

**МІНІСТЕРСТВА ОСВІТИ І НАУКИ УКРАЇНИ  
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ФАКУЛЬТЕТ СОЦІОЛОГІЇ ТА УПРАВЛІННЯ**

**КАФЕДРА ПОЛІТОЛОГІЇ**

**КВАЛІФІКАЦІЙНА РОБОТА**

**магістра**

на тему **ВИКОРИСТАННЯ ТЕХНОЛОГІЇ BIG DATA  
В ПОЛІТИЧНИХ ПРОЦЕСАХ**

Виконав: студент групи 8.0529  
спеціальності 052 Політологія  
освітньої програми 052 Політологія  
Демидов Денис Юрійович

Керівник: д.політ.н., доцент Хорішко Л. С.

Рецензент: к.політ.н., доцент Кальцева С. І.

Запоріжжя  
2020

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

Факультет соціології та управління  
Кафедра політології  
Рівень вищої освіти магістр  
Спеціальність 052 Політологія  
Освітня програма 052 Політологія

**ЗАТВЕРДЖУЮ**

Завідувач кафедри Є.Г. Цокур \_\_\_\_\_  
«10» грудня\_2020 року

**З А В Д А Н Н Я  
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

Демидову Денису Юрійовичу

1. Тема роботи: Використання технології Big Data в політичних дослідженнях керівник роботи Хорішко Лілія Сергіївна, д.п.н., доцент, затверджені наказом ЗНУ від «25» травня 2020 року № 605-с
2. Строк подання студентом роботи: 10 грудня 2020 року.
3. Вихідні дані до роботи Квіт С. Масові комунікації : підручник / С. Квіт . — 2-ге вид., випр. і допов. — Київ : Вид. дім "Києво-Могилянська академія", 2018. — 350 с.; Эрл Томас Основы Big Data. Концепции, алгоритмы и технологии / Томас Эрл, Ваджид Хаттак, Пол Булер ; [пер. с англ. А. Гладуна]. — Дніпро : Баланс Бізнес Букс, 2018. — 308 с.; Власенко Р. В. Концепція Big Data в Україні: перспективи застосування в державних органах / Р. В. Власенко // Держава та регіони. Сер. Держ. управління. — 2017. — Вип. 4. — С. 97-100; Матеріали міжнародного наукового симпозіуму. BIG DATA ANALYTICS: моделювання та інформаційні технології. Київ, 20 березня 2019 р. URL : <https://knute.edu.ua/file/NjY4NQ==/a8ee205bb8f2cd8272b7fe9668eb5a35.pdf>
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
  1. Розглянути сучасний стан наукової розробки тематики роботи.
  2. Проаналізувати основні поняття, що будуть використані під час дослідження.
  3. Дослідити існуючі методології дослідження використання технології Big Data в політичних процесах.
  4. Простежити історію виникнення та розвитку Big Data у світі.
  5. Узагальнити зміст та структурні складові технології Big Data.
  6. Розглянути юридичні та етичні аспекти використання

технології Big Data та даних, потрібних для її функціонування. 7. Вивчити зарубіжний досвід в застосування технології Big Data в політичній сфері. 8. Дослідити досвід України в залученні Великих Даних до політичних процесів. 9. Описати перспективи подальшого розвитку взаємодії технології Big Data та політичних процесів в Україні та світі.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) 2 діаграми, 3 зображення.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
Розділ 1	Хорішко Л. С., доцент кафедри політології	18.06.20	18.06.20
Розділ 2	Хорішко Л. С., доцент кафедри політології	23.09.20	23.09.20
Розділ 3	Хорішко Л. С., доцент кафедри політології	29.10.20	29.10.20

7. Дата видачі завдання 26 травня 2020

### КАЛЕНДАРНИЙ ПЛАН

Пор. №	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1.	Вибір та формулювання теми роботи	Березень 2020	<i>виконано</i>
2.	Опрацювання наукових джерел	Квітень 2020	<i>виконано</i>
3.	Робота над вступом	Травень 2020	<i>виконано</i>
4.	Робота над першим розділом	Червень 2020	<i>виконано</i>
5.	Робота над другим розділом	Вересень 2020	<i>виконано</i>
6.	Проведення дослідження	Жовтень 2020	<i>виконано</i>
7.	Робота над третім розділом	Листопад 2020	<i>виконано</i>
8.	Систематизація висновків	Грудень 2020	<i>виконано</i>

Студент \_\_\_\_\_ Д. Ю. Демидов \_\_\_\_\_  
 (підпис) (ініціали та прізвище)

Керівник роботи \_\_\_\_\_ Л. С. Хорішко \_\_\_\_\_  
 (підпис) (ініціали та прізвище)

**Нормоконтроль пройдено**

Нормоконтролер \_\_\_\_\_ Ю. Г. Мальована

**Декларація**  
**академічної доброчесності**  
**здобувача ступеня вищої освіти**  
**Запорізького національного університету**

Я, Демидов Денис Юрійович, студент 2 курсу магістратури, денної форми навчання, факультету соціології та управління, спеціальність 052 Політологія, адреса електронної пошти dendesoft@gmail.com:

– підтверджую, що написана мною кваліфікаційна робота на тему «Використання технології Big Data в політичних процесах» відповідає вимогам академічної доброчесності, опублікованих на сайті Запорізького національного університету станом на 15 грудня 2020 року та не містить порушень, що визначені у ст. 42 Закону України «Про освіту», зі змістом яких ознайомлений;

– заявляю, що надана мною для перевірки електронна версія роботи є ідентичною її друкованій версії;

– згоден на перевірку моєї роботи на відповідність критеріям академічної доброчесності у будь-який спосіб, у тому числі за допомогою інтернет-системи, а також на архівування моєї роботи в базі даних цієї системи зі збереженням за мною всіх авторських прав.

15.12.2020

ПІБ

Науковий керівник,

Доктор політичних наук, доцент кафедри політології Л. С. Хорішко

15.12.2020

## ЗМІСТ

ВСТУП .....	3
РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ BIG DATA В ПОЛІТИЧНІЙ СФЕРІ .....	7
1.1. Сучасний стан наукової розробки .....	7
1.2. Аналіз основних понять дослідження .....	12
1.3.    Методологія    дослідження    використання    технології Big Data в політичних процесах .....	16
РОЗДІЛ 2. ОСОБЛИВОСТІ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ BIG DATA В ПОЛІТИЧНИХ ПРОЦЕСАХ.....	20
2.1. Технологія Big Data: історія розвитку .....	20
2.2. Зміст та структурні складові технології Big Data .....	29
2.3. Юридичні та етичні аспекти використання технології Big Data в політичних процесах.....	36
РОЗДІЛ 3. ПРАКСЕОЛОГІЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ BIG DATA В УМОВАХ СУЧАСНОЇ ПОЛІТИЧНОЇ ДІЙСНОСТІ.....	43
3.1. Зарубіжний досвід використання Big Data.....	43
3.2. Досвід використання Big Data в Україні .....	50
3.3. Перспективи використання технології в Україні .....	58
ВИСНОВКИ.....	65
СПИСОК ЛІТЕРАТУРИ.....	69

## РЕФЕРАТ

Кваліфікаційна робота складається з 86 сторінок, 117 позицій у списку літератури.

BIG DATA, ХМАРНІ ОБЧИСЛЕННЯ, ШТУЧНИЙ ІНТЕЛЕКТ, ПОЛІТИЧНІ ПРОЦЕСИ, АНАЛІТИКА, СТАТИСТИЧНИЙ АНАЛІЗ, ПОЛІТИЧНИЙ МАРКЕТИНГ, ВИБОРЧА КАМПАНІЯ.

*Мета дослідження:* дослідити використання технології Big Data в політичних процесах.

*Об'єкт дослідження:* політичні процеси як.

*Предмет дослідження:* аспекти використання технології Big Data в сучасних політичних процесах.

*Методи дослідження:* порівняльний, історичний, системний та структурно-функціональний.

*Основна гіпотеза:* використання Big Data дає можливість значно ефективніше та з меншими витратами проводити дослідження та реалізовувати певні кроки в політичній сфері.

*Додаткова гіпотеза:* у перспективі Big Data стануть основою для досліджень в політичній сфері, але, в той же час, може суттєво змінитись підхід до використання певних масивів даних.

*Висновки:*

- Технології Big Data постійно та динамічно розвиваються паралельно з об'ємами інформації, які створює людство.

- Їх використання з кожним днем стає все більш доцільним економічно та технічно.

- З огляду на досвід в Україні та світі можна сказати, що використання Big Data в політичній та інших сферах буде лише посилюватись, як і результати його впливу.

## ABSTRACT

This thesis consists of 86 pages, 117 items in the references list.

BIG DATA, CLOUD COMPUTING, ARTIFICIAL INTELLIGENCE, POLITICAL OPPORTUNITY, ANALYTICS, STATISTICAL ANALYSIS, POLITICAL MARKETING, ELECTION CAMPAIGN.

*The aim of the research* is to study the use of Big Data technology in political processes.

*The research object* of our master's degree thesis are political processes.

*The research subject* are the aspects of using the Big Data technology in modern policy processes.

*The research methods* are comparative, historical, systemic and structural-functional.

*The main hypothesis* is that the use of Big Data technology makes it possible to carry out research and implement certain policy measures more efficiently and at a lower cost.

*An additional hypothesis:* Big Data will eventually become the basis for research in the political sphere, but at the same time, the approach to the use of certain data sets may change significantly.

Conclusions:

- The Big Data technologies are constantly and dynamically developing in parallel with the amount of information that humanity creates.
- Their use is becoming more economical and technical every day.
- Taking into account the Ukrainian and the world experience, it is possible to underline that the use of Big Data in political and other spheres will increase, as well as the results of its influence.

## ВСТУП

**Актуальність дослідження.** «Хто володіє інформацією, той володіє світом». Ця цитата Натана Ротшильда, англійського банкіра, відома сьогодні майже всім. Дійсно, адже ми живемо в епоху, коли інформація є одним із головних джерел влади та впливу.

Але її об'єми настільки збільшилися та продовжують зростати в геометричній прогресії. На противагу ціна обчислювальної потужності знижується, а хмарні технології стають доступнішими. Це призводить до того, що обробка та аналіз великих масивів даних стає важливим інструментом для гравців політичного процесу, починаючи від держави до певних політичних постатей [48].

Світ постав на порозі епохи Big Data, коли кожен стикається з ними щоденно. Кожної хвилини мільярди людей у всьому світу не просто натискають кнопки своїх смартфонів та набирають трильйони символів тексту. Вони генерують величезний обсяг інформації, залишають цифровий слід. Усе це можна легко акумулювати та проаналізувати, тим самим дізнавшись про неї значно більше, ніж про неї знають її найближчі родичи та вона сама.

Кожна особа може обманювати оточуючих, владу, близьких та самого себе, та свої дії — ніяким чином. Саме ця інформація й лягає в основу Big Data, які нам, як дослідникам, можна використовувати в найрізноманітніших сферах, а особливо це цікаво в політичних.

Наразі вони зустрічаються як в комерційному, так і некомерційному середовищі. Комерційні компанії майже в промислових масштабах використовують Big Data для отримання інформації про клієнтів, процесах, конкурентів і товари. Багато компаній застосовують цю інформацію, щоби справити хороше враження на користувача, для організації перехресних і додаткових продажів, а також для персоналізації своїх послуг, щоби ті були «ближчими» до кінцевого користувача.



Урядові організації також добре обізнані про цінність даних. Багато урядові організації не тільки використовують власних аналітиків для пошуку цінної інформації, але і викладають свої результати у відкритий доступ. Дані можуть використовуватися для глибокого аналізу так і побудови додатків для їх керування та дослідження. Data.gov.ua — база відкритих даних державних органів України (аналог Data.gov в США) — усього лише один із прикладів.

Фахівці з Big Data в урядових організаціях працюють над найрізноманітнішими проєктами, від виявлення випадків шахрайства та інших видів злочинної діяльності до оптимізації фінансування проєктів та проведення перепису населення.

Добре відомий приклад такого роду було надано Едвардом Сноуденом, який опублікував внутрішні документи Агентства національної безпеки США і Британського центру урядового зв'язку, з яких явно випливає, що Big Data застосовувалися для стеження за мільйонами громадян по всьому світу.

Ці організації зібрали 5 мільярдів записів даних з таких поширених додатків і сервісів, як Microsoft (Hotmail), Google (Google Mail), Yahoo!, Facebook, YouTube, Skype, AOL, Apple та Paltalk [52]. Потім методи Data Science були застосовані для фільтрації інформації.

Неурядові організації теж мають досвід використання Big Data. Вони використовують ці методи для залучення коштів і відстоювання власних цілей. Наприклад, у низки міжнародних проєктів, що займаються журналістськими розслідуваннями в політичній, економічній та війсьній сферах, як наприклад Bellingcat, активно використовують аналітику Big Data у своїй діяльності [107].

У той час, коли весь світ намагається максимально повно використовувати аналітику Big Data, а ІТ-спеціалісти (в тому числі у сфері обробки Big Data) з України посідають перші місця серед найпрофесійніших у світі [56], ми маємо детально вивчати можливість її використання, у тому числі в Україні і в досліджуваній нами політичній сфері.

Наше дослідження дасть змогу зрозуміти, яким чином розвивались технології Big Data і на якому етапі знаходиться сьогодні; як з ними працюють по всьому світі та в Україні і можливості застосування в подальшому в різних політичних сферах.

*Об'єкт дослідження:* політичні процеси як сфера використання технології Big Data.

*Предмет дослідження:* аспекти використання технології Big Data в сучасних політичних процесах.

*Мета роботи:* дослідити основні аспекти використання технології Великих Даних в політичній сфері як одного з напрямків поєднання із обчислювальними технологіями і машинного навчання.

Для досягнення поставленої мети ми визначили наступні завдання:

- розглянути сучасний стан наукової розробки тематики нашої роботи;
- проаналізувати основні поняття, що будуть використані під час цього дослідження;
- дослідити існуючі методології дослідження використання технології Big Data в політичних процесах;
- простежити історію виникнення та розвитку Big Data у світі;
- узагальнити зміст та структурні складові технології Big Data;
- розглянути юридичні та етичні аспекти використання технології Big Data та даних, потрібних для її функціонування;
- вивчити зарубіжний досвід в застосування технології Big Data в політичній сфері;
- дослідити досвід України в залученні Великих Даних до політичних процесів;
- описати перспективи подальшого розвитку взаємодії технології Big Data та політичних процесів в Україні та світі.

*Основна гіпотеза:* використання Big Data дає можливість значно ефективніше та з меншими витратами проводити дослідження та реалізовувати певні кроки в політичній сфері.

*Додаткова гіпотеза:* в перспективі Big Data стануть основою для досліджень в політичній сфері, але, в той же час, може суттєво змінитись підхід до використання певних масивів даних.

*Структура роботи* підпорядкована меті та завданням дослідження. Кваліфікаційна робота складається із переліку аббревіатур, вступу, 3 розділів (8 підрозділів), висновків, списку використаних джерел та літератури (117 найменувань), додатків, реферату на українській та англійській мові.

Обсяг основної частини роботи складає 68 сторінок машинописного тексту, загальний обсяг бакалаврської роботи — 86 сторінок.

## РОЗДІЛ 1

### ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ BIG DATA В ПОЛІТИЧНІЙ СФЕРІ

#### 1.1. Сучасний стан наукової розробки

Оскільки тема використання технології Big Data, зокрема в політичних процесах хоч і нова, але досить актуальна, її дослідженням займається багато вчених як в Україні, так і за кордоном.

Кандидат наук з державного управління Власенко Р. В. приділяє увагу дослідженню концепції Big Data, зокрема, можливостям її застосування у державних органах. Він розглядає місце та тенденції розвитку Big Data в Україні, аналізує передумови та проблеми її застосування в інтересах держави.

В своїх працях він приходиться висновків, що сьогодні Big Data в Україні — це молодий напрям досліджень, але при цьому країна має достатній технологічний, інфраструктурний, людський та правовий потенціал для подальшого розвитку цієї концепції як на рівні окремих суб'єктів господарювання, так і на загальнодержавному рівні.

В той самий час, Власенко відмічає, що сьогодні в Україна існує проблема із підготовкою якісних кадрів в цій сфері.

При цьому, він підкреслює думку, що ці технології здатні вирішувати практично будь-які задачі, пов'язані з попередженням та виходом із кризових явищ, а тому держава повинна працювати в цьому напрямку.

Нікерсон Д. з університету Нотр-Дам та Роджерс Т. з Гарвардської школи Кеннеді працюють в сфері використання Big Data саме в політичних кампаніях.

У своєму дослідженні вони спираються на те, що сучасні компанії розробляють бази даних детальної інформації про громадян для інформування виборчої стратегії та для керівництва тактичними зусиллями. Вони зауважили, що незважаючи на сенсаційні повідомлення про цінність даних про окремих

споживачів, найбільшої цінності інформаційні кампанії набувають завдяки поведінці та прямим реакціям самих громадян.

Вони приходять до висновків, що аналітики даних кампанії розробляють моделі, використовуючи цю інформацію, для формування прогнозів на індивідуальному рівні щодо ймовірності здійснення громадянами певної політичної поведінки, підтримки кандидатів та зміни їх підтримки, залежно від націленості на конкретні втручання в кампанію.

Використання цих прогнозних оцінок різко зросло з 2004 року, і їх залучення може принести значну перевагу політичним компаніям, які їх використовують.

У той же час їх широке використання стає менш ефективним, коли на одне електоральне поле виходить декілька гравців, а тому організаціям було б вигідно розділити електорат, щоб не дублювати зусилля, проте правові та політичні обмеження виключають таку можливість.

Проблемою етичних та правових аспектів використання персональних даних, в тому числі з соціальних медіа займається Щеглова І. А. Її роботи присвячені аналізу етичних і правових аспектів досліджень з використанням даних із соціальних медіа для прогнозування економічних, політичних, соціальних та інших процесів. Спираючись на кейси, що мають місце в дослідницькій практиці, вона аналізує наслідки порушень, що виникають в ході роботи з даними із соціальних мереж та пропонує шляхи подолання етико-правових бар'єрів у використанні таких даних.

В результатах її праці говориться про те, що великі дані стали одним з перспективних напрямків в науковому світі, але їх використання відкриває не тільки нові можливості перед вченими, а й ставить серйозні етичні та юридичні питання, на які поки не знайдені правильні відповіді.

