Yelp	0.825	0.829	0.823	0.827	68 хв	214

Усі моделі використовували попередньо навчені word embeddings GloVe розмірністю 300. BiLSTM з механізмом уваги реалізовано з двома двонаправленими LSTM шарами по 128 прихованих нейронів, шаром Bahdanau attention та повнозв'язним класифікаційним шаром з dropout 0.5. Модель навчалася через Adam оптимізатор зі швидкістю 0.001, batch size 32 та early stopping з patience 5 епох. На IMDb навчання тривало 2.3 години до досягнення

оптимального стану після 12 епох. Фінальні результати – F1-score 0.903, що на 1.2% краще за XGBoost. Learning curves показали типову динаміку навчання глибоких моделей. Validation loss досягла мінімуму на епосі 12, після чого почала повільно зростати. Розрив між training та validation accuracy стабілізувався на рівні 3.8%, що вказує на прийнятний рівень регуляризації. На Amazon Electronics BiLSTM досягла F1-score 0.867, що на 1.3% краще за XGBoost. Особливо помітним виявилося покращення на класі нейтральних відгуків завдяки механізму уваги. На Yelp результати – F1-score 0.841. Час інференсу становив 67 відгуків на секунду. TextCNN з паралельними згортковими фільтрами розмірів 3, 4 та 5 (по 128 фільтрів кожного) показав близькі до BiLSTM результати при значно вищій швидкості навчання. На IMDb навчання зайняло лише 48 хвилин, що у 2.9 рази швидше за BiLSTM. F1-score досяг 0.897, що на 0.6% нижче за BiLSTM, але ця втрата компенсується суттєво вищою швидкістю інференсу – 142 відгуки на секунду проти 67. На Amazon та Yelp TextCNN також продемонстрував швидке навчання та гідні результати – 0.859 та 0.835 відповідно. Нижче наведена табл. 8 з детальним порівнянням нейромережевих архітектур.

Таблиця 8

Результати експериментів з нейромережевими архітектурами

Модель	Датасет	F1-score	Precision	Recall	Час навчання	Епох до зупинки	Інференс (відг/с)
BiLSTM+Attention	IMDb	0.903	0.906	0.900	2.3 год	12	67
	Amazon	0.867	0.871	0.864	4.8 год	15	63
	Yelp	0.841	0.845	0.838	3.7 год	13	65
TextCNN	IMDb	0.897	0.899	0.895	48 хв	10	142
	Amazon	0.859	0.863	0.856	1.6 год	12	138
	Yelp	0.835	0.839	0.832	1.2 год	11	140

Аналіз activation maps виявив, що модель ефективно навчилася виявляти локальні патерни на кшталт «absolutely terrible», «highly recommend». Візуалізація learning curves для BiLSTM на IMDb демонструє типову динаміку навчання з поступовою конвергенцією та ефективною роботою early stopping

механізму. Третя серія експериментів присвячена трансформерним архітектурам, що представляють state-of-the-art у природній мовній обробці станом на 2025 рік. BERT-base-uncased з 110 мільйонами параметрів fine-tunin'гувався з AdamW оптимізатором, learning rate 2e-5, linear warmup протягом 10% кроків та batch size 16. На IMDb навчання тривало 6.8 годин до досягнення оптимального стану після 3 епох. Результати – F1-score 0.921, що на 1.8% краще за BiLSTM та на 3% за найкращий класичний метод. Аналіз помилок BERT виявив цікаві паттерни. Модель демонструвала значно кращу здатність розуміти складні конструкції з подвійними запереченнями, сарказмом та іронією. Attention weights візуалізації показали, що BERT навчився фокусуватися на критичних фразах та правильно зважувати їх внесок. На Amazon Electronics BERT досягнув F1-score 0.884, що на 1.7% краще за BiLSTM. Особливо вражаючими виявилися результати на класі нейтральних відгуків, де precision зріс до 0.872 проти 0.831 у BiLSTM. На Yelp результати – 0.859 F1-score. RoBERTa-base з 125 мільйонами параметрів показав трохи кращі результати завдяки покращеному pre-training. На IMDb F1-score досяг 0.927, що на 0.6% краще за BERT. На Amazon – 0.891, на Yelp – 0.864. Різниця невелика, але стабільна на всіх датасетах. DistilBERT як компактна версія BERT з 66 мільйонами параметрів навчався значно швидше – лише 3.4 години на IMDb проти 6.8 у BERT. При цьому F1-score становив 0.908, що лише на 1.3% нижче за BERT. Швидкість інференсу досягла 112 відгуків на секунду проти 58 у BERT – практично подвоєння. Результати експериментів з трансформерними моделями наведено в табл. 9.

