

Горбачук В.М.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

Лупей М.І.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

Ніколенко Д.І.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

Дунаєвський М.С.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

Сулейманов С.-Б.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

Батіг Л.О.

Інститут кібернетики імені В.М. Глушкова Національної академії наук України

ПРОЦЕСИ ГЕНЕРАЦІЇ ЦИФРОВИХ ДАНИХ ТА ЇХ ЗАСТОСУВАННЯ

Стаття присвячена процесам датафікації для перетворення широкого кола явищ в цифровізовані кількісно вимірювані одиниці інформації для їх аналізу. У зв'язку з масовим проникненням інтелектуальних технологій у повсякденне життя та підвищенням фільтрації соціальної взаємодії через платформи соціальних медіа й інші онлайн-сервіси, дані сьогодні генеруються і збираються від різноманітного спектру практик. Визначено, що датафікація та обчислювальні соціальні науки можуть запропонувати новітні уявлення про життя з збудованими цифровими технологіями. Водночас багато вчених із соціальних наук стверджують, що процес датафікації звужуватиме складність об'єкта дослідження і багатство соціального контексту, в якому відбувається власне датафікація. Датафікація стосується процесу, за допомогою якого суб'єкти, об'єкти і практики перетворюються в цифрові дані. У зв'язку з розвитком цифрових технологій, цифровізацією та великими даними багато фахівців стверджують, що датафікація посилюється з розгортанням нових вимірів соціального життя в цифрових просторах. Датафікація відтворює різного роду інформацію у кількісно вимірювані та зчитувані машиною дані з метою агрегації й аналізу. Датафікація також використовується як термін для опису логіки, за якою речі у світі – це джерела даних, що добуватимуться для уточнення чи продажу, за допомогою яких можна діставати уявлення про людську поведінку та соціальні явища. У статті розкрито, що цей термін часто використовується тими, хто прагне критикувати подібну логіку і відповідні процеси перетворення інформації. З'ясовано, що поняття датафікації спочатку застосовувалося вченими, які прагнули перевіряти, як цифровий світ змінюється з розвитком великих даних та економік даних. Коли власне датафікація набуває поширення, науковий підхід у ряді дисциплін і піддисциплін потребує відповідної концептуалізації, щоб краще розуміти зрушення до відтворення інформації як даних для аналізу цифрових (віртуальних) образів реальних об'єктів, виходячи за межі онлайн-платформ. Науковці даних прагнуть виробляти висновки з широкого кола питань, досліджуючи сліди даних, які люди залишають після себе в кіберпросторі.

Ключові слова: цифрові платформи, соціальні мережі, інфраструктури, конфіденційна інформація, вимірювані індикатори, датафікація.

Постановка проблеми. Оскільки все більше аспектів соціального життя почали генерувати цифрові дані, природа орієнтації цифрових інфраструктур істотно змінилися через можливості проведення аналізу цих даних для вироблення прибуткових опцій. Наприклад, існуючі спромож-

ність аналізу великих (наборів) даних і здатність поєднувати уявлення з множинних наборів даних (через дані пошукових систем (search engine), соціальні медіа, демографічну інформацію, історію переглядів (viewing history) на YouTube тощо) суттєво змінили роботу онлайн-платформ.

Аналітика великих даних використовується у приватному секторі для різноманітних цілей, включаючи фільтрацію цифрового контенту чи продукцію на ринкових онлайн-майданчиках у формах рекомендацій та, найголовніше, цільових рекламних оголошень. Крім питань цифровізації, виникають нові проблеми так званої датафікації, яка вважалася можливістю діставати безпрецедентний доступ до даних для соціальних досліджень.

Аналіз останніх досліджень і публікацій.

З розвитком аналітики даних з'явилася сучасна концепція датафікації, яка описує і популяризує новий підхід до соціальних досліджень [1]. Хоча датафікація відрізняється від цифровізації, остання часто є частиною процесу датафікації [2]. Аналогічно кількісно вимірювані об'єкти датафікуються, коли зводяться до елементів інформації, які можна обліковувати, агрегувати, обчислювати і відтворювати у зчитувані машиною елементи. Тоді процес датафікації не бере до уваги якісні характеристики об'єктів, які не вимірюються кількісно, втрачаючи можливі значні складні залежності в логіці великих даних [3–5]. Крім того, з'являються методологічні та епістемологічні питання до уявлень, виведених з економік нових даних, в яких існує нерівність, а також до великої вартості, якої можна досягати з цифровою медіацією (digital mediation) шляхом заохочення участі в соціальних взаємодіях і практиках обміну особистою інформацією онлайн [6–8].