Відзначає, що кілька років тому, коли вчені тільки починали втілювати в життя свої проекти, можна було легко завантажити дані з Facebook, не запрошуючи дозволу користувачів на доступ до їхнього профілю, і використовувати отриману інформацію для будь-яких цілей.

Сьогодні вже не можна так просто отримати доступ до даних на соціальних платформах.

Крім того, наукові журнали стали серйозніше відноситися до етики досліджень, що проводяться на основі таких даних.

Серьогін В. А., доктор юридичних наук Харківського університету ім. В. Н. Каразіна висвітлює проблемні аспекти забезпечення недоторканності приватного життя в умовах динамічного розвитку Big Data як потужної форми інтелектуального аналізу даних. В результаті дослідження отримує висновки про необхідність стимулювати бізнес-структури до впровадження нових бізнес-моделей, заснованих на розширенні прав споживачів. Пропонує доповнити чинне інформаційне законодавство положеннями, що стосуються особливостей обробки персональних і неперсональних даних з використанням БД, прийняти відповідні професійні кодекси поведінки в сфері обробки персональних даних

Сакулін С. робить акцент на тому, яку загрозу маніпулюванням масовою свідомістю несе в собі використання Big Data в політичних процесах.

Він дійшов висновків, що методологічно вивірена система Big Data відкриває нові можливості в дослідженнях психометрії, що, в свою чергу, створює нову платформу для маніпулювання суспільною свідомістю, яка використовується сьогодні для легітимації політичних груп.

Застосування психометричного портретування інтернет-користувачів електоральних груп спрямоване на виявлення тих громадян, хто не визначився з позицією та надають їм індивідуальні політичні меседжі. Така таргетована реклама може бути спрямована на формування «правильного» думки в питаннях зовнішньої і внутрішньої політики.

Сучасне суспільство стрімко занурюється в цифрову середу, залишаючи там все більше «слідів», в той же час обчислювальні алгоритми стають все більш досконалішими, через що а ні про яку приватність не може бути й мови.

Майбутнє світової політики знаходиться в руках психометрії, технологій Big Data і таргетованого контенту.

Свірідова Т. молода науковець з Національного університету «Львівська політехніка» досліджує практичні основи необхідності використання нових технологій в контексті проведення політичних кампаній. Вона розглядає вплив використання Big Data на процес виборчих кампаній.

Балашов Б. М. досліджує поняття Big Data та його походження, базові проблеми обробки та використання в політичній сфері, роблячи акцент на аналізі передвиборчої кампанії в США 2017 року.

Кандидат економічних наук Інституту економіки промисловості НАН України Чекіна В. Д. та її колега з Відділення економіки Президії НАН України Крізев С. І. працюють над аналізом впливу Big Data та аналітики на реальний і державний сектори економіки України, вивчають проблеми їх оцінювання з використанням статистичних методів.

Володенков С. В. присвятив свою роботу проблематиці застосування технологій Big Data в сучасній політичній практиці. Автор акцентував свою увагу на потенційних викликах і загрозах, пов'язаних з використанням цифрових масивів інформації в процесах суспільно-політичного розвитку.

Визначив ключові, на думку автора, виклики та загрози, пов'язані з формуванням суспільства цифрової нерівності, що характеризується активним застосуванням технологій цифрової маніпуляції і пропаганди на основі використання персональних даних.

Він констатує, що на сьогоднішній день існує значна кількість викликів і загроз у сфері суспільно-політичного розвитку, пов'язаних із застосуванням технологій Big Data.

Так, наприклад, існують також виклики і загрози, пов'язані з трендом на анонімізація Інтернету як реакцією з боку онлайн-користувачів на спробу «приватизації» персональних даних, використанням Big Data для втручання в суспільно-політичні процеси малих країн з боку технологічно розвинених держав, застосуванням технологій Big Data з боку недержавних терористичних формувань, а також цілий ряд проблем, пов'язаних з етичним аспектом забезпечення приватного життя користувачів.

На його думку, існує значний спектр інструментальних можливостей використання технологій Big Data для забезпечення конструктивного та ефективного суспільно-політичного розвитку в сучасних державах.

Багато в чому основні вектори застосування сучасних технологій роботи з цифровими масивами даних будуть визначатися в процесах реальної політичної практики, в рамках якої баланс інтересів різних акторів суспільно-політичних процесів може бути зміщений в ту чи іншу сторону.

Звертає увагу вчених-політологів на можливі складнощі, пов'язані з впровадженням в повсякденне політичну практику «розумних» технологій, здатних сформувати негативний потенціал в аспекті суспільно-політичного розвитку.

Крім того, зважаючи на міждисциплінарний характер дослідження, є велика кількість вчених, що проводять дослідження в частково-суміжних та релевантних тематиках.

Таким чином, ми можемо зробити висновок, що тема, незважаючи на свою відносну новизну досить часто стає об'єктом наукових досліджень, але через надзвичайно-стрімкі темпи розвитку з кожним днем з'являються нові виклики та можливості, технології та методики, що можуть бути досліджені з нуля, або більш детально.



## 1.2. Аналіз основних понять дослідження

Дослідження використання технології Big Data в політичних процесах передбачає визначення базового термінологічного полі дослідження, зважаючи на його комплексний та міждисциплінарний характер.

Весь тезаурус можна поділити на дві умовні групи, враховуючи його специфіку. До першої групи належать терміни, що описують політичні поняття, які можуть бути дотичними до політичних процесів, в яких можна використати Великі дані:

До другої групи віднесемо поняття, які пов'язані суто з технічною стороною роботи і мають бути розкриті більш детально у зв'язку з її захистом на гуманітарній спеціальності для доброго розуміння читачем: Data science, машинне навчання, Data mining, хмарне середовище ті інші.

Але почнемо все таки з найбільш вживаного в роботі поняття, винесеного у назву, але найбільш розвинутого, а саме — Big Data.

Вперше задокументоване використання цього терміну з'явилося в науковій роботі працівника NASA у 1997 році. Там говориться про «цікавий виклик для комп'ютерної системи: масив даних, що, як правило, досить великий [...] коли набори даних не вміщуються в основній пам'яті [...] і навіть на локальному диску та потребує залучення більшої кількості ресурсів» (приблизний переклад з англ.) [71].

Найбільш популярним сьогодні є визначення з дослідження, проведеного МакКінсі у 2011 році, де він описав Big Data як: «[...] набори даних, розмір яких перевищує можливості типових програмних засобів баз даних записувати, зберігати, керувати та аналізувати [...]» [90].

Оксфордський словник англійської мови надає наступне визначення цьому терміну: «набори інформації, занадто великі або занадто складні, щоб обробляти, аналізувати або використовувати стандартними методами» (приблизний переклад з англійської) [66].

Дослівний переклад українською — «Великі дані».

Тобто, фактично під Великими Даними розуміються такі дані, що перевищують стандартні об'єми, але як ми визначили в ході роботи, поняття таких об'ємів наразі досить розмите. Формального розділення між великими і звичайними даними немає.

З самого початку Big Data описували через три V:

- Volume — об'єм, оскільки інформації повинно бути дуже багато;
- Velocity — швидкість, бо їх об'єм постійно збільшується, його потрібно швидко обробляти;
- Variety — багатоманітність та неоднорідність, оскільки в роботу йдуть числа, тексти, графіка ті й інші типи даних, в тому числі неструктуровані.

В подальшому аналітики запропонували додати ще п'ять V до попереднього переліку, а саме:

- Veracity — достовірність;
- Viability — життєздатність;
- Value — цінність для економіки, науки та людства в цілому;
- Variability — мінливість;
- Visualization — можливість візуалізації результатів.

Виходячи з цих пунктів можна помітити, що науковці намагаються підкреслити практичність цих даних в майбутньому. Тобто, річ повинна йти не просто про великі об'єми даних, а такі, з яких потім можна отримати користь [20].

Наука про дані (англ. Data science) — розділ інформатики, що вивчає проблеми аналізу, обробки та публікації даних в цифровій формі. Завдяки популяризації Big Data виокремлена в практичну міжгалузеву сферу діяльності [28], а з початку 2010-х років вважається однією з найпривабливіших, високооплачуваних та перспективних професій у світі [72].

Аналіз сфери політики, політичних процесів, політичних технологій. Інформаційні та Біг Дата.

Машинне навчання (англ. machine learning) — сфера штучного інтелекту, що розуміє під собою розробку програм та комплексів, алгоритмів, здатних навчатись та приймати рішення, виходячи з власного досвіду. Машина запрограмована на обробку та аналіз даних, що надходять, щоб в подальшому використовувати опрацьовану базу для прогнозування трендів, в тому числі поведінки користувачів. В подальшому програма може виконуватись без отримання нових вказівок від людини [91].

Термін Data mining не має поки що сталого перекладу українською мовою, а тому більшість спеціалісти використовують саме англійську версію. Вона була введена Пітецьким-Шапіро Г. у 1989 році та означає сукупність методів виявлення в даних раніше невідомих, нетривіальних, практично корисних і доступних для подальшої інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності [23].

Хмарні обчислення (англ. cloud computing) — модель для забезпечення зручного доступу через Інтернет-мережу до пулу конфігурованих обчислювальних ресурсів, таких як сервери, сховища, служби чи програми, які можуть бути швидко надані з мінімальними зусиллями управління або взаємодії постачальника послуг [99]. Використання хмарних обчислень можуть значно зменшувати витрати на інфраструктуру інформаційних технологій та швидко масштабуватись за потреби користувача [31].

Також, протягом роботи ми будемо часто зустрічати такі поняття, як структуровані та неструктуровані дані. До структурованих відносяться такі, що мають певну жорстку структуру та можуть бути поміщені в базу даних. Неструктуровані не мають цієї самої чіткої структури, а тому для їх збереження використовуються спеціалізовані сховища.

Неструктуровані дані, як правило, представлені в формі тексту, котрий може містити такі дані, як дати, цифри, факти, закодоване зображення та ін. Це призводить до складнощів з аналізом, а тому для зручної обробки вимагають в подальшому переведення у структуровані дані [29].

Перейдемо до гуманітарних термінів, з якими ми стикаємося під час нашої роботи як дослідників. В першу чергу — це ще одне широке поняття

Ващенко К. О. та Корнієнко В. О. дають на характеризують політичний процес як специфічну ланку політології, що має неоднозначне тлумачення: Доуз Р. ототожнює його з поняттям політики в загалом; Парсонс Т. вбачає специфічність політичних процесів в результатах функціонування політичної системи, а Дарендорф Р. — в динаміці боротьби різних суб'єктів політики за статуси і владу. Мерріам Ч. додає сюди ще й політичну поведінку суб'єктів політики в ході досягнення власної мети. Виходячи з цього, робиться висновок, що в змісті будь-якого політичного процесу присутні зміни, діяльність людей та їх конфлікти, а також функціонують певні інститути і досягається результат [5].

Отже, політичний процес — це форма функціонування політичної системи суспільства, що змінюється у просторі і часі; сукупна діяльність суб'єктів політики, завдяки якій забезпечується функціонування та розвиток політичної системи [38].

Таким чином, ми виклали основні поняття та їх пояснення для подальшого маніпулювання ними в рамках нашого дослідження. Це дало нам змогу структурувати теорію, узагальнити та систематизувати певні системоутворюючі поняття, що могли в подальшому викликати непорозуміння через міждисциплінарний характер роботи.

При цьому, оскільки в дослідженні зустрічаються поняття, які мають досить розмиті, або не до кінця сформульовані тлумачення, ми визначили для себе те, як будемо сприймати його в рамках цього дослідження.

### **1.3. Методологія дослідження використання технології Big Data в політичних процесах**

Будь яка наука володіє своєю системою дослідницьких принципів, що виникають в процесі творчого освоєння людиною об'єктивної реальності.

Саме аналітична діяльність має в своїй основі низку різноманітних методів, кожен з котрих несе в собі сукупність певних правил, алгоритмів, принципів та прийомів аналітичної діяльності, яка складається в певну систему.

Наразі наука має декілька визначень поняттю «метод дослідження». Вони не суперечать один одному, а лише доповнюють чи уточнюють інші, хоча частіше за все просто несуть ту саму думку, але сформульовану іншими словами.

Ми візьмемо за основу поняття, викладене Рузавіним Г. І. про те, що метод пізнання, він же — метод дослідження — це певна специфічна процедура, яка складається з низки дій, за допомогою котрих здобувається та обґрунтовується нове знання в науці [53].

Оскільки наше дослідження є міждисциплінарним та охоплює одночасно як гуманітарну політологічну науку, так і технічну з статистичного аналізу і машинного навчання, то коло методів досить широке.

В контексті даної роботи нами були використана низка методів, основними з яких виступили наступні: порівняльний, історичний, системний та структурно-функціональний.

Метод порівняння. Відноситься до найпоширеніших методів наукового пізнання, оскільки дає можливість, порівнюючи певні ознаки та сутності, знаходити подібності та відмінностей, що властиві двом або більшій кількості об'єктів [10]. Є методом емпіричним.

В контексті нашого дослідження ми мали можливість провести низку порівнянь. Наприклад, порівнюючи кейси використання технології Big Data в Україні та за кордоном, ми маємо змогу знайти слабкі сторони та віднайти

шляхи їх ліквідації. В той самий час сильні сторони показують можливість взаємовигідного обміну досвідом із закордонними партнерами в цій галузі.

Порівнюючи досвід роботи с Великими Даними між політичними, некомерційними та урядовими структурами, ми можемо віднайти сильні та слабкі сторони кожного напрямку, усвідомивши які та поєднавши їх для ліквідації недоліків можна осмислено вивести більш сильну формулу роботи з Big Data.

Перейдемо до методів загальнонаукових, тобто таких, що використовуються як для емпіричного, так і теоретичного дослідження [10, 23].

Перш ніж вивчати сучасний стан технології, потрібно дослідити її розвиток та зрозуміти, від чого та яким шляхом наука дійшла до неї. Тут нам на допомогу приходить історичний метод. Саме він дозволить нам дослідити виникнення, становлення та розвиток технології, появи її понять і термінів у хронологічній послідовності [21, 19].

І вже розуміючи історію розвитку технології, ми зможемо не тільки аналізувати її стан на сьогодні, але й спрогнозувати подальший розвиток як в Україні, так і по всьому світу.

Також не обійшлося і без структурно-функціонального метода. Його сутність полягає в розбитті системних об'єктів на структурні елементи і визначення їх функцій в системі [21, 23].

Розібравшись в структурі технології та виділивши основні компоненти, ми можемо знайти сильні та слабкі сторони, що наразі наявні в цій технології; виокремити проблеми, що заважають масштабувати використання Big Data в політичних процесах, а що навпаки загрожує окремим громадянам чи цілим державам під час її використання.

Системний метод використано в роботі для розгляду процесу використання Big Data в якості системи, сукупності певних етапів, стадій, які досліджуються автономно з урахуванням подальшого узагальнення та узгодження особливостей кожного етапу із загальною динамікою політичного

процесу загалом. Завдяки цьому методи ми матимемо можливість прослідкувати всю складність використання Big Data в політичних процесах як державою, так і комерційними чи неурядовими організаціями, визначити роль та місце Big Data в сучасних політичних процесах по всьому світу.

Отже, ми бачимо, що дослідження використання технології Big Data в політичних процесах вимагає від нас використання цілої низки різноманітних науково-дослідницьких принципів та методів. Саме вищевикладені методи в основі нашого дослідження є найбільш актуальними для висвітлення даного питання та його глибокого, всеосяжного дослідження. Вони є необхідні, та все одно ми не можемо обмежуватись лише ними та по ходу роботи можемо тим чи іншим чином брати на озброєння інші методи. Тільки в такому випадку результати дослідження будуть максимально повними в рамках нашої мети.

Виходячи з викладеного нами в першому розділі нашої роботи, ми можемо зробити висновки, що тематика Big Data в сучасному світі є вельми актуальною та попит на неї зростає з кожним днем. Це призводить і до великого інтересу з боку наукової спільноти.

Ціла низка вітчизняних та закордонних науковців піднімають актуальні теми, що стосуються як технічних, юридичних чи етичних аспектів збирання і використання Big Data, так і економічні, політичні та інші наслідки їх аналізу; можливості використання в найрізноманітніших сферах.

При цьому, через швидкість, з якою ці технології трансформуються на розвиваються, питань не вивчених в цій сфері все ще значно більше, ніж досліджених, а їх кількість постійно зростає.

Ми визначили основні поняття, що будуть використовуватись в роботі і можуть викликати певні питання, розібравши їх з різних поглядів та обравши пояснення, на які будемо спиратись під час написання нашого диплому.

Наприкінці були окреслені базові методики наукового дослідження, які ми будемо активно застосовувати в наступних двох розділах, до яких віднесли: порівняння, історичний, системний та структурно-функціональний.

Маючи подібну теоретичну базу, дослідження може переходити в наступну фазу, оскільки ми, як науковці, вже розуміємо тему, з якою нам доведеться працювати та основні принципи і поняття, що будуть використані.



## РОЗДІЛ 2

### ОСОБЛИВОСТІ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ BIG DATA В ПОЛІТИЧНИХ ПРОЦЕСАХ

#### 2.1. Технологія Big Data: історія розвитку

Для того, щоб детально розібрати питання використання такої технології, як Big Data в сучасних дослідженнях і політичних науках доцільно буде провести деякий поверхневий огляд самої технології, історії її виникнення та розвитку на сучасному етапі.

Саме розуміння цих основ допоможе нам в подальшому працювати з темою таким чином, щоб охопити максимум деталей, які потрібні для вивчення тематики.

Власне кажучи, деякі найдавніші записи про застосування даних для аналізу та контролю ділової діяльності були зроблені близько 7000 років назад.

Це відбулось із запровадженням в Месопотамії обліку приросту врожаю та поголів'я.

Принципи статистичного аналізу продовжували зростати і вдосконалюватися. У 1663 році Грант Дж. записав та проаналізував інформацію про рівень смертності в Лондоні. Грант зробив це, намагаючись підвищити обізнаність про наслідки бубонної чуми, яка тривала в той.