Таблиця 9

Результати експериментів з трансформерними моделями

Модель	Датасет	F1-score	Час навчання	Інференс (відг/с)	GPU пам'ять (GB)
BERT-base	IMDb	0.921	6.8 год	58	8.4
	Amazon	0.884	13.2 год	54	8.6
	Yelp	0.859	10.5 год	56	8.5
RoBERTa-base	IMDb	0.927	7.1 год	52	9.2
	Amazon	0.891	13.8 год	49	9.4
	Yelp	0.864	11.0 год	51	9.3
DistilBERT	IMDb	0.908	3.4 год	112	4.8
	Amazon	0.872	6.7 год	108	5.0
	Yelp	0.847	5.3 год	110	4.9

Четверта серія експериментів присвячена доменній адаптації трансформерних моделей. Для кожного датасету створено domain-adapted версію BERT шляхом continued pre-training на нерозмічених відгуках з відповідного домену. Domain-adapted BERT для IMDb навчався додатково на корпусі з 200,000 нерозмічених відгуків про фільми. Continued pre-training тривав 2 епохи з masked language modeling задачею. Загальний час навчання склав 11 годин (4.2 години додаткового навчання + 6.8 годин fine-tuning). Результати – F1-score 0.938, що на 1.7% краще за vanilla BERT та на 1.1% за RoBERTa. Детальний аналіз виявив, що domain-adapted модель значно краще справляється з оцінкою технічних аспектів фільмів. Відгуки з описами кінематографії класифікувалися з точністю 94.3% проти 91.8% у стандартного BERT. Domain-adapted BERT для Amazon Electronics пройшов додаткове навчання на 500,000 нерозмічених відгуках про електронні пристрої. F1-score покращився до 0.907, що на 2.3% краще за стандартний BERT – найбільше покращення серед усіх доменів. Це пояснюється високою специфічністю технічної лексики. Domain-adapted BERT для Yelp навчався на 400,000 нерозмічених ресторанних відгуків. Результати – F1-score 0.882, що на 2.3% краще за vanilla BERT. Модель навчилася краще розпізнавати специфічну лексику, включно з розмовними виразами та метафорами. П'ята серія експериментів спрямована на оцінку здатності моделей узагальнювати знання між різними доменами. Кожна модель, навчена на одному датасеті, тестувалася

на двох інших без додаткового донавчання. Тестування моделі, навченої на IMDb, на інших доменах показало очікуване падіння точності. На Amazon Electronics F1-score склав 0.782 для BERT (падіння 13.9 п.п.) та 0.698 для XGBoost (падіння 19.3 п.п.). На Yelp результати були ще гіршими – 0.751 для BERT. Трансформерні моделі виявилися більш робастними до domain shift, демонструючи на 6–8% менше падіння порівняно з класичними методами. Тестування моделі Amazon на інших доменах дало неоднорідні результати. На IMDb F1-score для BERT становив 0.858 (падіння 2.6 п.п.), що значно менше. Це пояснюється тим, що модель, навчена на коротких технічних відгуках, може адаптуватися до довгих рецензій краще, ніж навпаки. Domain-adapted моделі виявили значно кращу робастність при крос-доменному тестуванні. Domain BERT для IMDb показав на Amazon F1-score 0.814 замість 0.782 у vanilla BERT – покращення на 3.2 п.п. Візуалізація domain shift ефекту демонструє значну деградацію при крос-доменному тестуванні, особливо для класичних методів. Детальний аналіз помилок класифікації проводився для найкращих моделей на датасеті Amazon Electronics як найскладнішому через тризначну класифікацію та незбалансованість класів. Confusion matrix для XGBoost виявила типові проблеми класичних методів. Найбільша плутанина спостерігалася між нейтральними та позитивними класами – 24.3% нейтральних відгуків помилково класифікувалися як позитивні. BERT продемонстрував суттєво кращу диференціацію класів. Плутанина між нейтральними та позитивними зменшилася до 14.8%, а між негативними та нейтральними – до 9.2%. Модель навчилася розуміти тонкі відмінності в інтенсивності висловлювань. Domain-adapted BERT показав найкращі результати з плутаниною між нейтральними та позитивними лише 11.3%. Мануальний аналіз 200 помилково класифікованих відгуків виявив кілька типових категорій: сарказм та іронія – 23% помилок; змішані відгуки з декількома аспектами – 31%; порівняння з очікуваннями – 18%; технічні описи без явної оцінки – 15%; культурні та контекстуальні відсилання – 13%. Порівняльний аналіз обчислювальних витрат критично