Постановка завдання. З 2010-х років більша увага приділяється до соціологічних уявлень, висунутих великими наборами даних, що іноді називають даталогічним чи обчислювальним поворотом. У соціальних науках теж виникло поживання навколо використання великих даних (особливо даних соціальних медіа) для проведення аналізу культури і з'ясування соціальних проблем. Датафікація стала поширеною ідеологією, відомою як датаїзм (dataism), де неявно припускаються цінності та уявлення агрегованих даних [5]. Ця ідеологія також надає значної довіри і легітимності інституціям, які збирають ці дані, незважаючи на фінансові інтереси таких інституцій та відповідні викривлення інформації. Справді, така ідеологія чітко простежується в заявах, зроблених на початку революції великих даних про експоненціальну спроможність великих даних пояснювати соціальне життя, коли деякі прихильники аналізу великих даних оголошували про завершення епохи соціальної теорії. Ці заяви були зроблені на основі того, що міркування про причини певних дій людей, тобто про

основоположні питання соціального наукового пошуку (inquiry), визнавалися недоречними при здатності великих даних виявляти моделі дій людей у масових масштабах. По суті, датафікація вважалася способом уникати зайвих складнощій суспільного життя та виявлення кореляцій без потреби змістовних пояснень. Завданням є перевірка подібних тверджень, які видозмінювалися після спостережень реальних прогностичних спроможностей великих даних.

Виклад основного матеріалу дослідження.

Дані, що агрегуються науковцями даних (data scientists), в основному, з'являються завдяки слідам даних, які генерують онлайн-взаємодії. Ці дані можна збирати й аналізувати шляхом датафікації соціального життя. Оскільки соціальні медіа-платформи перетворилися на хостинги більшості щоденних онлайн-взаємодій між людьми, то ці взаємодії стали пакетами (parcels) даних у формі лайків (likes), поширень (shares), кліків (clicks). Крім цих взаємодій, цифрове життя людей також включає низку діяльностей зі збору даних: перегляд веб-сторінок (web browsing), використання пошукових систем, потокове передавання (streaming) цифрового контенту, онлайн-купівля, взаємодії з рекламними оголошеннями та інші цифрові практики відображаються як фрагменти (pieces) даних, які можна сортувати для ідентифікації та аналізу трендів, а також, напевне, для створення подальших вирашних можливостей шляхом комерціалізації [9]. Не здійснюючи подібної діяльності, людина надає докладні геолокаційні дані від кожного свого цифрового пристрою чи кожної своєї цифрової мітки, які можна збирати й аналізувати. Така датафікація соціального життя суттєво змінила організацію цифрових платформ, спрямовану на пошук вирашних можливостей шляхом збору даних, а нагляд за онлайн-даними і метаданими (data surveillance, dataveillance) вбудовується у майже всі аспекти цифрової взаємодії.

Датафікація й аналітика даних поширюються від онлайн-взаємодії та використання соціальних медіа на інші сфери. Вчені виявили датафікацію здоров'я як у трендах технологій індивідуального самоконтролю (self-tracking) (фітнес-трекери, смарт-годинники тощо), так і способах, якими клінічна практика стає все більше керованою даними (data-driven), особливо способами, якими уряди працюють з медичною інформацією [10]. Датафікація теж вплинула на освіту, оскільки дітей все більше контролюють у школах шляхом радіочастотної ідентифікації (radio-frequency identification, RFID) шкільної форми, відео-

спостереження (замкненого телебачення; системи закритої трансляції телебачення; closed-circuit television, CCTV) з підтримкою розпізнавання облич (facial recognition-enabled) та онлайн-моніторингу класної роботи [11].

RFID використовує електромагнітні поля для автоматичної ідентифікації та відстеження міток (tags), прикріплених (attached) до об'єктів. Система RFID складається з мініатюрного радіотранспондера (RFID-мітки), радіоприймача (radio receiver) та передавача (transmitter). Коли спрацював (triggered) електромагнітний імпульс запиту (interrogation pulse) від розташованого поблизу пристрою зчитування (reader device; рідер; зчитувач; інтеррогатор) RFID, RFID-мітка передає радіосигналом цифрові дані, зазвичай ідентифікаційний інвентарний номер (identifying inventory number) назад до рідера (цей номер можна використовувати, наприклад, для відстеження товарних запасів у ланцюгах постачання). Пасивні мітки живляться (powered) енергією від запитувальних (interrogating) радіохвиль рідера RFID. Активні мітки живляться від батареї, а тому можуть зчитуватися на порівняно більшій відстані від рідера RFID (до сотень метрів). На відміну від штрих-коду (barcode), тег не обов'язково має знаходитися в межах прямої видимості (line of sight) зчитувача, а тому може вбудовуватися у відстежуваний об'єкт (tracked object). RFID є одним з методів автоматичної ідентифікації та збору даних (automatic identification and data capture, AIDC).