Винахідник мав рацію: сучасні технології, що стали доступні кожному, спровокували інформаційний вибух. Так, у своїй книзі «Критичний шлях» футурист і винахідник Річард Бакмінстер Фуллер підрахував, що до 1900 року людські знання подвоювались приблизно кожне століття. У період з початку ХХ ст. до кінця Другої Світової війни інформація стала подвоюватись раз на 25 років, а у 1982 році — кожен рік.

Згідно з оцінками корпорації ІВМ, наразі знання людства та наявна інформація подвоюється кожні 12 годин.

Схематично це інформаційне цунамі відображене в роботах Марка Розенберга [103]. Побачити схему можна нижче.

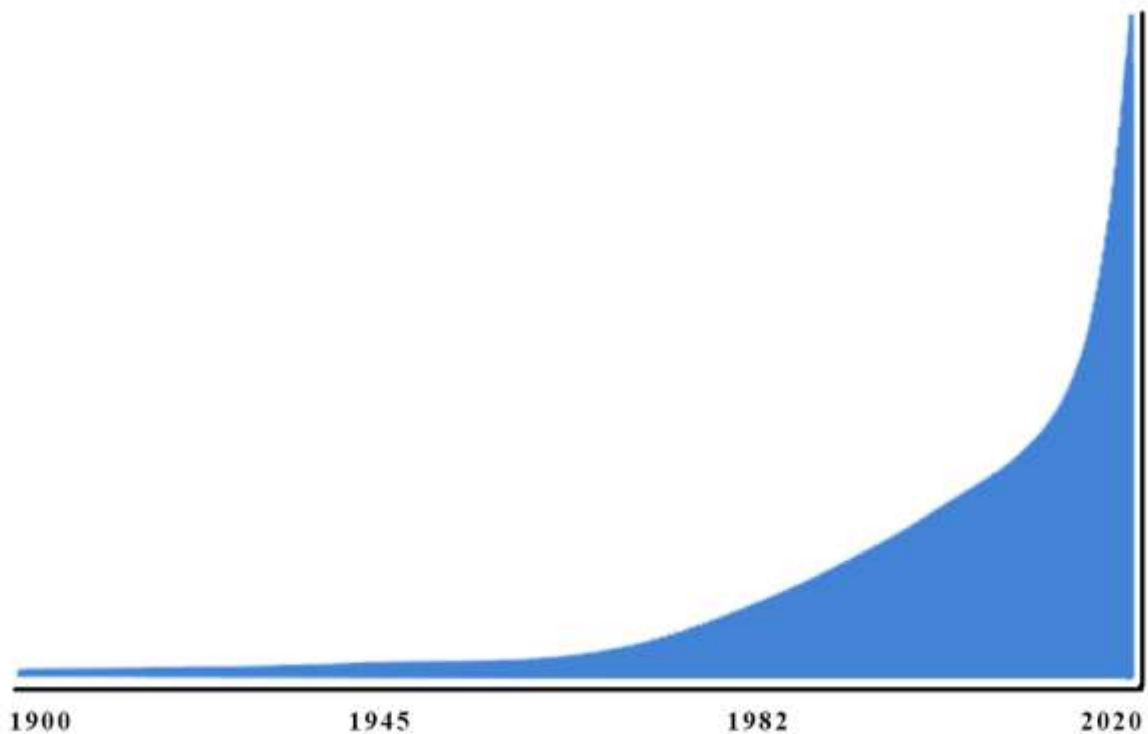


Рис. 2.2 Схематичне відображення інформаційного цунамі

В той самий час, хоча сама інформація експоненціально зростає, період напіввиведення знань зменшується. Тобто, період часу, який повинен пройти до того, як половина знань або фактів буде замінена або виявиться неправдивою [83].

У 1928 році Пфляймер Ф., німецько-австрійський інженер, винаходить метод магнітного зберігання інформації на стрічці [78]. Принципи, які він розробляє, використовуються і сьогодні в більшості цифрових систем, оскільки дані зберігаються магнітно на жорстких дисках.

Під час Холодної Війни перед США та СРСР постає питання шифрування та дешифрування повідомлень. Це стає поштовхом для створення перших систем, що можуть самостійно, в автоматичному режимі збирати та обробляти інформацію.

В 1965 році США планують побудувати [63] перший у світі центр обробки даних, який може зберігати одночасно 742 мільйони податкових декларацій та 175 мільйонів відбитків пальців на магнітних носіях.

А в 1991 році інформатик Бернерс-Лі Т. оголосив про народження того, що стане Інтернетом в тому сенсі, в якому ми знаємо його сьогодні. В групі комп'ютерної мережі Usenet він викладає технічну специфікацію взаємопов'язаної глобальної мережі даних, що буде доступна з будь-якого місця та для будь-кого.

У 1996 році збереження даних на цифрових носіях вперше стає економічно більш доцільним, ніж на паперових [94]. Цей момент можна умовно назвати одним з вирішальних, оскільки починаючи з нього все більше інформації починає переводитись та зберігатись в цифровому вигляді.

В жовтні 1997 року Майкл Кокс та Девід Еллспорт публікують доповідь, де аналізують роботу з пейджинговими даними таких об'ємів, що це може скласти проблему для тодішніх комп'ютерів [71]. Саме в ньому вони вперше використали термін Big Data в тому розумінні, про яке говоримо ми сьогодні, хоча і в значно менших масштабах.

У 2005 році вчені повідомляють, що ми спостерігаємо народження "Web 2.0" [96] — створеної користувачами мережі, де більшість контенту генеруватимуть вже користувачі послуг, а не самі постачальники. Це досягається інтеграцією традиційних вебсторінок у стилі HTML з величезними внутрішніми базами даних, побудованими на SQL. 5,5 мільйона

людей вже використовують Facebook, запущений роком раніше, для завантаження та обміну власними даними з друзями.

Того року також було створено Hadoop — фреймворк з відкритим кодом, створений спеціально для зберігання та аналізу наборів великих даних. Його гнучкість робить його особливо корисним для управління неструктурованими даними (голосом, відео, необробленим текстом тощо), які користувачі все частіше створюють та збирають.

Паралельно зі зростанням об'єму інформації відбувається і зниження вартості її зберігання та обробки, що також впливає на темпи росту інтересу до технології Big Data.

Середня вартість жорстких дисків за останні роки зазнали суттєвого зменшення. Так, у 2000 році середня вартість 1 гігабайту складала 7 доларів, а в 2020 — 0,018 долара. Динаміку можна чудово побачити на діаграмі, побудованій нижче, спираючись на дані з мережі Інтернет [76].



Рис. 2.3 Діаграма динаміки середньої ціни на жорсткі диски (за 1 гігабайт)

Крім того, продовжує розвиватись технологія зберігання інформації на стрічці [88]. Наразі її нерентабельність використання в невеликих компаніях та на ПК пов'язана з ціною на пристрої зчитування та запису з таких носіїв, але самі носії значно дешевші та мають більший об'єм. Великі компанії, такі як Google [80], Microsoft [93] використовують саме цю технологію.

Так, через досягнення найменшого розміру гранули, яка придатна для магнітного запису на HDD (супермагнетична межа [106]), виробники жорстких дисків не можуть вмістити більше даних на тій самій площині.

В той самий час запис на плівці це далеко від супермагнетичної межі. У 2017 році компанія IBM, разом з Sony досягти щільність запису в 20 разів більшу, ніж будь які інші сучасні накопичувачі.

Наприклад, плівковий накопичувач Lto Ultrium 8 Data Cartridge одного з останніх поколінь, що дозволяє зберігати до 30 терабайт інформації, можна придбати від 200 доларів [64], тобто ціна одного гігабайту складає менше 0,003 долара. Накопичувачі більш старого покоління можна придбати за ціною від 0,001 долара за гігабайт [77]. Середня ціна за гігабайт на традиційному магнітному накопичувачі станом на 2020 рік становить близько 0,032 долара, а диск об'ємом у 16 терабайт обійдеться від 500 доларів [104].

На противагу їм розвивається ще один важливий напрямок — твердотільні накопичувачі, які хоч і мають наразі менший об'єм, але дозволяють проводити зчитування та запис інформації з більшою швидкістю.

Наразі плівковий носій максимального об'єму можна придбати від 114 доларів США, що може у зв'язці з накопичувачем IBM TS1155 зберігати до 20 терабайт інформації.

Завдяки динамічному розвитку технології, вона все частіше почала використовуватись в найрізноманітніших сферах життя.

Так, для підвищення прибутку в торгівлі компанія Walmart щогодини обробляє понад мільйон транзакцій клієнтів, які імпортуються до баз даних (за оцінками, вони містять понад 2,5 петабайта даних). Завдяки цій системі, що має внутрішню назву Retail Link постачальники бачать точну кількість своїх

продуктів на кожній полиці кожного окремого магазину в теперішню хвилину. Також система демонструє темпи продажу по годинах, днях, за останній рік.

Започаткована в 1990-х роках Retail Link дає постачальникам повне розуміння того, коли і як продається їхня продукція, а також з якими іншими товарами потрапляє до кошика. Це дозволяє постачальникам покращувати об'єми продажу завдяки коригуванню маркетингових стратегій.

Ця технологія дозволяє WalMart змінити бізнес-модель роздрібною торгівлі. У деяких випадках компанія навіть залишає управління запасами продукції на складах в руках своїх постачальників і не приймає на себе права власності на продукти до моменту їх продажу. Це дозволяє знизити інвентарний ризик і витрати.

Windermere Real Estate, приватна компанія с нерухомості в США, яка є однією з найбільших в своєму роді на Заході США [114] стала однією з перших на ринку надавати таку ціну для потенційних покупців інформацію, як приблизний час їзди автівкою на роботу в залежності від часу доби, дня тижня та т. і. Ці дані надаються компанією Inrix.

Підрахунки поїздок компанії базуються на аналізі з навігаційних систем понад 100 мільйонів водіїв, які анонімно повідомляють систему про своє місцезнаходження та швидкість.

Він оновлює свої оцінки часу їзди кожні 90 днів, щоб відобразити мінливі дорожні умови — наприклад, великий проєкт шосе, який створює об'їзд, що заїжджає на щомісячні поїзди на місяці. У компанії навіть є алгоритми, які викорінюють аномалії, такі як люди, які зупиняються, щоб замінити запасну шину, сказав Кевін Форман, генеральний менеджер з геоаналітики Inrix [115].

Науковці також намагаються використати цю технологію на повну. Так, експерименти Великого адронного колайдера фіксуються завдяки близько 150 мільйонів датчиків, що передають дані 40 мільйонів разів на секунду. За секунду відбувається майже 600 мільйонів зіткнень всередині, що продукує колосальні об'єми даних. Навіть зберігаючи менше ніж 0,001% даних потік

даних станом на 2012 рік складав 25 петабайт на рік. Якщо б фіксували 100% даних, то потік даних становив би майже 150 млн петабайт на рік.

Раніше, коли тільки розпочався проєкт нобелівського лауреата Джейсона Ватсона що до Проєкту геному людини, розробка його кодування займала 10 років. З використанням технологій Big Data цей результат можна досягти менш ніж за добу [75].

Центр кліматичної симуляції НАСА зберігає більше 32 петабайт кліматичних досліджень та завдяки суперкомп'ютеру моделює важливі кліматичні ситуації, як то: прогноз зміни клімату з 1880 по 2100 роки; створення інтерактивних тривимірних візуалізацій циклонів; ретроспективний аналіз погоди та клімату, тощо [73].

Пандемія COVID-19, що охопила весь світ, також зробила великий поштовх для використання Big Data в медицині. Технологія допомагає провести обчислювальний аналіз, щоб виявити закономірності, тенденції, асоціації та відмінності. Це також може допомогти у розкритті інформації про поширення та контроль над цим вірусом. Завдяки можливості детального збору даних, великі дані можна використовувати з вигодою, щоб мінімізувати ризик розповсюдження цього вірусу [110].

Технологія великих даних може зберігати величезну кількість інформації про цих людей, інфікованих вірусом COVID-19. Це допомагає детально зрозуміти його природу. Отримані дані можна додатково повторно навчити для розробки майбутніх профілактичних методів. Кілька методів цифрових даних, включаючи місцезнаходження пацієнта, близькість, поїздки, про які повідомляють пацієнти, супутню захворюваність, фізіологію пацієнта та поточні симптоми, можна оцифрувати та використовувати для отримання ділової інформації на рівні громади та демографії [82].

Також Big Data часто використовуються спортсменами для поліпшення тренувань та розуміння суперників. Збираючи статистичні дані, в тому числі зі спортивних датчиків, можна детально коригувати тренування власних команд, а також передбачити слабкі сторони конкурентів. Також треба



вказати, що зарплата гравців команд часто визначаються даними, які були зібрані протягом усього сезону та демонструють їх ефективність.

Аналізуючи дані гравців, рекрутери або спортивні менеджери та тренери використовують прогнозу аналітику, щоб визначити майбутню ефективність гравця та визначити його вартість, а отже, ціну, яку вони готові заплатити за контракт чи оренду.

Використовується прогностична аналітика, щоб визначити, яких гравців наймати чи замінювати — можливо, вони досягли піку своєї кар'єри.

Аналітика також використовується з боку бізнесу для прогнозування продажів квитків, визначення ціни на матчі, управління стосунками з фанатами та покращення рентабельності інвестицій [98].

В цій главі ми детально розібрали історію виникнення поняття Big Data, зрозуміли основні тенденції розвитку. Окремо зосередились на ефективних прикладах використання в різних сферах, крім політичних та державних, яким буде приділено більше уваги в окремому розділі дослідження.

Виходячи з вищевикладеного можна зробити висновки, що технологія Big Data — це досить широке поняття, яке й досі не має точного визначення, але використовується майже в усіх сферах сучасного життя, поліпшуючи ефективність роботи аналітиків та обробки даних в десятки та сотні разів.

З кожним днем технології обробки, збору та зберігання великих даних стають все доступнішими, а тому вони будуть тільки зростати та еволюціонувати а подальшому.

## 2.2. Зміст та структурні складові технології Big Data

Процес обробки Big Data зазвичай складається з шести кроків. Розглянемо кожен з них для того, щоб в подальшому мати розуміння етапів роботи аналітиків в цій сфері.

- Визначення мети дослідженн
- Збір даних чи отримання від третіх осіб
- Підготовка; в цій фазі дослідники підвищують кількість, якість даних для подальшої обробки. Складається з трьох підфаз: очистка даних для видалення некоректних значень ті усунення розбіжностей між джерелами; інтеграція для охоплення більшої інформації з декількох джерел та перетворення, що гарантує в подальшому збереження даних в зручному форматі [107, 27].

- Дослідження даних для досягнення більш глибокого розуміння
- Моделювання даних чи побудова моделі для отримання відповіді на запити дослідження
- Відображення та автоматизація результатів, що може в подальшому існувати самостійно.

Компанія McKinsey, яка спеціалізується на вирішенні завдань стратегічного управління з використанням технологій Великих Даних виділяє 11 основних технік та методів аналізу, які можна застосовувати по відношенню до Big Data [90].

*Data Mining*. В розділі що до понятійних засад нашого дослідження ми дійшли висновку, що це сукупність методів виявлення в даних раніше невідомих, нетривіальних, практично корисних і доступних для подальшої інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності.

В основі цього, фактично, глибинного, інтелектуального аналізу даних лежать, в тому числі: навчання певним асоціативним правилам, класифікація шляхом поділення на категорії та підкатегорії, кластеризація та подальший кластерний аналіз, пошук різних відхилень, тощо.

Прикладом може служити розповсюджена ситуація, коли у нас є досить велика база даних та ми розуміємо, що в базі існує певна «прихована інформація». Ці дані повинні бути раніше невідомими, нетривіальними, практично корисними та доступними до інтерпретації.

*Краудсорсінг* — залучення до вирішення тих чи інших проблем виробничої діяльності широкого кола осіб для використання їх знань, досвіду, творчих здібностей як субпідрядників на добровільних засадах з використанням інформаційних технологій. Таким чином ми маємо безліч переваг, до яких можна віднести: доступ до талановитих кадрів по всьому світу без бар'єрів, зменшення часу на вирішення певної проблеми та інші.

Мабуть, найбільш відомим прикладом, котрим хоч раз в житті користувалась будь-яка людина, що має доступ до Інтернету — це Вікіпедія, найбільша в світі електронна енциклопедія, що повністю наповнюється силами волонтерів [33]. Так само створюються картографічні бази OpenStreetMaps [35] та навіть проводяться супутникові археологічні експедиції [11].

*Змішування та інтеграція даних* — це процес інтеграції кількох джерел даних для отримання більш послідовної, корисної і точної інформації, ніж та, що базується тільки на одному з джерел.

Прикладом такого синтезу можна вважати аналіз тональності тексту, що часто використовується під час політичної аналітики для оцінки ставлення користувачів Мережі до події чи особи. Дані як для навчання штучного інтелекту, так і для аналізу видобуваються з багатьох джерел та інтегруються для підвищення якості результату [50].

*Машинне навчання* як основа побудови штучного інтелекту є характерною ознакою роботи з Big Data. На підставі даних, що є результатом рішення великої кількості типових задач формується ядро, яке в подальшому повинно продовжити самостійне вдосконалення моделі під час самостійного аналізу і роботи.

Таким чином, ми отримуємо систему, що постійно самовдосконалюється та покращує власні результати з обробки Big Data та їх використання. Прикладом може слугувати система прогнозування погоди на основі статистичних даних. Ядром стає база спостереження за метеорологічними умовами на протязі певного проміжку часу, наприклад — 100 останніх років. Система працює, виробляючи прогноз погоди, та паралельно збираючи та зберігаючи нові дані. Ці дані збільшують дані для самоаналізу, і вже на базі успішних чи помилкових прогнозів система коригується та самоудосконалює власний алгоритм.

*Штучні нейронні мережі*, мережевий аналіз, оптимізація, в тому числі генетичні алгоритми (genetic algorithm — евристичні алгоритми пошуку, які використовуються для вирішення завдань оптимізації та моделювання шляхом випадкового підбору, комбінування і варіації шуканих параметрів з використанням механізмів, аналогічних природному відбору в природі) [47].

*Розпізнавання образів* — система класифікації та ідентифікації предметів, явищ, процесів, сигналів, ситуацій тощо, котрі характеризуються кінцевим набором певних ознак та властивостей [54]. Прикладами можуть бути системи розпізнавання автомобільних номерів, облич, документів, штрихкодів та інших образів.