важливий для вибору оптимальної моделі в продакшн-середовищі. Час навчання варіюється від секунд для Naive Bayes до годин для трансформерних моделей. Швидкість інференсу визначає throughput системи – Naive Bayes обробляє 612 відгуків на секунду, XGBoost – 198, BiLSTM – 63, BERT – 54, DistilBERT – 108. Нижче представлена комплексна візуалізація компромісу між точністю та обчислювальними витратами. Узагальнюючи результати всіх серій експериментів, можна сформулювати чіткі рекомендації щодо вибору оптимальної моделі. Для швидкого прототипування та MVP найкраще підходить XGBoost з TF-IDF векторизацією. Модель навчається за 40–90 хвилин, досягає точності 0.85–0.89 F1-score та забезпечує швидкість інференсу 198–243 відгуків/секунду. Для продакшн-систем середнього масштабу (10,000–100,000 відгуків/день) оптимальним є DistilBERT. Модель забезпечує F1-score 0.87–0.91, що лише на 1–2% нижче за повний BERT, при подвоєній швидкості інференсу та вдвічі меншому споживанні GPU пам'яті. Для критичних застосувань з вимогами максимальної точності слід використовувати Domain-adapted BERT або RoBERTa. Domain-adapted BERT досягає найвищої точності 0.88–0.94 F1-score завдяки спеціалізованому pre-training. Для систем з обмеженими ресурсами найкращий вибір – SVM з лінійним ядром або TextCNN. SVM працює на CPU з точністю 0.81–0.88 F1-score без потреби у GPU. Для мультидоменних систем рекомендується RoBERTa, що демонструє найменше падіння при domain shift (5.7% в середньому проти 15.5% у BERT навченого на одному домені). Таким чином, проведені експериментальні дослідження підтвердили основні гіпотези про перевагу трансформерних моделей для класифікації тональності відгуків при збереженні практичної цінності класичних методів для певних сценаріїв. Результати демонструють, що вибір оптимальної моделі має базуватися на балансі між вимогами до точності, доступними обчислювальними ресурсами, обсягом даних та специфікою бізнес-завдання. Комплексний аналіз результатів дозволяє сформулювати систематичні рекомендації для вибору моделі. Ключові критерії включають: пріоритет точності чи швидкості; доступність GPU; обсяг

розмічених даних; необхідність інтерпретованості; критичність латентності; вартісні обмеження. Матриця рекомендацій: Високий пріоритет точності (F1 > 0.90): Domain BERT або RoBERTa при наявності GPU та великих даних; BERT або DistilBERT при середніх даних; максимум XGBoost без GPU. Баланс точності та швидкості: DistilBERT для real-time інференсу; BERT для batch обробки; XGBoost для дуже великих обсягів. Обмежені ресурси: XGBoost без GPU; SVM при малому бюджеті; TextCNN або quantized DistilBERT для мобільних пристроїв. Систематичне порівняння восьми моделей на трьох датасетах дозволяє сформулювати наступні ключові висновки. По-перше, спостерігається чітка кореляція між складністю моделі та точністю класифікації. Трансформерні архітектури (F1-score 0.86–0.91) перевершують нейромережеві моделі (0.84–0.90), які кращі за класичні методи (0.76–0.89). Проте різниця у точності не перевищує 2–4 п.п., тоді як обчислювальні витрати відрізняються на порядки. По-друге, доменна адаптація демонструє стабільний позитивний ефект з приростом точності 1.7–2.3%. Покращення найбільш помітне на технічних доменах з специфічною термінологією. По-третє, стабільність результатів є не менш важливою за абсолютну точність. Трансформерні моделі демонструють найнижчу variance (std=0.006–0.009) та найшвидшу конвергенцію. По-четверте, калібрація ймовірностей значно відрізняється. Domain BERT (ECE=0.035) та RoBERTa (ECE=0.038) демонструють найкращу калібрацію, тоді як Naive Bayes (ECE=0.127) систематично overconfident. По-п'яте, економічний аналіз виявляє нелінійну залежність між інвестиціями у модель та ROI. Для сценаріїв з високою вартістю помилок інвестиція у Domain BERT окупається багаторазово. Для застосувань з низькою критичністю простіші моделі забезпечують кращий економічний ефект. Таким чином, порівняльний аналіз підтверджує, що оптимальний вибір моделі має базуватися на комплексній оцінці технічних характеристик, економічних факторів та специфічних вимог бізнес-контексту, а не виключно на метриці точності класифікації.