CCTV, відоме як відеоспостереження (video surveillance), – це система передавання інформації, яка полягає у використанні відеокамер або телевізійних камер для передачі сигналу у певне місце на обмежений набір моніторів та/або пристроїв запису. CCTV відрізняється від телевізійного мовлення (broadcast television) тим, що сигнал не передається у відкритому режимі, хоча може застосовуватися з'єднання (links) точка-точка (point-to-point, P2P), точка-багато точок (point-to-multipoint, P2MP), сітчасті дротові (mesh wired) чи бездротові з'єднання. Сітчаста мережа (mesh network, meshnet) – це топологія локальної мережі, в якій кожний вузол інфраструктури (мости, комутатори (switches), інші пристрої інфраструктури) з'єднується безпосередньо, динамічно та неієрархічно з якнайбільшою кількістю інших вузлів і кооперується з ними для ефективної маршрутизації даних до клієнтів і від клієнтів. Незважаючи на те, що всі відеокамери входять до CCTV, відеоспостереження найчастіше застосовується до нагляду в зонах, які потребу-

ють додаткової безпеки чи постійного (ongoing) моніторингу, чого зазвичай не потребують відеокамери відеотелефонії. Такими зонами є банки, банкомати, казино, вокзали, аеропорти, військові об'єкти, кордони, крамниці тощо.

Галузь промислового Інтернету речей (Industrial Internet-of-Things, IIoT) корисна для отримання джерел даних, а сфера машинного навчання – для обробки отриманих даних. Ланцюги постачання даних стають об'єктами блокчейн-застосунків.

По-перше, аналітика периферійних даних здійснюється на рівні інструментів з бездротовою передачею лише важливих доступних даних (і затримкою решти даних). Алгоритми аналітики даних можуть завантажуватися на програмовану користувачем вентиляну матрицю (ПКВМ; field-programmable gate array, FPGA) в інструменті, забезпечуючи також прогнозне технічне обслуговування. FPGA – це інтегральна схема, призначена для конфігурування на робочому полі (field-programming) клієнтом або розробником після її виготовлення. Дана конфігурація FPGA зазвичай вказується, використовуючи мову опису апаратного забезпечення (АЗ) (hardware description language, HDL), подібну до мови, що використовується для інтегральної схеми конкретного застосування (application-specific integrated circuit, ASIC). Якщо раніше для задання конфігурації використовували електросхеми (circuit diagrams), то зараз для цього зазвичай використовують електронні засоби автоматизації проектування. ПКВМ містять масив програмованих логічних блоків та ієрархію з'єднань (interconnects), які можна реконфігурувати, щоб підключати (wire) блоки разом. Логічні блоки можна реконфігурувати для виконання складних комбінаційних функцій чи роботи як простих логічних вентилів (наприклад, AND та XOR). У більшості ПКВМ логічні блоки також включають елементи пам'яті, які можуть бути простими тригерами (flip-flops) чи досконалішими блоками пам'яті. Багато ПКВМ можна перепрограмувати для втілення різних логічних функцій, що дозволяє гнучко реконфігурувати обчислення при їх виконанні комп'ютерним програмним забезпеченням (ПЗ). ПКВМ є важливими для розвитку вбудованих систем завдяки своїй спроможності розпочинати розробку ПЗ одночасно з розробкою АЗ, уможлилювати моделювання продуктивності системи (system performance) на початку її розробки, допускати різні випробування (trials) системи та проектні версії (design iterations) до завершення процесу вибору системної архітектури.

По-друге, аналітика на зашифрованих (encrypted) даних передбачає кіберзахист з шифруванням усієї інформації, але зашифровані дані теж можуть бути матеріалом роботи для оператора.

По-третє, технологія блокчейну разом з IoT дозволяє досягати консенсусу інструментів у випадках, коли деякі з них несправні, свідомо (deliberately wrong) (скажімо, внаслідок кібератак на деякі вузли) чи несвідомо (faulty). Тому доступ до пристроїв видається лише дозволеним особам (permissioned entities), як опубліковано на блокчейні.