Наразі вони все глибше входять в наше життя, оскільки дозволяють пришвидшити обробку величезних масивів даних та позбавитись людського фактору.

*Прогнозна аналітика* — клас методів аналізу даних, що бере за основу прогнозування майбутньої поведінки об'єктів і суб'єктів з метою прийняття оптимальних рішень.

Використовуючи методи інтелектуального аналізу даних, теорії ігор та аналізуючи поточні і історичні факти система може з великою долею вірогідності передбачати майбутні події. Прикладом такого використання можна назвати систему соціально-кредитного рейтингу, впроваджувального в

Китаї та розглянутої нами в роботі в розділі закордонного досвіду використання технології.

*Імітаційне моделювання або симуляція* — метод, в якому досліджуваний об'єкт замінюється моделлю, яка з великою точністю описує реальну та імітує її. Іже з нею проводяться експерименти для одержання інформації. Є різновидом експериментальних випробувань. Використовуються у випадках, коли розроблення аналітичної моделі є неможливим чи недоцільним.

Прикладом може стати симуляція дослідниками компанії Microsoft навантаження на державні системи охорони здоров'я у випадку пандемії [79]. Достатньої кількості даних до спалаху коронавірусної інфекції в тоді ще не було, а тому вірогідно, що саме імітація використовувалась під час аналітичної роботи. Вона показала неготовність найрозвинутіших країн світу до пандемії, але всі попередження були проігноровані [8].

*Просторовий аналіз* або просторова статистика включає будь-який із формальних методів, що вивчають сутності з використанням їх топологічних, геометричних або географічних властивостей.

Просторовий аналіз включає різноманітні методики, багато з яких все ще перебувають у своєму ранньому розвитку, використовуючи різні аналітичні підходи і застосовуються в таких різноманітних галузях, як астрономія з її дослідженнями розміщення галактик у космосі.

У більш простому розумінні просторовий аналіз — це техніка, що застосовується до структур у людському масштабі, особливо при аналізі географічних даних.

Першим відомим застосувань просторового аналізу була карта доктора Джона Сноу з Лондона, який на звичайній карті виявив скупчення випадків холери у 1854 році (див. додаток А).

З появою комп'ютеризованих геоінформаційних систем та технологій Big Data просторовий аналіз став масовим явищем. Сучасним прикладом можна назвати побудову за аналогією карти захворюваності COVID19 [70].

*Статистичний аналіз.* До нього можна віднести аналіз часових рядів та спліт-тестування (А/В-тестування).

Спліт-тестування добре відоме в маркетингу. Під час нього створюється контрольна група та кілька тестових, в кожній з яких змінено один чи декілька показників. Таке тестування дозволяє з'ясувати, як зміна тих чи інших показників впливає на кінцевий показник.

Часто використовується під час підготовки політичної реклами та брендингу, коли тестовим групам демонструються, наприклад, брендбук з різними логотипами чи кольоровими гамами. В масштабах Big Data ті самі методи можуть використовуватись за допомогою систем статистики на офіційному сайті, чи в рекламній компанії через соціальні мережі.

*Візуалізація аналітичних даних* — переведення інформації в малюнки, діаграми, карти, інфографіку з використанням інтерактивних можливостей та анімації як для отримання результатів, так і для використання в якості вихідних даних для подальшого аналізу. Дуже важливий етап аналізу великих даних, що дозволяє представити найважливіші результати аналізу в найбільш зручному для сприйняття вигляді [67].

Технологій обробки великих даних наразі також досить багато, а тому ми звернемо увагу на основні у 2020 році, що були виділені компанією Analytics Steps Infomedia [109].

*Штучний інтелект* — про нього ми вже казали в нашому дослідженні і не раз. Це технологія, що дозволяє алгоритму чи машині самовдосконалювати свій алгоритм. Технологія постійно розвивається та використовується вже майже в усіх галузях, починаючи від гри в шахи [74], закінчуючи судочинством [2].

*База даних NoSQL* — дозволяє зберігати неструктуровані дані та забезпечує швидшу продуктивність, ніж традиційні SQL, а також забезпечує гнучкість під час роботи з різноманітними типами даних у величезному масштабі [111]. Прикладами можуть бути бази даних MongoDB, Redis та Cassandra [24].

*Мова програмування R* — безкоштовне програмне забезпечення, що широко використовується для статистичних обчислень, візуалізації, уніфікованих середовищ для розробки, таких як Eclipse та Visual Studio. Однією з сильних сторін R є простота, з якою можна створювати добре розроблені графіки якості публікації, включаючи математичні символи та формули, де це необхідно [100].

*Озеро даних (DataLake)* — метод зберігання даних системою або репозиторієм в натуральному форматі, який передбачає одночасне зберігання даних в різних схемах і форматах, тобто без обов'язкового перетворення неструктурованої інформації на структуровану. Ви можете зберігати дані як є, без необхідності попередньо структурувати дані та запускати різні типи аналітики — від інформаційних панелей та візуалізацій до обробки великих даних, аналітики в реальному часі [112].

*Блокчейн* — надзвичайно захищена екосистема та неймовірний вибір для різних застосувань великих даних у галузях банківської справи, фінансів, страхування, охорони здоров'я, роздрібної торгівлі тощо. Ця база даних зберігає впорядкований ланцюжок записів (так званих блоків), що постійно [3; 68]. Дані захищено від підробки та спотворення.

Технологія блокчейн все ще перебуває в процесі розробки, однак багато продавців різних організацій, таких як AWS, IBM, Microsoft, включаючи стартапи, пробували численні експерименти, щоб представити можливі рішення у побудові технології блокчейн.

*База даних в пам'яті (IMDB)* зберігається в оперативній пам'яті комп'ютера (RAM) і управляється системою управління базами даних у пам'яті. Раніше звичайні бази даних зберігалися на жорстких дисках чи SSD-накопичувачах, що мають в рази меншу швидкість доступу до інформації.

Бази даних в пам'яті швидші, ніж бази даних, оптимізовані для використання дискових накопичувачів, оскільки доступ до диска повільніше, ніж доступ до пам'яті, а внутрішні алгоритми оптимізації простіші і виконують менше інструкцій центрального процесора. Доступ до даних з пам'яті

виключає час на пошук при їх запиті, що забезпечує більш швидку і передбачувану роботу, ніж при використанні дискового накопичувача [84]).

Але оскільки всі дані збираються та контролюються в основній пам'яті повністю, існує велика ймовірність втратити дані внаслідок відмови процесу або сервера.

*Apache Spark* — це цілісна обчислювальна система з набором бібліотек для паралельної обробки даних на кластерах комп'ютерів. На даний момент Spark вважається самим активно розробляються засобом з відкритим кодом для вирішення подібних завдань, що дозволяє йому бути корисним інструментом для будь-якого розробника або дослідника-фахівця, зацікавленого у великих даних. Spark підтримує безліч широко використовуваних мов програмування (Python, Java, Scala і R), а також бібліотеки для різних завдань, починаючи від SQL і закінчуючи трансляціями і машинним навчанням, а запустити його можна як з ноутбука, так і з кластера, що складається з тисячі серверів [55; 87].

В цій главі ми дослідили основні методи та техніки, а також інструменти для обробки та збереження Big Data на сучасному етапі. Нам довелось охопити лише найбільш популярні технології, оскільки насправді їх набагато більше і дослідити всі в рамках однієї, навіть більш вузько тематичної роботи проблематично.

В той самий час це демонструє, що екосистема Великих даних постійно зростає, нові технології з'являються, а старі покращуються з кожним днем, що розширює можливості для досліджень, зменшуючи фінансовий та освітній поріг входження аналітиків в цю сферу.

В той самий час, динамічний розвиток технологій демонструє попит, який тільки зростає.

Виходячи з цього ми можемо припустити, що тенденція є сталою і Big Data в усіх сферах життя, в тому числі політичних процесах будуть лише набирати обертів.



### 2.3. Юридичні та етичні аспекти використання технології Big Data в політичних процесах

Один з важливих моментів, які потрібно розглядати в рамках цієї роботи — це етичність та юридичні аспекти збору, зберігання, обробки тієї інформації, яка потрібна для роботи з Big Data. В цій главі ми повинні розібрати: які дані відносяться до персональних, безособових та таких, що мають обмежений доступ. Повинні розібрати аспекти використання кожного з цих типів даних та розглянути типові законодавчі аспекти, які наразі набирають чинності у світі з цих питань.

Працюючи з Big Data гостро постають проблеми юридичних та етичних аспектів, оскільки більша частина з них можуть належати окремим фізичним чи юридичним особам, зокрема відноситись до персональних даних (тобто такими, за якими можна ідентифікувати особу-власника). А тому одночасно з питаннями обробки таких даних набуває актуальності так звана етика даних, оскільки збільшується і масштаб впливу.

Особисті дані стосуються наборів даних, що описують людину, починаючи від фізичних атрибутів і закінчуючи її уподобаннями та поведінкою. Приклади персональних даних включають: дані про геном, розташування GPS, письмове спілкування, розмовне спілкування, списки контактів, звички перегляду Інтернету, фінансові операції, витрати в супермаркетах, податкові платежі, судимість, запис об'єктивів камери ноутбука та мобільного телефону, записи мікрофона пристрою, звички водіння через автомобільні трекери, мобільні та медичні записи, заняття фітнесом, харчування, вживання речовин, серцебиття, режим сну та інші життєво важливі показники. Колектив персональних даних однієї особи формує цифрову ідентичність (або, можливо, цифрове Alter ego є більш доречним). Цифрова ідентичність охоплює всі наші персоналізовані дані, що представляють і пов'язані з нашим фізичним та ідеологічним «я». Різниця між категоріями даних не завжди є чіткою. Наприклад, Дані про здоров'я та банківські дані переплітаються, оскільки про поведінку та спосіб життя можна

зробити висновок за допомогою банківських даних, і це надзвичайно важливо для прогнозування ризику хронічних захворювань.

Отже, банківські дані — це також дані про здоров'я. Дані про стан здоров'я можуть вказувати на те, скільки людина витрачає на охорону здоров'я, отже, дані про здоров'я — це також банківські дані. Ці перекриття існують і між іншими категоріями даних, наприклад, дані про місцезнаходження, дані вебперегляду, дані про податки — це, по суті, все про фізичних осіб.

Право власності передбачає визначення прав та обов'язків щодо власності. Поняття власності на дані пов'язане зі здатністю людини контролювати та обмежувати обмін власними даними. Якщо одна людина записує свої спостереження за іншою особою, якій належать ці спостереження? Спостерігач чи спостерігається? Які обов'язки спостерігач і спостерігається покладають один на одного? Оскільки масові масштаби та систематизація спостережень за людьми та їх думками в результаті Інтернету, ці питання стають все більш важливими для вирішення. Рабство, право власності на людину, заборонено у всіх визнаних країнах. Питання власності персональних даних потрапляє на невідому територію між корпоративною власністю, інтелектуальною власністю та рабством.

Якщо фізична або юридична особа бажає використовувати персональні дані, потрібна поінформована та чітко виражена згода на те, які персональні дані кому, коли і з якою метою передаються від суб'єкта даних. Суб'єкт інформації має право знати, як використовуються їх дані.

Транзакція даних не може використовуватися як переговорна монета для непов'язаного або зайвого питання згоди, наприклад, покращення маркетингових рекомендацій під час спроби зателефонувати до родича. Хоча існують послуги, в яких необхідний спільний доступ до даних, ці транзакції не слід перебільшувати і повинні міститися в контексті. Наприклад, особі потрібно обмінюватися даними, щоб отримати адекватні медичні рекомендації, однак, такі медичні дані не повинні автоматично надходити до

медичного страхування . Врешті-решт людина повинна прийняти рішення щодо своїх даних. Це окремі транзакції даних, які слід розглядати як такі. Пряма згода на прийняття передачі права власності на дані, оскільки використовується програма чату, не вважається дійсною.

Повний обсяг та обсяг операції повинні бути чітко деталізовані для особи, якій має бути надана розумна можливість брати участь у процесі оцінки того, чи бажають вони брати участь. Час є критичним. тобто ці питання слід вирішувати в спокійну хвилину з часом, щоб задуматися, а не в той момент, коли робиться термінова закупівля або виникає надзвичайна медична допомога.

Хоча особа могла б дати згоду на безперервне зберігання, завжди повинна бути можливість відкликати цей дозвіл для майбутніх операцій.

В той же час, наразі суспільство лише приходить до формування єдиних принципів використання Big Data. Найбільш всеохоплюючим документом, який наразі використовується можна вважати Загальний регламент захисту персональних даних (далі — GDPR), що було прийнято у 2016 році та набрав чинності у 2018 на території Європейського союзу.

Перш за все GDPR спрямовано на те, щоб надати громадянам контроль над особистими персональними даними та уніфікувати регулятивні рамки для міжнародних економічних відносин в рамках ЄС та ЄЕС.

Основними принципами GDPR є [32]:

- законність, справедливість та прозорість — всі дані можуть збиратись тільки на законних підставах;
- обмеження метою — обробка повинна виконуватись лише для тих цілей, Про які було заявлено суб'єкту даних. Всі ці завдання повинні бути чітко прописані в політиці конфіденційності.
- мінімізація даних — збиратись та використовуватись мають лише той мінімум даних, який потрібен для виконання поставлених цілей;
- точність — персональні дані повинні бути точними ті не можуть вводити в оману;

- обмеження зберігання даних — їх не можна зберігати довше, ніж потрібно. Повинен проводитись аудит для видалення невикористовуваних даних;

- цілісність і конфіденційність та безпека — всі дані повинні зберігатись в безпечному місці та мати обмеження в доступі для сторонніх осіб;

- підзвітність — відповідальність за обробку персональних даних та виконання всіх інших принципів покладається; призначення посадових осіб щодо захисту даних.

Цей закон стосується не тільки обробників інформації з країн — учасниць ЄС, але і будь-якої юридичної особи, що обробляє персональні дані осіб з ЄС. За невиконання закону накладається штраф до 20 млн. євро або до 4% від річного фінансового обороту компанії за попередній фінансовий рік в залежності від того, що більше [101].

Положення стало взірцем для багатьох національних законів поза межами ЄС, наприклад в Чилі, Японії, Бразилії, Аргентині, Кенії та локальних законів, як наприклад в штаті Каліфорнія, США [89].

Коли ми говоримо про інформацію, яку використовує держава, то може йти мова і про секретну інформацію з обмеженим доступом. Така інформація вимагає не меншого контролю за доступом до неї, умов зберігання та обробки, ідентифікації людей що уповноважені на роботу з такими даними. Адже розсекречення такої інформації може нести загрозу державній безпеці чи окремим її громадянам.

В наші дні демократична відкритість означає більше, ніж те, що громадяни можуть голосувати через регулярні проміжки часу на вільних і справедливих виборах. Вони також розраховують отримати доступ до урядових даних.

Держава довгий час було найбільшим генератором, збирачем і користувачем даних. Він веде записи про кожного народження, шлюб і смерть, збирає дані з усіх аспектів економіки і веде статистику ліцензій, законів і погоди. Однак до недавнього часу всі ці дані були надійно заблоковані. Навіть

коли вони є загальнодоступними, їх було важко знайти, а агрегування великої кількості друкованої інформації, як відомо, є культовим.

Але тепер громадяни та неурядові організації в усьому світі наполягають на отриманні доступу до загальнодоступних даних на національному, державному і муніципальному рівнях, а іноді державні чиновники їх з ентузіазмом підтримують.

Америка лідирує по доступу до даних. У свій перший повний робочий день президент Барак Обама видав президентський меморандум, в якому главам федеральних агентств пропонувалося надати якомога більше інформації, і закликав їх діяти з чітким припущенням.

В той же час, наприклад в Україні діє закон «Про доступ до публічної інформації» [16] та закон «Про інформацію» [15], який чітко визначає перелік інформації, яка не може бути засекреченою чи мати обмежений доступ, не можуть становити державної чи комерційної таємниці. Це так звана відкрита інформація, яку можуть збирати, обробляти чи публікувати всі без винятка.

До такої інформації відносяться дані:

- про бюджетні кошти та їх розпорядження;
- про право володіння, користування чи розпорядження державним та комунальним майном;
- про умови отримання державних чи комунальних коштів чи майна;
- декларації про доходи осіб та членів їх сімей, якщо вони претендують або обіймають посаду державного службовця, службовця органу місцевого самоврядування першої чи другої категорії;
- про стан довкілля, якість та безпеку харчових продуктів і предметів побуту;
- про катастрофи, аварії, надзвичайні ситуації що стались чи можуть статись;
- про стан здоров'я населення, його рівень життя, соціальне забезпечення тощо;
- про факти порушення прав і свобод людини і громадянина;

- про незаконні дії державних органів, органів місцевого самоврядування, посадових та службових осіб.

Окрім того, в ст. 29 сказано, що інформація з обмеженим доступом може бути поширено, якщо вона є суспільно необхідною, а шкода від її приховування переважає над шкодою від її приховування.

Підводячи підсумки цієї глави, ми можемо зрозуміти, що використання Big Data з кожним днем хоч і стає більш доступним у фінансовому та технічному плані, та пропорційно стають більш складними для використання з юридичних та етичних норм.

І якщо раніше головними перешкодами були державна монополія на інформацію, закритість та обмеженість доступу до певної з них, то наразі це зовсім не так. Навіть в Україні більшість інформації, що раніше належали виключно державі тепер є в публічному доступі і заборонити її отримання не можна за законом: це і судові документи, декларації чиновників, дані метеостанцій, вимірювання основних показників безпечного навколишнього середовища, тощо. А якщо порівнювати з деякими країнами Заходу — то тут навіть більше «простору» для некомерційних та приватних організацій, оскільки в багатьох країнах вони є платними (наприклад, доступ до судових документів в США), чи захищені авторським правом (інформація статистичних служб Великобританії захищена і не може бути використана для комерційних досліджень, хоч і є відкритою).

А от робота з персональними даними стає все більш складною. Раніше вона зовсім не контролювалась і більшість компаній збирали їх відкрито або таємно [19]. Тепер це є неприпустимим, а з кожним роком громадяни отримують все більше прав на свої персональні дані.

Роблячи висновки з розділу, ми можемо зрозуміти також і те, що Big Data — це не черговий ажітаж на IT-ринку, а системний, якісний перехід до нового етапу роботи з даними, заснованих на знаннях. За ефектом його можна порівняти з появою доступної комп'ютерної техніки в кінці минулого століття.