Висновки. У сучасних умовах стрімкого зростання обсягів цифрових даних автоматизація аналізу відгуків користувачів в корпоративному маркетингу набуває критичного значення для бізнесу. Дослідження 2024-2025 років фіксують, що понад 92% споживачів вивчають онлайн-відгуки перед здійсненням покупки, компанії отримують десятки тисяч текстових відгуків щодня. Традиційний ручний аналіз характеризується високими витратами, суб'єктивністю інтерпретації та неможливістю масштабування. Застосування методів природної мовної обробки відкриває нові можливості для ефективного аналізу великих обсягів текстових даних та формування об'єктивних висновків про настрої користувачів.

Проведений комплексний аналіз проблематики показав, що еволюція методів NLP від статистичних підходів до трансформерних моделей демонструє суттєвий прогрес. Статистичні методи (BoW, TF-IDF) забезпечують точність 70-75%, векторні представлення (Word2Vec, GloVe) – 78-84%, рекурентні архітектури (LSTM, BiLSTM) – 84-90%, трансформерні моделі (BERT, RoBERTa) досягають 90-96% точності.

Встановлено, що класичні методи забезпечують швидку обробку при обмеженій точності, нейромережеві моделі досягають балансу між точністю та складністю, а трансформерні архітектури демонструють найвищу точність при значних обчислювальних вимогах корпоративного маркетингу.

Розроблено теоретичні та методичні основи використання NLP для аналізу відгуків, що включають комплексний підхід до попередньої обробки текстових даних з урахуванням специфіки користувацьких текстів. Правильна попередня обробка мови в корпоративному маркетингу підвищує якість аналізу на 8-15%. Створено набір моделей різної складності: класичні методи з F1-score 0.79-0.82, нейромережеві архітектури з F1-score 0.87-0.90, трансформерні моделі з F1-score 0.91-0.94, domain-adapted BERT з F1-score 0.94.

Розроблено комплексну методику оцінювання ефективності, що враховує точність класифікації, обчислювальну ефективність, статистичну значущість результатів та робастність до різних типів даних.

Експериментальне порівняння одинадцяти моделей на базі 150 000 відгуків підтвердило, що domain-adapted BERT досягає найвищої точності класифікації (F1 = 94.2%), що на 15 відсоткових пунктів перевищує результати класичних методів. Критичним результатом стало виявлення оптимального балансу між точністю та швидкістю. DistilBERT продемонстрував найкраще співвідношення якості (F1=91.1%) та продуктивності (890 відгуків/сек), що робить його універсальним рішенням для більшості комерційних застосувань. Система забезпечує прискорення аналізу у 120-150 разів порівняно з ручною обробкою при збереженні високої точності та об'єктивності оцінок.

Аналіз використання обчислювальних ресурсів показав суттєві відмінності між моделями. Класичні методи вимагають мінімальної пам'яті (0.3-1.8 GB) та швидко навчаються, трансформерні моделі потребують 1.4-6.1 GB пам'яті. Час навчання варіює від 12 хвилин для Naive Bayes до 24.7 години для DA-BERT.

Практичне тестування на реальних бізнес-процесах підтвердило, що система обробляє 1000 відгуків за 1.1 секунди, тоді як команді аналітиків потрібно 2-3 години. Узгодженість оцінок між моделлю та експертами становить 94.8% для DA-BERT, що перевищує міжекспертну узгодженість (91.2%).

Економічна ефективність впровадження виявилася значною. Автоматизація аналізу 100 000 відгуків на місяць скорочує витрати на персонал на $12 000-15 000 при вартості інфраструктури близько $800-1200 щомісяця. Чиста економія становить понад $10 000 на місяць, окупність досягається за 4-6 місяців. Система виявила на 45% більше критичних проблем порівняно з вибірковим ручним аналізом. За три місяці пілотного впровадження на платформі з 2 млн користувачів компанія покращила показники задоволеності клієнтів на 18%.

Сформульовано практичні рекомендації щодо впровадження для організацій різного масштабу. Для стартапів з обсягом до 10 000 відгуків доцільні класичні методи (витрати $50-100/міс). Середні компанії (10 000-100 000 відгуків) отримають максимальну вигоду від DistilBERT (витрати $200-400/міс). Великі платформи з мільйонами відгуків мають інвестувати в DA-BERT або RoBERTa

(витрати $1000-2000/міс), оскільки навіть покращення точності на 2% забезпечує суттєві переваги при масштабі.

Розроблено шестиетапний процес впровадження: аналіз вимог та підготовка даних; вибір та fine-tuning моделі; валідація та тестування; інтеграція з існуючими системами; pilot-тестування на реальних даних; масштабування та continuous learning.