Те, що вчені передбачали ще на початку ХХ ст. стало реальністю. Кожна людина добровільно, усвідомлюючи це чи ні генерує величезні масиви інформації кожен день. Розпорядниками цієї інформації можуть бути як приватні особи, так і держава чи комерційні організації. І отримавши в свої руки ці дані вони будуть старатись максимально використати їх в своїх цілях.

Тепер, коли фінансово та технічно обробка та зберігання Big Data стали доступні кожному бажаючому, держава та великі корпорації втрачають монополію на роботу з ними.

Це накладає відбиток на їх подальший розвиток. Десятки різних технологій з'являються на ринку для поліпшення роботи з Big Data. Ними користуються в усіх сферах.

Одночасно, постає питання, що раніше не піднімалось: кому належать права на ці дані? З часом все більше країн звертаються до практики, що вся персональна інформація, яка часто лежить в основі досліджень в політичній сфері, належить виключно людині, до якої вона відноситься.

Законодавство країн та наддержавних утворень поступово змінюються, адаптуючись до сьогодення і починають захищати право на збереження та контроль над власними персональними даними, паралельно захищаючи їх від несанкціонованого доступу.

## РОЗДІЛ 3

### ПРАКСЕОЛОГІЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ BIG DATA В УМОВАХ СУЧАСНОЇ ПОЛІТИЧНОЇ ДІЙСНОСТІ

#### 3.1. Зарубіжний досвід використання Big Data

Використання Big Data в політичних процесах за кордоном та в Україні почалось масово на початку 2010-х років. Тоді, усвідомивши можливості, які перекривали попередні та йшли вперед, політичні партії, рухи, громадські та урядові організації, цілі країни взяли на озброєння аналітику з використанням великих об'ємів інформації.

В першій главі ми розглянемо декілька найбільш вдалих прикладів використання технології за кордоном для того, щоб проілюструвати можливості та результати такого підходу.

У сфері політики, враховуючи нездатність агентств, що займаються опитуваннями, завжди зробити якісну вибірку та провести її повноцінне опитування, важливим є пошук нових методів прогнозування результатів виборів та референдуму. Одним з таких способів стало використання сервісу Google Trends.

Google Trends — це вебсайт Google, який аналізує популярність найпоширеніших пошукових запитів у Пошуку Google в різних регіонах та на різних мовах. Вебсайт використовує графіки для порівняння обсягу пошуку різних запитів з часом.



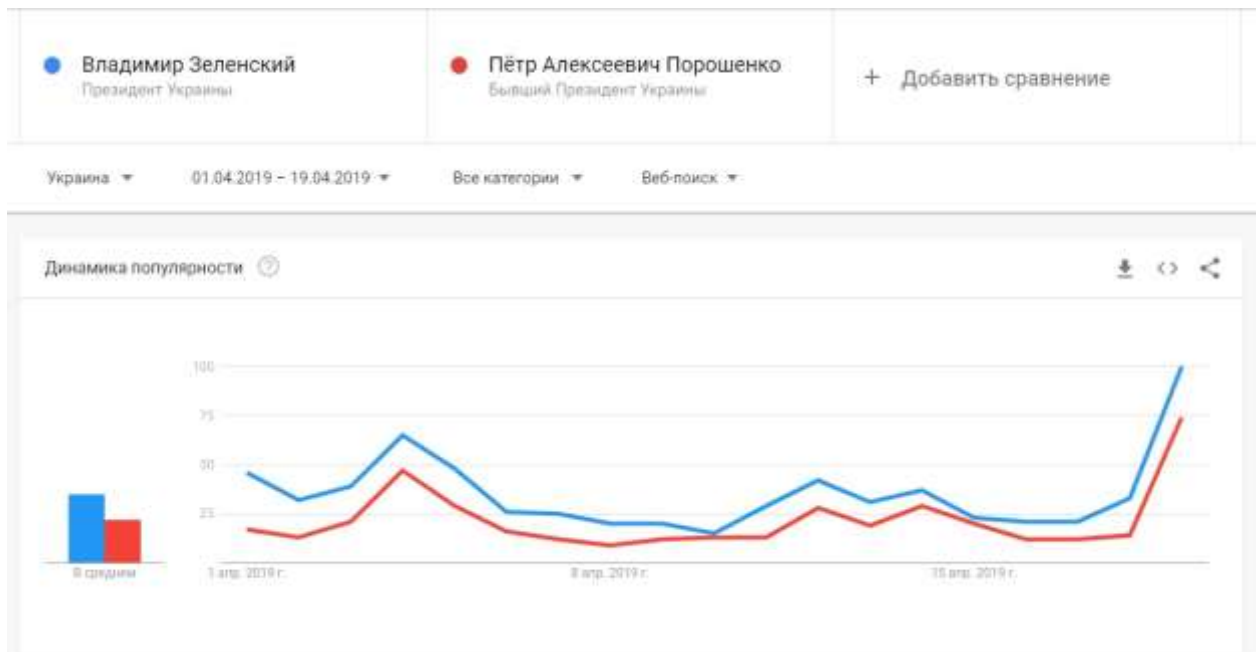


Рис. 3.1 Приклад аналізу передвиборчої популярності основних кандидатів на посаду Президента України у 2019 році перед другим туром виборів серед пошукових запитів.

За останні 10 років метод прогнозування референдуму за сервісом аналітики пошукових запитів був використаний не один раз.

Дослідники результатів відмітили, що під час оцінювання саме результатів референдуму, коли є не більше двох варіантів, цей метод зарекомендував себе досить позитивно. Так, на шести окремих випадках похибка була в межах статистичної від результатів традиційних опитувань та результатів референдумів [92]. До цих референдумів відносяться:

- шотландський референдум 2014 року;
- грецький референдум 2015 року;
- референдум Великобританії 2016 року;
- референдум Угорщини 2016 року;
- референдум Італії 2016 року;
- референдум Туреччини 2017 року;

Прогнози з використанням даних Google Trends у деяких моментах давали кращі, а в деяких — гірші порівняно з традиційними агентствами

опитування. Незважаючи на різницю у відсотках від офіційного результату, дослідження підтвердило, що результати використання Google Trends для намірів голосувати та прогнозів результатів є точними та, у багатьох випадках, кращими, ніж офіційні опитування. Вони стверджують, що цей метод буде застосований найближчим часом органами опитування та політичними дослідниками в цій галузі, оскільки це знижує фінансові витрати на дослідження та допоможе якщо не замінити, то підвищити ефективність та точність традиційних опитувань.

Однак деякі обмеження існують. Спочатку цей метод не міг застосовуватися в регіонах, де обмежене використання Інтернету (наприклад — Китай та Іран [65]), в тих, що мають низькі оцінки свободи преси (як то Республіка Білорусь чи В'єтнам [61]), або в тих, що мають низький рівень проникнення в Інтернет (наприклад — Еритрея чи ДР Конго [69]), де можлива маніпуляція даними (в Україні близько 30% коментарів, що залишають під політичними публікаціями в соціальних мережах публікуються ботами) та ін.

Крім того, зразок не може бути доведеним репрезентативним. Не всі користувачі Інтернету голосують, і не всі виборці використовують Інтернет для пошуку відповідних ключових слів на референдумі, отже, не кожне звернення може бути пов'язане з голосуванням на референдумі. Крім того, обмеження інструменту полягає в тому, що дані, отримані в різні часові моменти за один і той же часовий проміжок, можуть дещо відрізнятись.

А от в прогнозуванні політичних уподобань на виборах, а особливо там, де немає великого відриву у кандидатів, під час альтернативного дослідження виборів у США 2008 та 2010 роках [113] продемонстрував досить низький результат.

В цей самий час саме США можна назвати однією з перших країн, де були масово використані Big Data під час виборчих перегонів.

Технології, пов'язані з Big data успішно застосовувалися вже в ході обох передвиборчих кампаній Барака Обами — і в 2008-2009, і 2012-2013 роках, з метою оптимізації шляхів залучення пожертвувальців, підвищення ефективності

реклами, тактики проведення на місцях передвиборчих заходів з урахуванням специфічних настроїв конкретних аудиторій виборців, аналізу відносини громадської думки в цілому саме до даного кандидата від Демократичної партії США [117]. В ході передвиборчої кампанії 2012 року команда з 18 фахівців з розсилання електронної пошти в штабі Обама протестували більше десяти тисяч текстових варіантів такої розсилки з 18 різновидами ключових словосполучень. У підсумку найбільш успішна з комбінацій принесла в передвиборчу скарбничку Обама 2 673 278 доларів, Найменш успішна — тільки 403 603 доларів. Розроблений на основі даного аналізу алгоритм керованого процесу збору пожертвувань на передвиборчу кампанію Обама дозволив акумулювати в цілому 1,123 млрд доларів, в т. ч. 4,4 млн осіб перерахували через сайт 690 млн доларів [26].

В листопаді 2016 року, невдовзі після того, як Дональд Трамп переміг на президентських виборах, CNN повідомили, що виборчий штаб 45го президента співпрацював з Cambridge Analytica, яка була найвідомішою у світі політтехнологічною компанією, що в першу чергу покладається на технології Big Data. У грудні того ж року швейцарське видання Das Magazin випустило розслідування про те, що перемогу Трампу принесли великі дані — точніше, адресна агітація в мережі [81].

В Cambridge Analytica протестували користувачів соцмереж по 5 тис. Різних параметрів. Серед них — політичні пристрасті, кандидата на попередніх виборах, улюблені телешоу і серіали, бажані товари і ін. Потім людей розділили на невеликі групи і кожній групі показували своє звернення з агітацією — в соцмережах, фільмах або передачах, які вони дивляться (не тільки на ТБ, але і в мережі).

Як стверджує Міхаель Козінські, один з тих, хто почав створювати психометричні портрети по аналізу діяльності людини в соціальних мережах стверджує: проаналізувавши 68 лайків в Facebook йому достатньо, щоб визначити колір шкіри з вірогідністю 95%, його гомосексуальність (88% вірогідності) та якій партії США (Демократичній чи Республіканській) він

відає перевагу. Модель розвивається, і зараз вона може дізнатись релігійні вподобання, інтелектуальний розвиток, пристрасть до алкоголю та курінню та навіть те, чи розводились батьки до його повноліття [45].

Наразі модель може дізнатись все це після аналізу 10 лайків, а після 300 буде знати про нього більше, ніж його батьки, партнер та він сам про себе зможе розповісти.

Цікаво те, що аналітика даних на виборах в цих випадках ніколи не була санкціонована самим Facebook, принаймні офіційно. Дані збирались через додатки в цій мережі, де користувачам пропонували пройти різні опитування. Подібні додатки, як то «Дізнайся який ти фрукт» чи «Хто ти в Гаррі Поттері» наразі розповсюджуються і в українському сегменті. Але їх основна мета була збір інформації, на що користувачі добровільно надавали згоду, не вчитуючись в політику використання даних в цих додатках (а часто вони запитують і список друзів, фотографій, перелік уподобань та лайків, історію активності, тощо).

Ще в 2015 році Олександр Коган — професор психології Кембриджського університету — створив опитувальник *thisismydigitalife* і запустив його на платформі Facebook. Користувачі самі давали дозвіл на збір і передачу даних і навіть отримували гроші за проходження тесту. Зібрані дані він передав Cambridge Analytica. Всього, за оцінкою Facebook, компанія отримала дані про 87 млн користувачів [4].

З отриманими даними компанія змогла значно детальніше проробити мікротаргетинг реклами кандидатів, одночасно працюючи як з його електоратом, так і пригнічуючи бажання йти на вибори електорат Хілларі Клінтон [45].

Звісно, що не можна казати, що тільки Big Data принесли перемогу Трампу на виборах. Це досить хибна думка, яку роздмухують його конкуренти. Але, в той же час, не можна забувати, що така деталізація даних допомогли в рази знизити бюджет, який витрачається на рекламу, адже вони

могли значно краще розпізнавати, кому яку рекламу треба демонструвати та бити значно привільніше.

Використання технологій Big data в передвиборчих кампаніях означає можливість обробки практично в режимі реального часу величезних масивів різномірної інформації з колосальних за чисельністю різнопланових джерел і представлення результатів такої обробки в вигляді, що дозволяє приймати стратегічні рішення фактично по ходу подій, що відбуваються, замість того, щоб займатися пост-аналізом вже dokonаних явищ. Іноді роками тривають вивчення стану і настроїв громадської думки в рамках всієї країни, численні трудно систематизуємі зрізи думок самих різних цільових аудиторій, так чи інакше по ряду ключових позицій застарівають в тій чи іншій мірі на момент вироблення стратегії передвиборчої боротьби. Big data дає реальну можливість передвиборчим штабам мати комплексне сиюминутне уявлення про настрої [26],

Група дослідників використала неоднорідні дані, зібрані з різних Інтернет-платформ, для прогнозування загальних виборів у Тайвані 2016 року і дійшли важливих висновків.

По-перше, громадські думки в Інтернет-засобах масової інформації домінують над традиційними опитуваннями у Тайвані щодо прогнозування виборів як з точки зору передбачуваної сили, так і своєчасності. Але офлайн-опитування все ще можуть функціонувати для зменшення упередженості думок в Інтернеті.

По-друге, хоча онлайн-сигнали зближуються з наближенням дня виборів, просте повідомлення "Подобається" у Facebook незмінно є найсильнішим показником результату виборів.

Китай, в свою чергу, створює так звану систему «Соціального кредиту». Система тестується до 2020 року 43 муніципалітетами, кожен зі своїми критеріями, системою листів або балів та назвою: У Сучжоу це називається Plum Blossom Social Credit; в Сямені, Жасмин. Майже всі використовують дані із соціальних мереж або додатків для смартфонів, окрім складного

відеоспостереження. До 2020 року більшість великих міських громадських просторів Китаю будуть обладнані камерами розпізнавання обличчя за системою Skynet. У багатьох сільських районах проєкт "Гострі очі" дозволяє людям підключати свої телевізори чи смартфони до камер спостереження біля в'їзду до їх сіл [102].

Ідея полягає в тому, щоб завдяки збору та аналізу даних про поведінку формувати динамічний соціальний кредит кожного громадянина, по аналогу з кредитним рейтингом в банківській системі. Системи відеоспостереження, поєднанні з розпізнаванням обличчя мають змогу детально вивчати соціальні звички кожної особи: де і як часто вона буває, чи вчасно ходить на роботу, чи не переходить дорогу на червоне світло, поступається місцем в громадському транспорті тощо. Телефони та комп'ютери надають інформацію про інтереси людини, які сайти відвідує. Платіжні системи передають все, що потрібно знати про фінансову складову життя та його витрати. Разом ці та ще сотні параметрів формують особистий рейтинг [105].

Надалі цей рейтинг буде впливати на відношення до цих громадян та надання їм тих чи інших послуг. Так, вже почали діяти заборона на купівлю авіа- та залізничних квитків тим, кого було визнано «ненадійними», в тому числі вживали їжу у невстановлених місцях, голосно розмовляли чи включали музику в літаку чи потязі. Розглядається можливість виключення дітей з неблагонадійних сімей з черг на приватні навчальні заклади [97].

Таким чином, виходячи викладеного в цій главі матеріалу, ми можемо наглядно побачити, що по всьому світу вже активно використовуються технології Big Data, в тому числі в політичних та державних процесах.

Держава, як то відбувається в Китаї, отримує змогу вводити систему соціального рейтингу, яка фактично буде працювати без людського фактору. Більшість інформації виробляється під час автоматичної аналітики та обробки даних, призначаючи певний бал рейтингу громадянам. І вже спираючись на цей рейтинг влада може будувати ти чи інші відносини з людиною, карати чи надавати преференції.

Політтехнологи масово використовують персональні дані як отримані в результаті співпраці з компаніями, так і неофіційно, а часто з порушенням закону. Маючи цю інформацію та аналізуючи її, вони можуть більш детально взаємодіяти з аудиторією та економити бюджет компанії.

Громадські об'єднання отримали змогу працювати з відкритими даними для просування своїх ідей чи контролю за діями влади. Це — додаткова гарантія підтримки основних демократичних ідей, які базуються на повній відкритості всіх політичних процесів в країні.

### **3.2. Досвід використання Big Data в Україні**

Як громадян України, в першу чергу під час проведення дослідження нас повинні цікавити як досвід використання технології в нашій державі, так і можливості в цьому напрямку.

В цій главі ми розберемо основні засади доступу до інформації в Україні на додаток до тих, що були розглянуті в главі про юридичні аспекти та вивчено досвід використання Big Data в Україні некомерційними організаціями та владою.

У 2011 році Верховна Рада прийняла закон "Про доступ до публічної інформації", в якому говорилося про можливість безоплатного отримання публічної інформації за запитом до державного органу [16].

У 2015 році парламент змінив профільне законодавство. У тому ж році Державне агентство з питань електронного урядування презентувало єдиний портал відкритих даних data.gov.ua [43].

У 2016 році Україна приєдналася до "Хартії відкритих даних", що поклало на неї зобов'язання робити урядові дані відкритими суспільству за замовчуванням [42].

10 серпня 2018 року була запущена оновлена версія національного порталу даних. Він повинен стати єдиною платформою для всіх наборів відкритих даних України. Нова версія порталу працює на платформі SKAN. Це найпопулярніша в світі платформа для е-урядів [30].

Одними з перших Великі Дані в Україні почали використовувати великі компанії та корпорації, такі як: банківські установи, мобільні оператори, інтернет-сайти, та ін.

Вони через специфіку своєї роботи мали великі об'єми даних, що стосувались своїх клієнтів та їх дій. Використовуючи їх аналіз в маркетингових цілях, компанії знаходять більше точок дотику з існуючими та потенційними клієнтами.

Так, бази даних мобільних операторів можна вважати однією з найбільш бажаних. Вони зберігають інформацію про дзвінки, повідомлення, трафік, переміщеннях, рухах грошей на рахунку і багато іншого.

Наприклад, в Київстарі розповідають, що щоденний потік даних, доступний для аналізу, становить понад 8 терабайт на добу. Це безцінна інформація не тільки про технічний стан мережі, а й поведінці клієнта. А розуміння поведінки монетизуються дуже добре.

Недарма ними часто користуються навіть банки щоб зрозуміти платоспроможність своїх клієнтів для встановлення кредитного ліміту [22].

Оператори використовують аналіз великих даних, щоб утримувати існуючих клієнтів і збільшувати користування будь-якої послугою. У Vodafone кажуть, що раніше для цього ґрунтувалися на сотнях показників в розрізі одного номера. Зараз — на десятках тисяч [58]

Ще один інструмент мобільних операторів — так звані «теплові карти» (heat maps). Адже в базі зберігаються і дані про переміщення конкретних сім-карт. Якщо вибірка досить велика, оператор може показати їх маршрути, не розкриваючи особистих даних абонентів.

Банкам і ритейлерам цікаво дізнатися, де живе і як переміщається їх потенційний клієнт, щоб правильно розташувати відділення, магазини, банкомати. Міській владі і комунальним підприємствам це допомагає оптимізувати нинішні маршрути міського транспорту, запускати нові, розуміти, де будувати об'єкти інфраструктури. Або навіть завантажити аеропорти.



У 2015 році Київська міська державна адміністрація звернулась до Київстар для того, щоб ті допомогли зібрати дані про міграцію та населення Києва та міст-сателітів. Але дуже швидко ідея трансформувалась і ці самі дані були використані для оптимізації транспортної інфраструктури міста .



Рис. 3.2 Приклад візуалізації даних, що використовується КМДА для покращення транспортної інфраструктури та побудованої на базі інформації стільникового оператора [9].

Модель будувалась з січня по вересень 2015 року. В них враховувались дані стільникової мережі, популярність зупинок, завантаженість транспорту. В результаті влада отримала детальну інформацію, яка розкрила «слабкі» місця діючої системи: виявилось, що найпопулярніші міські маршрути — це транзитні поїздки на короткі дистанції, а найбільші пробки збираються далеко від центра міста; тобто інформація повністю протилежна тій, якою керувались раніше [9].

Також моніторинг допоміг виявити найбільш популярні зупинки, прибрати дублюючі маршрути, виокремити дороги для обладнання окремої полоси для громадського транспорту, тощо.

Наразі влада міста продовжує вивчати аналітику для того, щоб реалізувати мету: витрати часу для того, щоб дістатись з однієї точки Києва до будь-якої іншої не повинні перевищувати 60 хвилин. Такі дослідження значно детальніші, ніж традиційні вимірювання спостереженням для побудови транспортної схеми, та не потребує залучення такої кількості людино годин.

Крім того, моніторинг показав найбільш популярні зупинки і дублюються маршрути, дозволив зіставити самі «людні» маршрути з дорогами, по яких рухається найбільша частина громадського транспорту.

Що до використання даних приватним, громадським сектором чи компаніями, що не мають власних джерел, велике значення мали Закон України Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних [39] та Постанова Кабінету Міністрів України №835-1100 [41], що були прийняті у 2015 році та зобов'язали державні органи надавати всю публічну інформацію в формі відкритих даних і регулярно оновлювати їх. Наразі більшість з них акумулюються на відповідному вебпорталі. Станом на 12 грудня 2020 року там міститься 35 458 наборів даних [27].

31 жовтня 2019 року Верховна Рада прийняла в першому читанні Проект Закону про публічні електронні реєстри [40]. В разі прийняття цього Закону він повинен вступити в силу з 1 січня 2022 року [6].

Одним з найбільш масштабних прикладів використання Великих Даних в Україні за останні роки на нашу думку можна вважати оцінку чисельності наявного населення України, який було проведено у 2019–2020 роках за ініціативи 19-го Міністра Кабінету Міністрів України Дмитра Дубілета.

Передумовою цього заходу була ситуація, що склалась в Україні через брак інформації про населення. Традиційний перепис населення, що рекомендовано проводити кожні 10 років [46, 7], не відбувався з 2001 року та через певні обставини відкладався кожного разу на певний термін.

Наявних статистичних даних недостатньо, особливо в умовах тотального дублювання та втрат реєстрів. Окрім того, багато громадян

покинули домівки під час трудової еміграції, деякі території України були тимчасово втрачені через окупацію.

Після перемоги на виборах 6-го Президента України Володимира Зеленського, сформований ним уряд запланував [18] проведення традиційного перепису населення на грудень 2020 року з оприлюдненням результатів в січні 2021 року [44]. На це з бюджету заплановано видання 3,5 млрд. гривень.

Тому, в той час, як в Києві проводився пробний перепис населення [14], Дубілет заявив про можливість провести оцінку наявного на підконтрольній Києву території населення завдяки сучасним технологіям, використовуючи великі масиви даних.

23 січня команда Дмитра Дубілета презентувала оцінку чисельності наявного населення України станом на 1 грудня 2019 року [34].

Оцінка була проведена з використанням великого обсягу даних з різних джерел:

- дані мобільних операторів Київстар, Vodafone і Lifecell, які передали інформацію про кількість номерів телефонів, якими користуються громадяни України в період з лютого по березень 2019 р. [37, 6];
- все ті ж оператори стали базою для статистичного обстеження домогосподарств. Оскільки кожен засіб стільникового зв'язку має особистий, унікальний ідентифікатор (IMEI-номер), це дало можливість врахувати кількість мобільних номерів, що перебувають у власності однієї особи [37, 6];
- ми припускаємо, що враховувались також реєстрація кожного номера до тієї чи іншої базової станції. Оскільки цей інструмент дозволяє визначати розташування абонента з точністю до 100 метрів [51], можна вирахувати переміщення декількох мобільних телефонів одночасно, тобто коли одна людина користується одночасно декількома телефонами;
- були задіяні дані реєстрів, в тому числі: пенсійного фонду, дітей до 14 років. Завдяки особам віком більше 60 років, що отримують

пенсійні виплати, була вирахована частка населення за статевовіковою структурою в цій групі, після чого дані екстраполювалися на дані статево-вікової структури [37, 7];

- в повному об'ємі проаналізовані дані державного реєстру фізичних осіб. За допомогою РНОКПП (реєстраційний номер облікової картки платника податків) були встановлені зв'язки кожного громадянина з отриманням доходів чи соціальних виплат, звернень до державних установ, наявність в нього паспорта громадянина для виїзду за кордон, тощо [37, 8];
- отримана інформація про перетин державних кордонів України, кількість громадян на консульському обліку, за нашими оцінками, були використані для розрахунку показників міграції.

Фактично, було здійснено отримання державою з різних джерел, в тому числі — приватних певну кількість даних у великому об'ємі із різними структурами та не пов'язані між собою.

Ці дані були структуровані, було встановлено ключі для зв'язку між собою. Потім цей масив даних було проаналізовано, що призвело до можливості проведення оцінки наявного населення що найменше за наступними показниками:

- 1) розподіл населення за віком;
- 2) визначення статево-вікової структури;
- 3) розподіл за статтю та регіонами.

За словами Дмитра Дубілета, проведення такої оцінки було безкоштовним для державного бюджету, що незважаючи на обмеження та недоліки оцінки чисельності наявного населення перед традиційним переписом робить його перспективним інструментом для поверхневої аналітики та перевірки даних, що надаються державною службою статистики. Окрім того, в разі реформування принципу зберігання інформації державними органами та створення єдиних реєстрів складність такого аналізу зменшується, як і похибка результатів.

Така аналітика не є повноцінною з певної низки причин, але використовується вже в низці інших країн, в тому числі для підготовки до виборів, референдумів чи традиційного перепису населення як інструмент визначення наявного населення в певний проміжок часу.

Через відсутність великих грошових витрат, може проводитись з більшою періодичністю, ніж традиційний перепис.

А от трохи раніше, коли Михайло Федоров ще не був міністром, а лише працював у штабі Зеленського, вони напрацювали чудовий кейс з використання Big Data в політичній агітації.

Для створення психометричного портрета однієї була відкрита «анкета волонтера», де користувачам мережі пропонували відповісти на низку питань, в тому числі з персональними даними в обмін на можливість в подальшому працювати з партією, бути волонтером на виборах, чи навіть балотуватись від них у Парламент. Таких анкет було зібрано більше 400 тисяч.

## Стань моєю командою!

Змінити країну можна тільки разом із тобою! Не стій осторонь —  
приєднуйся до моєї команди! Я їду! Ти зі мною?

- 1 **Подай заявку**
- 2 Підтверди пошту
- 3 Заповни анкету
- 4 Отримай зворотній зв'язок

Даю згоду на обробку персональних даних

Рис. 3.3 Частина анкети для збору персональних даних, що була розміщена на сайті команди кандидата в Президенти України Володимира Зеленського

Наразі нам не вдалось знайти відкритої інформації про те, хто і скільки з заповнивши ці анкети справді були залучені до команди та чи є там ті, хто потім потрапив у парламент, але точно відомо, що аналіз цих даних серйозно вплинув на перемогу Володимира Зеленського.

За словами Федорова, анкети були оброблені та всі люди розділені на 32 сегменти. З них вилучили 7 ключових, з якими і працювали в подальшому [59]. Таким чином повторюється ситуація, як і за кордоном: люди добровільно надавали персональні дані, але не завжди розуміли як і для чого вони це робили.

І вже далі можна відстежити різницю між рекламними компаніями кожного з основних кандидатів в Президенти у 2019 році: Володимир Зеленський рекламував від 35 до 70 різних рекламних повідомлень для різних аудиторій, а Петро Порошенко — одним і тим самим повідомленням по всій аудиторії [58], при чому як лояльній до нього, так і не лояльній, не кажучи вже про розподіл на мікрогрупи.

Так, згідно з дослідженням Київської школи економіки та британського Open Data Institute, відкриті дані у 2017 році принесли економіці України більше 700 млн доларів США, тобто 0,67% від ВВП [30].

Дослідивши в цій главі етапи розвитку відкритих даних та використання технологій Big Data в нашій державі, ми можемо зробити висновки, що Україна є однією з держав, що активно розвивають цю сферу та використовують в багатьох сферах.

Якщо розглянути кейси, реалізовані командою шостого Президента України під час виборчої агітації, можна побачити, що наші спеціалісти проробили роботу навіть кращу, ніж західні колеги під час виборів в США та агітаційної компанії Brexit.

Після приходу уряду «технократів» до влади в Україні активізувалась робота з Big Data, що допомагає приймати ти чи інші політичні рішення, економлячи при цьому фінанси держави.

В той же час, це все знаходиться на початковому етапі і однозначно вимагає більш детального аналізу і постійного вдосконалення, щоб поставити цю технологію на служіння державі в повному обсязі. Вже зараз розглядаються ідеї впровадження технології Штучного Інтелекту на основі Big Data в медичну, судову, оборону сфери та ін. [57].

### **3.3. Перспективи використання технології в Україні**

Як ми виявили в попередній главі, Україна має не аби які амбіції у використанні технологій на базі Big Data в майбутньому, чому обіцяє сприяти влада з технократичним урядом на чолі.

Саме тому важливо в цій главі розглянути перспективи та можливості подальшого розвитку технології в Україні.

Фахівців з Big Data та штучного інтелекту сьогодні навчають в шести українських ВНЗ [36]:

- Львівський національний університет імені Івана Франка;
- Національний університет «Львівська політехніка»;
- Національний технічний університет України «Київський політехнічний інститут»;
- Харківський національний університет радіоелектроніки;
- Одеський національний політехнічний університет;
- Український Католицький Університет;

Держава заохочує молодих спеціалістів до розвитку своїх навиків в сфері роботи з Великими Даними у сферах, що можуть допомогти державі та громадянам взаємодіяти із владою чи аналізувати її дії.

Так, починаючи з 2017 року в Україні проводиться національний конкурс IT-проектів на основі відкритих даних Open Data Challenge. Призовий фонд у 2020 році склав 3,5 млн. грн [95].

До участі в конкурсах запрошуюються розробники, дослідники та активісти для розробки сервісів та продуктів, що сприятимуть вирішенню

проблем суспільства. За попередні роки учасники конкурсу розробили багато сервісів, таких як:

- За Parkanom — автоматична аналітика та оцінку ризик-фактору чиновника завдяки його деклараціям, зв'язкам з різними особами та компаніями, тощо;

- Відкрита влада — аналітична система для моніторингу та виявленню корупційної складової серед державних посадовців, політиків, комунальних підприємств;

- Суд на долоні — інструмент, що допомагає працювати з судовими рішеннями, перетворюючи дані Єдиного державного реєстру судових рішень та інші відкриті дані у структуровану систему, з якою зручно працювати.

Мотивація молодих команд науковців та розробників дає їм потенціал для розвитку, а державі допомагає в побудові інструментів, які можна використовувати, в тому числі в політичній сфері за мінімальний кошт, оскільки розробка кожного такого сервісу, з урахуванням всіх бюрократичних процедур зайняло б великий проміжок часу та міг коштувати набагато більше, ніж загальний призовий фонд конкурсу.

Взагалі, використання технологій Big Data в усіх сферах буде тільки поглиблюватись одночасно з тенденцією, що набирає обертів. З різними цілями, але вони вже використовуються в усіх державах світу, хоча і в різних масштабах.

Якщо взятись за детальне використання великих даних, країна може покращити життя своїх громадян в рази, нічого при цьому докорінно не змінюючи, навіть рівень корупції, що є одним з найвищих в Європі. В рамках дослідження пропонуємо звернути увагу на деякі сфери та вивчити можливі шляхи подальшого використання Big Data в них.

Судочинство. Прозвучали перші думки про використання штучного інтелекту в судах першої інстанції. З першого погляду ідея виглядає абсурдною, але якщо зрозуміти, що комп'ютерна система, на відміну від



суддів [49], не піддається корупційним схемам, підкупу, тиску та позбавлена людського фактору — стає більш зрозумілим, що ідея може бути реалізована.

В країні, де процент виправдувальних вироків не перевищує становить в різні роки від 0,44% до 0,76% [7], в той час як в країнах з розвинутою судовою системою складає 25 — 30% [13]. більшість людей може схилитись не те, що до штучного інтелекту, а навіть до жеребкування підкиданням монетою, а ні ж довіряти суддям.

Можемо говорити про те, що впровадження автоматичного судочинства шляхом аналізу справи штучним інтелектом в судах першої інстанції дозволить значно скоротити кількість суддів, зменшити навантаження на них (наразі це близько 600 справ на рік в одного судді [25]) та прибрати величезний пласт корупційних схем, побудований в судах.

Наприклад, в період у 2018 році до судів першої інстанції надійшло 3,3 млн справ. З них лише 36,5% — кримінальні. Всі інші — адміністративні, цивільні, господарські та що до адміністративних правопорушень. Величезна частка з них могла бути оброблена автоматично [25].

Це вже не кажучи про те, що наразі люди можуть очікувати рішення суду по своїй справі роками, в тому числі знаходячись в місцях позбавлення волі без будь-якого вироку.

До апеляційних інстанцій надійшло лише 447 тис справ. Звісно, що процент звернень до апеляційних інстанцій зросте, але не настільки великий процент. Частина буде також нівелюватись тими зверненнями, що виникали через суб'єктивне відношення суддів чи корупцію.

Станом на 13 грудня 2020 р. Єдиний державний реєстр судових рішень має в своїй базі 9 1425 390 документів. Це — приклад неструктурованих Великих Даних, що без проблем можуть бути використані для створення та навчання ядра штучного інтелекту і подальшого його роботи в судовій сфері.

Головною проблемою може стати зворотній бік технологій, а саме — можливість стороннього втручання в алгоритми системи для впливу на її результати [17]. В Україні, як в одній з найбільш корумпованих країн таке

питання постає ще більш гостро, оскільки подібні втручання вже були, наприклад, в систему нерухомості, яка відкрито втручалась в роботу для отримання хабара державними службовцями від громадян [12].

Охорона здоров'я. Навіть сьогодні більшість даних, що лікарі генерують про пацієнтів, як то: результати обстежень, аналізів, оглядів, діагнози, тощо — в паперовому вигляді, а ті, що вводяться в електронному — розрізненні по різним системам.

При цьому, створення так званої Мапи здоров'я на підставі таких даних не є новою.

Подумайте: коли ви викликаєте таксі зі свого смартфона, система розпізнає ваше місцезнаходження та ім'я, побажання що до типу авто, чи повинен салон бути для курців чи ні, чи є у вас домашні тварини та безліч інших даних, які дозволяють надати вам найбільш якісно послугу.

Коли людина викликає швидку допомогу, то вони не знають ні адреса, ні телефону з якого дзвонять. Якщо пацієнт непритомний, то вони не зможуть дізнатись які супутні діагнози в нього є, чи має він алергічну реакцію, номер його страхового полісу, тощо.

Якщо всі дані, що стосуються здоров'я людини будуть потрапляти в єдину базу та зберігатись там, то по перше — доктор, до якого звернувся пацієнт чи лікар швидкої допомоги зможе одразу отримати всю інформацію про пацієнта і почати лікування більш швидко та якісно.

В той самий час, штучний інтелект, обробляючи всі дані пацієнтів та знаходячи певні тенденції, зможе знаходити різні групи ризику, чим значно пришвидшить терміни постановки діагнозу, що зазвичай відіграє головну роль в якісному та ефективному лікуванні.

Якщо доповнити ці дані інформацією з мобільних телефонів та фітнес-девайсів людини, то ми зможемо оперувати ще й звичками, фізичною активністю пацієнта та його заняттям, переміщенням.

Це також є доцільним під час прогнозування чи постанови діагнозу, та набирає актуальності під час епідемій, коли можна бути точно зрозуміти, чи контактувала людина з хворим чи ні.

Карта здоров'я це платформа, яка автоматично відстежує та аналізує кілька каналів даних на 15 мовах — включаючи офіційні документи, новини, соціальні медіа, міжнародні організації охорони здоров'я, урядові вебсайти та блоги медичних працівників — для створення онлайнової візуалізації тенденцій захворювань. Цей підхід може сприяти ранньому виявленню спалахів хвороби, забезпечуючи швидкі та цілеспрямовані реакції [116].

Одне з подібних рішень було застосовано під час спалаху Еболи в Західній Африці у 2014 році.

Комунальні послуги. Дані, отримані з мобільних пристроїв, можуть бути ефективно використані для аналізу поведінки людей та покращення інфраструктури відповідно до неї. На дворі 2020 рік, але в Україні все ще робиться за радянським взірцем: так, як потрібно, а не так, як зручно.