Показано важливість організаційних аспектів – формування кросфункціональної команди, навчання персоналу, управління очікуваннями стейкхолдерів. Розроблені блок-схеми вибору моделі та етапів впровадження надають практичний інструментарій для прийняття рішень.

Методичне забезпечення дослідження базується на сучасному технологічному стеці Python 3.10+ з бібліотеками NLTK, spaCy, Hugging Face Transformers, PyTorch, TensorFlow. Систематичний підхід до організації експериментів через Weights & Biases, Jupyter Notebooks та Docker забезпечує відтворюваність результатів та ефективну ітерацію при розробці моделей.

Практична цінність результатів полягає у можливості безпосереднього застосування для створення систем автоматизованого аналізу відгуків у електронній комерції, сервісних компаніях, розробці програмного забезпечення, корпоративному маркетингу та аналітиці. Впровадження дозволяє скоротити час аналізу в 120-150 разів, забезпечити об'єктивність результатів та виявляти на 45% більше критичних проблем.

Рекомендації щодо вибору оптимальних моделей дозволяють компаніям різного масштабу впроваджувати ефективні системи з урахуванням обмежень ресурсів та вимог до точності.

Результати дослідження підтверджують, що сучасні методи NLP, особливо трансформерні архітектури з доменною адаптацією, є потужним інструментом для автоматизації аналізу відгуків користувачів.

Комплексний підхід, що поєднує правильну попередню обробку даних, вибір адекватних моделей відповідно до специфіки завдання, ретельну оцінку якості та

структуроване впровадження, дозволяє досягти точності, що наближається до людської, при значно вищій швидкості обробки в корпоративному маркетингу.

Розроблені практичні рекомендації та методики впровадження створюють основу для масштабного застосування NLP-технологій у бізнес-процесах організацій різних галузей та масштабів, забезпечуючи суттєві конкурентні переваги через об'єктивний та всебічний аналіз відгуків користувачів в корпоративному маркетингу.

References:

Ковпака Анастасія, Мосійчук Ірина, Клімова Інна. (2021). Інструменти інноваційного маркетингу в системі управління підприємством. *Економіка. Управління. Інновації,* Вип. 2(29). 15 с. 2021. https://doi.org/10.35433/ISSN2410-3748-2021-2(29)-4.

Артур Струнгар. (2024). Вплив штучного інтелекту на стратегії цифрового маркетингу: поточні можливості та перспективи розвитку. *Економіка та суспільство,* Вип. 62. 10 с. 2024. https://doi.org/10.32782/25240072/-2024-62-160.

Бутенко В.М., Тоюнда А.І. (2022). Формування маркетингової стратегії підприємства. *Підприємництво та інновації*. (24), С. 61-67. 2022. https://doi.org/10.32782/2415-3583/24.10.

Ірина Вікторівна Іванова, Тетяна Михайлівна Боровик, Таміла Григорівна Залозна, Альона Юріївна Руденко. (2025). Використання штучного інтелекту в маркетингу. *Маркетинг і цифрові технології*. Вип. 9(1). С. 103-116. 2025. https://doi.org/10.15276/mdt.7.2.2023.3.

Joseph T. (2024). Natural Language Processing (NLP) for Sentiment Analysis in social media. *International Journal of Computing and Engineering*: Vol. 6, no. 2, pp. 35–48. 2024. https://doi.org/10.47941-/ijce.2135.

Mahander Kumar, Lal Khan, Hsien-Tsung Chang (2025). Evolving techniques in sentiment analysis: a comprehensive review. *PeerJ Computer Science*. 61 p. 2025. https://doi.org/10.7717/peerj-cs.2592.

Zhang W., Deng Y., Liu B., Pan S., Bing L. (2024). Sentiment Analysis in the Era of Large Language Models: A Reality Check. *Findings of the Association for Computational Linguistics*. pp. 881–906. 2024. https://doi.org/-10.48550/arXiv.2305.15005.

Viktoria Adamyk, Oleksandr Ivanovskyi. (2025). Use of artificial intelligence in the marketing system: current trends and challenges. *Herald of Economics*. no. 1, pp. 230-43. 2025. https://doi.org/10.35774/-visnyk2025.01.230.

Hnoievyi V. H., Koren O. M. (2025). Modern trends in digital marketing and their influence on the formation of a marketing strategy. *Akademichnyi ohliad*. no. 1(54), pp. 49–55. 2025. https://doi.org/10.32342/2074-5354-2021-1-54-5.

Калініченко Андрій Петрович. (2024). Вплив штучного інтелекту на стратегії просування компаній. *Економіка та суспільство*. Вип. 69. 7 с. 2024. https://doi.org/10.32782/2524-0072/2024-69-67.