Наприклад, велика кількість ДТП за участю пішоходів трапляється через перехід проїзної частини в недозволеному місці. Але переходять там люди не тому, що вони погані, а тому що там їм зручніше. Якщо аналітика показує, що всі переходять дорогу за 15 метрів від переходу (бо, наприклад, там поблизу знаходиться крамниця) — то чому потрібно продовжувати обладнувати перехід там, де його за якихось обставин призначили ще у 60-х роках минулого століття?

Окрім того, дані з мобільних пристроїв та супутникових систем можуть бути використані для оцінки кількості та якості вже наданих послуг. Так, наприклад, в Індії використовуються супутникові знімки для оцінки постачання електроенергії у віддалені села [85].

Громадська безпека. Держава може використовувати великі дані, розпізнавання образів та прогностичну аналітику для розкриття злочинів і прийняття кращих рішень щодо їх подальшого попередження.

Так, наприклад, в багатьох країнах відсутня дорожня поліція, так як всі дорожні порушення фіксуються автоматично. Подібний проєкт наразі вже імплементують в Україні [60].

Що стосується прогностичної моделі, то в США існує комерційна система PredPol, до якої звертаються як державні, так і приватні правоохоронні структури. Аналізуючи дані про злочини, виклики поліції, пересування громадян, система визначає, де і коли з більшою вірогідністю відбудуться злочини, що дозволяє ефективно розподіляти свої ресурси та запобігати ним [62]. Бразилія, Великобританія та Нідерланди використовують подібні підходи.

Поєднуючи таку аналітику зі змінами в громадський простір (як то освітлення вулиць, відмова від підземних переходів, відкриття громадських закладів) та зміни в маршрутах патрулювання поліції можна значно зменшити рівень злочинності в країні.

Боротьба з корупцією. Не треба забувати, що і корупцію завдяки великим даним подолати значно простіше. В принципі, саме боротьба з нею і стала основною рушійною силою для відкриття багатьох публічних даних в Україні. Вже зараз їх аналіз призводить до виявлення десятків схем, які, нажаль, не знаходять продовження в судах. Це знав відсилає нас до проблеми судочинства.

Таким чином ми можемо побачити, що Big Data не просто входять в наше життя, а стають його частиною. Майже в кожній сфері вони можуть бути імплементовані та приносити користь при розумному підході та врахуванні ризиків.

Такими ризиками можуть бути: можливість несанкціонованого втручання з різною метою та маніпулювання. Саме останній пункт є одним із найбільш проблемних для політичних компаній, оскільки використовуючи Big Data політики отримують значно більше можливостей щодо впливу на думки громадян.

Несанкціоноване втручання ж — значно ширша проблема. Тут можна казати і про бажання певної особи чи групи вплинути на результати аналізу Великих Даних, чи самі дані, які можуть бути використані в різних цілях, у тому числі — кримінальних. Про це говорить і прикра статистика постійного витоку даних із великих корпорацій, яку часто замовчують [108] та банальна торгівля цими даними на чорному ринку [1].

Отже, одночасно з дослідженням перспектив використання Big Data в Україні та інших країнах світу і їх імплементації потрібно приділяти велику увагу їх безпеці та захисту від стороннього втручання, як саму інформацію, так і результат її обробки.

## ВИСНОВКИ

Епоха Big Data вже наступила. В основі багатьох аналітичних та прогностичних процесів по всьому світу, у різних сферах вже лежать великі об'єми інформації, що зростають із кожною хвилиною та потребують детального вивчення.

Технологія Big Data — це досить широке поняття, яке й досі не має точного визначення, але використовується майже в усіх сферах сучасного життя, поліпшуючи ефективність роботи аналітиків та обробки даних у десятки та сотні разів.

З кожним днем технології обробки, збору та зберігання великих даних стають усе доступнішими, а тому вони будуть тільки зростати та еволюціонувати в подальшому.

Ми виклали основні поняття та їх пояснення для подальшого маніпулювання ними в рамках нашого дослідження. Структурувавши теорію, узагальнивши та систематизувавши певні системоутворюючі поняття, ми змогли знизити складності дослідження, що виходили з міждисциплінарного характеру роботи.

Також, дослідження використання технології Big Data в політичних процесах вимагало від нас використання цілої низки різноманітних науково-дослідницьких принципів та методів. Саме їх застосування в основі нашого дослідження стали актуальними для висвітлення цього питання та його глибокого, всеосяжного вивчення.

Дослідивши основні методи та техніки, а також інструменти для обробки та збереження Big Data на сучасному етапі, ми дійшли до висновків, що екосистема Big Data постійно зростає, нові технології з'являються, а старі покращуються з кожним днем, що розширює можливості для досліджень, зменшуючи фінансовий та поріг знань для входження аналітиків в цю сферу.

Також, динамічний розвиток технологій демонструє попит, який тільки зростає з кожним днем, оскільки тенденція є сталою і використання Big Data в

усіх сферах життя, у тому числі політичних процесах будуть лише набирати обертів.

Та під час роботи з даними аналітики все частіше стикаються з новими викликами. Якщо раніше це були технічні та фінансові перепони, то тепер на перший план усе частіше виходять юридичні та етичні аспекти з використання Big Data, оскільки в їхній основі можуть лежати персональні, конференційні дані або дані з обмеженим доступом, а результати аналітики можуть нести загрозу державі чи певним громадянам, дадуть змогу маніпулювання чи впливу на настрої населення.

Раніше в цих питаннях головними перешкодами були: державна монополія на інформацію, закритість та обмеженість доступу до певної з них. Наразі це зовсім змінюється. Навіть в Україні більшість інформації, що раніше належали виключно державі тепер є в публічному доступі й заборонити її отримання не можна за законом.

Якщо порівнювати з деякими країнами Заходу — то тут, навіть, більше «простору» для некомерційних та приватних організацій, оскільки в багатьох країнах вони є платними (наприклад, доступ до судових документів у США), чи захищені авторським правом (інформація статистичних служб Великобританії захищена й не може бути використана для комерційних досліджень, хоч і є відкритою).

А от робота з персональними даними стає все більш складною. Раніше вона зовсім не контролювалась і більшість компаній збирали їх відкрито або таємно. Тепер це є неприпустимим, а з кожним роком громадяни отримують все більше прав на свої персональні дані.

Розібравши певний закордонний та вітчизняний досвід у використанні Big Data, у тому числі в політичних та державних процесах, ми наглядно побачили можливості та перспективи технології.

Наприклад, у Китаї держава вводить систему соціального рейтингу, яка фактично буде працювати без людського фактору. Більшість інформації виробляється під час автоматичної аналітики та обробки даних, призначаючи

певний бал рейтингу громадянам. І вже спираючись на цей рейтинг влада може будувати ти чи інші відносини з людиною, карати чи надавати переваги.

Політтехнологи масово використовують персональні дані як отримані в результаті співпраці з компаніями, так і неофіційно, а часто з порушенням закону. Маючи цю інформацію та аналізуючи її, вони можуть більш детально взаємодіяти з аудиторією та економити бюджет компанії.

Громадські об'єднання отримали змогу працювати з відкритими даними для просування своїх ідей чи контролю за діями влади. Це — додаткова гарантія підтримки основних демократичних ідей, які базуються на повній відкритості всіх політичних процесів в країні.

Україна є однією з держав, що активно розвивають цю сферу та використовують в багатьох сферах.

Командою шостого Президента України під час виборчої агітації проробили, можливо, навіть більш глибоку роботу, ніж західні колеги під час виборів у США та агітаційної компанії Brexit.

Після приходу уряду «технократів» до влади в Україні активізувалась робота з інтеграції Big Data в усі сфери державного управління, що допомагає приймати ті чи інші політичні рішення, економлячи при цьому бюджетні кошти.

Водночас, це все знаходиться на початковому етапі і однозначно вимагає більш детального аналізу і постійного вдосконалення, щоб поставити цю технологію на служіння державі в повному обсязі. Але при тому Big Data не просто входять в наше життя, а стають його частиною. Майже в кожній сфері вони можуть бути імплементовані та приносити користь при розумному підході та врахуванні ризиків.

До ризиків можна віднести: можливість несанкціонованого втручання з різною метою та маніпулювання.

Отже, одночасно з дослідженням перспектив використання Big Data в Україні та інших країнах світу і їх імплементации потрібно приділяти велику



увагу їх безпеці та захисту від стороннього втручання, як саму інформацію, так і результат її обробки.

Проаналізувавши всі аспекти та особливості використання технології Big Data в політичних процесах, можна стверджувати, що підтверджуються основна та додаткова гіпотеза дослідження. Використання Big Data дає можливість значно ефективніше та з меншими витратами проводити дослідження та реалізовувати певні кроки в політичній сфері, а в перспективі Big Data стануть основою для досліджень у політичній сфері, але, водночас, може суттєво змінитись підхід до використання певних масивів даних.

## СПИСОК ЛІТЕРАТУРИ

1. Адошев В. Наши с вами персональные данные ничего не стоят. Хабр. URL: <https://habr.com/ru/post/423947/> (дата звернення: 13.12.2020).
2. Бекетов А. Искусственный интеллект в суде. Euronews. URL: <https://ru.euronews.com/2019/01/28/eu-robojudge-courts> (дата звернення: 13.12.2020).
3. Блокчейн. Вікіпедія. URL: <https://uk.wikipedia.org/wiki/Блокчейн> (дата звернення: 13.12.2020).
4. Большие данные и большая политика: как технологии влияют на выборы. РБК. URL: <https://trends.rbc.ru/trends/industry/5ef4c3299a7947db501485f8> (дата звернення: 13.12.2020).
5. Ващенко К. О., Корнієнко В. О. Політичний процес. URL: [https://web.posibnyky.vntu.edu.ua/fmib/13vashenko\\_politologiya\\_dlya\\_vchitelya/22..htm](https://web.posibnyky.vntu.edu.ua/fmib/13vashenko_politologiya_dlya_vchitelya/22..htm) (дата звернення: 13.12.2020).
6. В Украине может появиться Реестр реестров: за основу принят проект. ЮРЛИГА. URL: [https://jurliga.ligazakon.net/news/190564\\_v-ukraine-mozhet-roavitsya-reestr-reestrov-za-osnovu-prinyat-proekt](https://jurliga.ligazakon.net/news/190564_v-ukraine-mozhet-roavitsya-reestr-reestrov-za-osnovu-prinyat-proekt) (дата звернення: 26.03.2020).
7. В Украине в течение последних 5 лет сохраняется критически низкий процент оправдательных приговоров — исследование. Украинский кризисный медиа-центр. URL: <https://uacrisis.org/ru/74463-analysis-of-court-decisions> (дата звернення: 13.12.2020).
8. Герасюкова М. Билл Гейтс предупредил Трампа о пандемии еще четыре года назад. Газета.Ru. URL: [https://www.gazeta.ru/tech/2020/05/12/13080223/feel\\_terrible.shtml](https://www.gazeta.ru/tech/2020/05/12/13080223/feel_terrible.shtml) (дата звернення: 13.12.2020).
9. Глущенко Н. Большие данные большого города: как Big Data меняет жизнь Киева. AIN.UA. URL: <https://ain.ua/special/big-data-in-kyiv/> (дата звернення: 13.12.2020).
10. Грабченко А. І., Федорович В. О., Гаращенко Я. М. МЕТОДИ НАУКОВИХ ДОСЛІДЖЕНЬ. Х. : Національний технічний університет «Харківський політехнічний інститут», 2009. URL:

[http://repository.kpi.kharkov.ua/bitstream/KhPI-](http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/5823/1/Hrabchenko_Metody_naukovykh_2009.pdf)

[Press/5823/1/Hrabchenko\\_Metody\\_naukovykh\\_2009.pdf](http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/5823/1/Hrabchenko_Metody_naukovykh_2009.pdf) 13.12.2020.

11. Демидов Д. Ю. Супутникова археологія як новий погляд на залучення ентузіастів до пошуку історичних пам'яток.: *Problems and achievements of modern science, Корк* , 19. С. 8–9.

12. Дзюба А. В Минцифре обнаружили “странный скрипт от предшественников.” LIGA.net. URL: <https://tech.liga.net/technology/novosti/v-mintsifre-obnarujili-strannyy-skript-ot-predshestvennikov> (дата звернення: 13.12.2020).

13. Диденко А. Оправдательные приговоры. Информационный портал Харьковской правозащитной группы. URL: <http://khpg.org/index.php?id=1602496346> (дата звернення: 13.12.2020).

14. З 1 грудня в частині Києва починається “пробний перепис населення.” Українська правда. URL: <https://www.pravda.com.ua/news/2019/12/1/7233530/> (дата звернення: 26.03.2020).

15. Закон України “Про інформацію.” Верховна Рада України. URL: <https://zakon.rada.gov.ua/laws/show/2657-12#Text> (дата звернення: 13.12.2020).

16. Закон України “Про доступ до публічної інформації.” Верховна Рада України. URL: <https://zakon.rada.gov.ua/laws/show/2939-17#Text> (дата звернення: 13.12.2020).

17. Заплата Т. С. Искусственный интеллект в вопросе вынесения судебных решений, или ИИ-судья. *Вестник Университета имени О. Е. Кутафина*. 2019. No. 56. С. 160–168. URL: <https://cyberleninka.ru/article/n/iskusstvennyu-intellekt-v-voprose-vyneseniya-sudebnyh-resheniy-ili-ii-sudyu> (дата звернення: 13.12.2020).

18. Кабмін остаточно узгодив терміни проведення перепису населення | Українська правда. URL: <https://www.pravda.com.ua/news/2019/11/6/7231111/> (дата звернення: 26.03.2020).

19. Как Facebook собирает и продает данные пользователей. TAdviser. URL: [https://www.tadviser.ru/index.php/Статья:Как\\_Facebook\\_собирает\\_и\\_](https://www.tadviser.ru/index.php/Статья:Как_Facebook_собирает_и_)

продает\_данные\_пользователей (дата звернення: 13.12.2020).

20. Какие термины помогут ориентироваться в мире больших данных. Российская газета. URL: <https://rg.ru/2017/10/03/kakie-terminy-pomogut-orientirovatsia-v-mire-bolshih-dannyh.html> (дата звернення: 13.12.2020).

21. Каламбет С. В. Конспект лекцій з дисципліни «Методологія наукових досліджень». Дніпродзержинськ : Дніпродзержинський державний технічний університет, 2015. URL: <http://www.dstu.dp.ua/Portal/Data/7/36/7-36-kl28.pdf> (дата звернення: 13.12.2020).

22. Карпусь В. При оценке надёжности заёмщика банки начинают использовать данные операторов сотовой связи. ИТС.ua. URL: <https://itc.ua/blogs/pri-otsenke-nadyozhnosti-zayomshhika-banki-nachinayut-ispolzovat-dannyye-operatorov-sotovoy-svyazi/> (дата звернення: 13.12.2020).

23. Колодчак О. М. Интеллектуальный анализ данных. 2013. С. 49–58. URL: [http://www.irbis-nbuv.gov.ua/cgi-bin/irbis\\_nbuv/cgiirbis\\_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE\\_FILE\\_DOWNLOAD=1&Image\\_file\\_name=PDF/VNULPKSM\\_2013\\_773\\_11.pdf](http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/VNULPKSM_2013_773_11.pdf).

24. Лисовенко А. NoSQL базы данных: понимаем суть. Хабр. URL: <https://habr.com/ru/post/152477/> (дата звернення: 13.12.2020).

25. Мамченко Н. Количество судей в окружных админсудах предлагают уменьшить на 17%, а в местных хозсудах — на 25%: новые нормативы нагрузки для определения штата. Судебно-юридическая газета. URL: <https://sud.ua/ru/news/publication/169337-kolichestvo-sudey-v-okrzhnykh-adminsudakh-predlagayut-umenshit-na-17-a-v-mestnykh-khozsudakh-na-25-novye-normativy-nagruzki-dlya-opredeleniya-shtata> (дата звернення: 13.12.2020).

26. Меньшиков П. В. Big data и предвыборная кампания Д.Трампа. Портал МГИМО. URL: <https://mgimo.ru/about/news/experts/big-data-i-predvybornaya-kampaniya-trampa/> (дата звернення: 13.12.2020).

27. Наборы данных. Data.gov.ua. URL: <https://data.gov.ua/dataset> (дата звернення: 13.12.2020).

28. Наука о данных. Википедия. URL: <https://ru.wikipedia.org/wiki/>

Наука\_о\_данных (дата звернення: 13.12.2020).

29. Неструктурированные данные. Википедия. URL: [https://ru.wikipedia.org/wiki/Неструктурированные\\_данные](https://ru.wikipedia.org/wiki/Неструктурированные_данные) (дата звернення: 13.12.2020).

30. Новак Н. Сотни миллионов долларов в год за дату: как открытые данные изменили Украину. Экономическая правда. URL: <https://www.epravda.com.ua/rus/publications/2018/09/4/640114/> (дата звернення: 13.12.2020).

31. Облачные вычисления. Википедия. URL: [https://ru.wikipedia.org/wiki/Облачные\\_вычисления](https://ru.wikipedia.org/wiki/Облачные_вычисления) (дата звернення: 13.12.2020).

32. Общий регламент по защите данных. Википедия. URL: [https://ru.wikipedia.org/wiki/Общий\\_регламент\\_по\\_защите\\_данных](https://ru.wikipedia.org/wiki/Общий_регламент_по_защите_данных) (дата звернення: 02.12.2020).

33. Описание. Википедия. URL: <https://ru.wikipedia.org/wiki/Википедия:Описание> (дата звернення: 13.12.2020).

34. Оприлюднено результати оцінки чисельності наявного населення України. Кабінет Міністрів України. URL: <https://www.kmu.gov.ua/news/oprilyudneno-rezultati-ocinki-chiselnosti-nayavnogo-naselennya-ukrayini> (дата звернення: 26.03.2020).

35. О проекте. OpenStreetMap. URL: <https://www.openstreetmap.org/about> (дата звернення: 13.12.2020).

36. Освітні програми в сфері штучного інтелекту. Міністерство та Комітет цифрової трансформації України. URL: [https://thedigital.gov.ua/lms\\_ai](https://thedigital.gov.ua/lms_ai) (дата звернення: 13.12.2020).

37. Оцінка чисельності наявного населення України. К. URL: [http://www.ukrstat.gov.ua/Noviny/new2020/zmist/novini/OnU\\_01\\_12\\_2019.pdf](http://www.ukrstat.gov.ua/Noviny/new2020/zmist/novini/OnU_01_12_2019.pdf).

38. Політичний процес. Вікіпедія. URL: [https://uk.wikipedia.org/wiki/Політичний\\_процес](https://uk.wikipedia.org/wiki/Політичний_процес) (дата звернення: 13.12.2020).

39. Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних від 09.04.2015 № 319-VIII. URL:

<https://zakon.rada.gov.ua/laws/show/319-19> (дата звернення: 26.03.2020).

40. Проект Закону про публічні електронні реєстри. Офіційний портал Верховної Ради України. URL: [https://w1.c1.rada.gov.ua/pls/zweb2/webproc4\\_1?pf3511=66772](https://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=66772) (дата звернення: 26.03.2020).

41. Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних від 21.10.2015 № 835. URL: <https://zakon.rada.gov.ua/laws/show/835-2015-п> (дата звернення: 26.03.2020).

42. Про затвердження плану дій з реалізації принципів Міжнародної хартії відкритих даних. Кабінет Міністрів України. URL: <https://www.kmu.gov.ua/npas/pro-zatverdzhennya-planu-dij-z-realizaciyi-principiv-mizhnarodnoyi-hartiyi-vidkritih-danih> (дата звернення: 13.12.2020).

43. Про портал. Data.gov.ua. URL: <https://data.gov.ua/pages/about> (дата звернення: 13.12.2020).

44. Про проведення у 2020 році Всеукраїнського перепису населення. К. URL: [http://www.ukrcensus.gov.ua/ukr/laws/2016/kmu\\_581-p.pdf](http://www.ukrcensus.gov.ua/ukr/laws/2016/kmu_581-p.pdf) (дата звернення: 26.03.2020).

45. Расследование Das Magazin: как Big Data и пара ученых обеспечили победу Трампу и Brexit. The Insider. URL: <https://theins.ru/politika/38490> (дата звернення: 13.12.2020).

46. Рекомендации Конференции европейских статистиков по проведению переписей населения и жилищного фонда 2020 года. Женева. .

47. Резниченко Р. В., Тимашова Л. А. Економіко - математичне моделювання соціально - економічних систем. *Економіко-математичне моделювання соціально-економічних систем: Зб. наук. пр.* 2013. Р. 200–211. URL: [http://irbis-nbuv.gov.ua/cgi-bin/irbis\\_nbuv/cgiirbis\\_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE\\_FILE\\_DOWNLOAD=1&Image\\_file\\_name=PDF/emmses\\_2013\\_18\\_15.pdf](http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/emmses_2013_18_15.pdf).

48. Сабрі К. Н. ТЕХНОЛОГІЯ BIG DATA В ЕЛЕКТОРАЛЬНОМУ ПРОЦЕСІ. *Young Scientist*. 2018. Vol. 6, No. 58. Р.

49. Симуха А., Шепель Т. Злочинна організація суддів. Українська

правда. URL: <https://www.pravda.com.ua/articles/2014/08/21/7035201/> (дата звернення: 13.12.2020).

50. Сметанин С. Анализ тональности текстов с помощью сверточных нейронных сетей. Хабр. URL: <https://habr.com/ru/company/mailru/blog/417767/> (дата звернення: 13.12.2020).

51. Стільниковий зв'язок. Вікіпедія. URL: [https://uk.wikipedia.org/wiki/Стільниковий\\_зв'язок](https://uk.wikipedia.org/wiki/Стільниковий_зв'язок) (дата звернення: 26.03.2020).

52. Сулейманов С. Американская разведка призналась в сборе данных интернет-пользователей. Lenta.ru. URL: <https://lenta.ru/articles/2013/06/10/bigbro> (дата звернення: 13.12.2020).

53. Сутність поняття «методи дослідження». studies in ua. URL: <http://studies.in.ua/ru/pedagogika-shpargalki/1404-sutnst-ponyattya-metodi-dosldzhennya.html> (дата звернення: 13.12.2020).

54. Теорія розпізнавання образів. Державний університет телекомунікацій. URL: [http://www.dut.edu.ua/ua/news-1-576-8731-teoriya-rozpiznavannya-obraziv\\_kafedra-shtuchnogo-intelektu](http://www.dut.edu.ua/ua/news-1-576-8731-teoriya-rozpiznavannya-obraziv_kafedra-shtuchnogo-intelektu) (дата звернення: 13.12.2020).

55. Уварова А. Apache Spark: гайд для новичков. Medium. URL: <https://medium.com/nuances-of-programming/apache-spark-гайд-для-новичков-959145ef6167> (дата звернення: 13.12.2020).

56. Умения и предпочтения программистов разных стран: исследование от DataArt. DataArt. URL: <https://www.dataart.ru/media-mentions/umeniya-i-predpochteniya-programmistov-raznykh-stran-issledovanie-ot-dataart/> (дата звернення: 13.12.2020).

57. Федоров М. Сьогодні на Уряді прийнята Концепція розвитку сфери штучного інтелекту в Україні. Офіційна Facebook-сторінка Михайло Федорова. URL: <https://www.facebook.com/mykhailofedorov.com.ua/posts/237058284424090> (дата звернення: 13.12.2020).

58. Шишацкий Е. 8 технологических революций Украины. Революция четвертая: Big Data. Liga.net. URL: [https://project.liga.net/projects/tech\\_revolutions\\_bigdata/](https://project.liga.net/projects/tech_revolutions_bigdata/) (дата звернення: 13.12.2020).

59. Шишацкий Е., Юрасов С. Михаил Федоров, куратор диджитал-кампании Зеленского. Интервью. LIGA.net. URL: <https://tech.liga.net/technology/interview/pochemu-poroshenko-proigral-intervyu-s-onlayn-strategom-zelenskogo> (дата звернення: 13.12.2020).

60. Яровая М. С 1 июля камеры начнут фиксировать не только превышение скорости, но и другие нарушения ПДД. AIN.UA. URL: <https://ain.ua/2020/06/26/novovvedeniya-v-pdd-gerashchenko/> (дата звернення: 13.12.2020).

61. 2020 World Press Freedom Index. Reporters without borders. URL: <https://rsf.org/en/ranking/2020> (дата звернення: 13.12.2020).

62. About PredPol. PredPol. URL: <https://www.predpol.com/about/> (дата звернення: 13.12.2020).

63. Allen R., Scott P. Data Center plan called privacy invasion. *The Lewiston Daily Sun*. 1966. P. URL: <https://news.google.com/newspapers?id=ZGogAAAAIBAJ&sjid=3GYFAAAAIBAJ&pg=933,5465131&dq=data-center&hl=en> (дата звернення: 14.12.2020).

64. Amazon.com: Lto Ultrium 8 Data Cartridge: Computers & Accessories. URL: [https://www.amazon.com/Lto-Ultrium-8-Data-Cartridge/dp/B077LK25K9/ref=sr\\_1\\_1?dchild=1&keywords=LTO-8+12%2C8TB&qid=1585224149&sr=8-1](https://www.amazon.com/Lto-Ultrium-8-Data-Cartridge/dp/B077LK25K9/ref=sr_1_1?dchild=1&keywords=LTO-8+12%2C8TB&qid=1585224149&sr=8-1) (дата звернення: 26.03.2020).

65. Annual Freedom on the Net report. Freedom House. URL: <https://freedomhouse.org/countries/freedom-net/scores?sort=asc&order=Total Score and Status> (дата звернення: 13.12.2020).

66. Big-data noun. OxfordLearnersDictionaries. URL: <https://www.oxfordlearnersdictionaries.com/definition/english/big-data?q=big+data> (дата звернення: 13.12.2020).



67. Big Data. IT.UA. URL: <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>. (дата звернення: 10.12.2020).

68. Blockchains - The great chain of being sure about things. The Economist. URL: <https://www.economist.com/briefing/2015/10/31/the-great-chain-of-being-sure-about-things> (дата звернення: 13.12.2020).

69. Country Comparison by Internet users. The World Factbook of the Central Intelligence Agency. URL: <https://www.cia.gov/library/publications/the-world-factbook/fields/204rank.html> (дата звернення: 13.12.2020).

70. COVID-19 Map. Johns Hopkins Coronavirus Resource Center. URL: <https://coronavirus.jhu.edu/map.html> (дата звернення: 13.12.2020).

71. Cox M. Application-Controlled Demand Paging for Out-of-Core Visualization. *8th IEEE Visualization '97 Conference*. 1997. P. 235–244. URL: [https://www.evl.uic.edu/cavern/rg/20040525\\_renambot/Viz/parallel\\_volviz/paging\\_outofcore\\_viz97.pdf](https://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf) (дата звернення: 14.12.2020).

72. Davenport T. H., Patil D. J. Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review. URL: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century> (дата звернення: 13.12.2020).

73. DeWitt S., Cohen J. NASA Goddard Introduces the NASA Center for Climate Simulation. NASA. URL: <https://www.nasa.gov/centers/goddard/news/releases/2010/10-051.html> (дата звернення: 13.12.2020).

74. Deep Blue. IBM. URL: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deerblue/> (дата звернення: 13.12.2020).

75. Delort P. Harnessing data as a new source of growth: Big data analytics and policies. Paris. URL: [https://www.oecd.org/sti/ieconomy/Session\\_3\\_Delort.pdf#page=6](https://www.oecd.org/sti/ieconomy/Session_3_Delort.pdf#page=6) (дата звернення: 13.12.2020).

76. Disk Drive Prices 1955+. URL: <https://jcmitt.net/diskprice.htm> (дата звернення: 26.03.2020).

77. Exabyte 307265 8mm D8 160m XL Helical Scan 7/14GB Data Tape Cartridge: Home Audio & Theater. Amazon. URL: <https://www.amazon.com/Exabyte-307265-160m-Helical->

Cartridge/dp/B006L26ETW (дата звернення: 26.03.2020).

78. Fritz Pfleumer. Wikipedia. URL: [https://en.wikipedia.org/wiki/Fritz\\_Pfleumer](https://en.wikipedia.org/wiki/Fritz_Pfleumer) (дата звернення: 14.12.2020).

79. Gates B. The next outbreak? We're not ready. YouTube. URL: [https://www.youtube.com/watch?v=6Af6b\\_wyiwI&ab\\_channel=TED](https://www.youtube.com/watch?v=6Af6b_wyiwI&ab_channel=TED) (дата звернення: 13.12.2020).

80. Gmail back soon for everyone. Official Gmail Blog. URL: <https://gmail.googleblog.com/2011/02/gmail-back-soon-for-everyone.html> (дата звернення: 26.03.2020).

81. Grassegger H., Krogerus M. Ich habe nur gezeigt, dass es die Bombe gibt. Das Magazin. URL: <https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/> (дата звернення: 13.12.2020).

82. Haleem A., Javaid M., Khan I. H. Significant Applications of Big Data in COVID-19 Pandemic. *Indian Journal of Orthopaedics*. 2020. Vol. 54, No. 4. P. 526–528. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7204193/> (дата звернення: 13.12.2020).

83. Half-life of knowledge. Wikipedia. URL: [https://en.wikipedia.org/wiki/Half-life\\_of\\_knowledge](https://en.wikipedia.org/wiki/Half-life_of_knowledge) (дата звернення: 26.03.2020).

84. IMDB (In-memory Database). Национальная библиотека им. Н. Э. Баумана. URL: [https://ru.bmstu.wiki/IMDB\\_\(In-memory\\_Database\)](https://ru.bmstu.wiki/IMDB_(In-memory_Database)) (дата звернення: 13.12.2020).

85. India Lights. Nightlights.io. URL: <http://india.nightlights.io/#/nation/2006/12> (дата звернення: 13.12.2020).

86. Kennedy J. B. An interview with Nikola Tesla by John B. Kennedy. Colliers. URL: <http://www.tfcbooks.com/tesla/1926-01-30.htm> (дата звернення: 14.12.2020).

87. Kumawat D. C. Apache Storm architecture: Real-time Big data analysis engine for streaming data. Medium. URL: <https://medium.com/analytics-steps/apache-storm-architecture-real-time-big-data-analysis->

engine-for-streaming-data-4fc34ce0adae (дата звернення: 13.12.2020).

88. Lantz M. Why the Future of Data Storage is (Still) Magnetic Tape - IEEE Spectrum. IEEE Spectrum. URL: <https://spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape> (дата звернення: 26.03.2020).

89. Lucarini Francesca. GDPR vs CCPA: What are the main differences? Advisera. URL: <https://advisera.com/eugdpracademy/blog/2020/04/13/gdpr-vs-ccpa-what-are-the-main-differences/> (дата звернення: 02.12.2020).

90. Manyika J., Chui M., Brow B. Big data: The next frontier for innovation, competition, and productivity. McKinsey & Company. URL: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation> (дата звернення: 13.12.2020).

91. Machine learning. Cambridge University Press. URL: <https://dictionary.cambridge.org/ru/словарь/английский/machine-learning> (дата звернення: 13.12.2020).

92. Mavragani A., Tsagarakis K. P. Predicting referendum results in the Big Data Era. *Journal of Big Data*. 2019. Vol. 6, No. 1. P. 3. URL: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0166-z>.

93. Microsoft revives MAID with Pelican, but tape can still sleep easy - StorageBuzz. URL: <https://www.computerweekly.com/blog/StorageBuzz/Microsoft-revives-MAID-with-Pelican-but-tape-can-still-sleep-easy> (дата звернення: 26.03.2020).

94. Morris R. J. T., Truskowski B. J. The evolution of storage systems. *IBM Systems Journal*. 2003. Vol. 42, No. 2. P. 205–217. URL: <http://ieeexplore.ieee.org/document/5386860/> (дата звернення: 14.12.2020).

95. Open Data Challenge 2020. URL: <https://odc.in.ua/> (дата звернення: 29.03.2020).

96. O'Reilly T. What Is Web 2.0. O'Reilly Media. URL: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> (дата звернення: 14.12.2020).

97. Osborne S. China could ban children from “untrustworthy” families from schools under social credit system. The Independent. URL: <https://www.independent.co.uk/news/world/asia/china-social-credit-system-school-ban-family-travel-a8821371.html> (дата звернення: 13.12.2020).

98. Pazvakavambwa R. Predictive analytics, big data transform sports. ITWeb. URL: <https://www.itweb.co.za/content/XnWJad7bZbe7bjO1> (дата звернення: 13.12.2020).

99. Peter M., Grance I. The NIST Definition of Cloud Computing. National Institute of Standards and Technology Special Publication. .

100. R: What is R? The R Foundation. URL: <https://www.r-project.org/about.html> (дата звернення: 13.12.2020).

101. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da. Brussels. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (дата звернення: 02.12.2020).

102. René R., Ling X. Discipline and Punish: The Birth of China’s Social-Credit System. The Nation. URL: <https://www.thenation.com/article/archive/china-social-credit-system/> (дата звернення: 13.12.2020).

103. Rosenberg M. Marc My Words: The Coming Knowledge Tsunami. Learning Solutions Magazine. URL: <https://learningsolutionsmag.com/articles/2468/marc-my-words-the-coming-knowledge-tsunami> (дата звернення: 26.03.2020).

104. Seagate IronWolf 16TB NAS Internal Hard Drive HDD — 3.5 Inch SATA 6GB/S 7200 RPM 256MB Cache for Raid Network Attached Storage (ST16000VN001): Computers & Accessories. Amazon. URL: <https://www.amazon.com/Seagate-IronWolf-16TB-Internal-Drive/dp/B07SGGWYC1> (дата звернення: 26.03.2020).

105. Social Credit System. Wikipedia. URL: [https://en.wikipedia.org/wiki/Social\\_Credit\\_System](https://en.wikipedia.org/wiki/Social_Credit_System) (дата звернення: 13.12.2020).

106. Superparamagnetism. Wikipedia. URL: <https://en.wikipedia.org/wiki/Superparamagnetism> (дата звернення: 26.03.2020).
107. Toler A. How to Uncover Corruption Using Open Source Research. Bellingcat. URL: <https://www.bellingcat.com/resources/2016/09/05/how-to-uncover-corruption-using-open-source-research/> (дата звернення: 13.12.2020).
108. Tunggal A. T. The 50 Biggest Data Breaches . UpGuard. URL: <https://www.upguard.com/blog/biggest-data-breaches> (дата звернення: 13.12.2020).
109. Tyagi N. Top 10 Big Data Technologies. Analytics Steps. URL: <https://www.analyticssteps.com/blogs/top-10-big-data-technologies-2020> (дата звернення: 13.12.2020).
110. Wang C. J., Ng C. Y., Brook R. H. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA - Journal of the American Medical Association*. 2020. Vol. 323, No. 14. P. 1341–1342. URL: <https://jamanetwork.com/journals/jama/article-abstract/2762689> (дата звернення: 13.12.2020).
111. What is NoSQL? NoSQL Databases Explained. MongoDB. URL: <https://www.mongodb.com/nosql-explained> (дата звернення: 13.12.2020).
112. What is a data lake? Amazon Web Services. URL: <https://aws.amazon.com/ru/big-data/datalakes-and-analytics/what-is-a-data-lake/> (дата звернення: 13.12.2020).
113. Wihbey J. Predicting U.S. elections through search volume in Google Trends. Journalist's Resource. URL: <https://journalistsresource.org/studies/politics/elections/elections-search-volume/> (дата звернення: 13.12.2020).
114. Windermere Real Estate. Wikipedia. URL: [https://en.wikipedia.org/wiki/Windermere\\_Real\\_Estate](https://en.wikipedia.org/wiki/Windermere_Real_Estate) (дата звернення: 13.12.2020).
115. Wingfield N. Predicting Commutes More Accurately for Would-Be Home Buyers. The New York Times. URL: <https://bits.blogs.nytimes.com/2013/03/12/predicting-commutes-more-accurately-for-would-be-home-buyers> (дата звернення: 13.12.2020).

116. World Bank. Data big in action for government. World Bank. Dubai. URL: [www.mckinsey.com/business-functions/digital-mckinsey/our-insights/](http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/). (дата звернення: 09.12.2020)

117. Zhang Y. How data analytics helped Obama win the 2012 US presidential election. South China Morning Post. URL: <https://www.scmp.com/yp/learn/college-uni-life/university-programmes/article/3071524/how-data-analytics-helped-obama-win> (дата звернення: 13.12.2020).