Розуміння Інтернету речей (Internet of Things, IoT) та зв'язку 5G зсуває парадигму мобільних обчислень від централізованих мобільних хмарних обчислень (Mobile Cloud Computing, MCC) до мобільних периферійних обчислень (Mobile Edge Computing, MEC). Головна особливість MEC – переміщення мобільних обчислень, управління мережею та зберігання даних до країв (edges) мережі (наприклад, до базових станцій і точок доступу), щоб забезпечувати на мобільних пристроях (з обмеженими ресурсами) роботу застосунків, чутливих до затримок (latency-critical) або потребуючих багато обчислень. Очікується, що перехід до парадигми MEC дозволить істотно зменшувати затримки та енергоспоживання мобільними пристроями, відповідаючи на ключові виклики для реалізації бачення 5G. Ці сподівання мотивували зусилля науковців і промисловців до подальшої розробки технології MEC. Основним напрямком досліджень MEC є поступове злиття дисципліни бездротового зв'язку та мобільних обчислень для широкого спектру нових проєктів від методів розвантаження обчислень (computation offloading) до мережевих архітектур. Тому заслуговує уваги сучасний стан досліджень MEC, зокрема, менеджмент спільних обчислювальних ресурсів і радіоресурсів, розгортання системи MEC, MEC з підключеним кешем (cache-enabled), менеджмент мобільності для MEC, зелені (екологічні) MEC, MEC з урахуванням приватності (privacy-aware), стандартизація MEC, типові сценарії застосування MEC.

Нової комп'ютерною парадигмою стали хмарні обчислення (ХО) з централізацією обчислень, зберігання даних, мережевого менеджменту у хмарах – центрах обробки даних, магістральних (backbone) IP-мережах, стільникових опорних (core) мережах. Ці наявні у хмарах ресурси можна використовувати для надання еластичних обчислювальної потужності та можливостей зберігання даних, щоб підтримувати пристрої кінцевих користувачів з обмеженими ресурсами. ХО стали

рушійною силою швидкого зростання багатьох Інтернет-компаній. Наприклад, хмарні послуги стали найприбутковішим сектором для Amazon (заснована у 1994 р.; AMZN у лістингу біржі NASDAQ), сприявши успіху Dropbox (заснована у 2007 р.; DBX у лістингу біржі NASDAQ).

Водночас спостерігається новий тренд у комп'ютерних технологіях, коли функції хмар частіше зміщуються до країв (периферії) мережі. Найближчим часом очікується розгортання десятків мільярдів крайових (периферійних) пристроїв, швидкості процесорів яких зростають експоненціально за законом Мура. Парадигма MEC полягає у тому, що збір (harvesting) великої кількості незайнятих (idle) обчислювальних потужностей та місць для зберігання даних, розподілених на краях мережі, може давати достатні спроможності для здійснення на мобільних пристроях завдань, чутливих до затримок і потребуючих багато обчислень. Позаяк тривалі затримки поширення сигналів залишаються ключовим недоліком ХО, MEC з ближчим доступом до користувачів широко вважаються ключовою технологією для втілення різних бачень Інтернету наступного покоління – тактильного Інтернету (з часом реакції у мілісекундах), IoT, соціальних мереж, Інтернету себе (Internet of Me, IoM), який постачає в Інтернет дані про кожний окремих людський організм. У 2010-х роках популярність мобільних пристроїв та експоненціальне зростання мобільного Інтернет-трафіку впливали на значне просування у галузях бездротового зв'язку і створення мереж: досягнення у дрібностільникових (small-cell) мережах, багатоантенний зв'язок і зв'язок на міліметрових хвилях можуть забезпечувати користувачам гігабітний бездротовий доступ у системах наступного покоління. Високошвидкісний та високонадійний бездротовий (air) інтерфейс дозволяє запускати обчислювальні послуги мобільних пристроїв у віддаленому хмарному центрі обробки даних (ЦОД), створюючи MCC. Це передбачає значну відстань для поширення сигналу від кінцевого користувача до ЦОД, а також помітну затримку для мобільних застосунків. Оскільки MCC не підходять для широкого кола новопосталих мобільних застосунків, чутливих до затримок, то розробляються нові мережеві архітектури для кращої інтеграції ХО у мобільні мережі.

У бездротових системах 5G розгортатимуться надщільні периферійні пристрої, включаючи малостільникові (small-cell) базові станції (base stations, BSs) з малими робочими зонами, бездротові точки доступу (access points, APs), ноутбуки,

планшети, смартфони, причому обчислювальна потужність кожного пристрою буде порівнянною з потужністю сервера, виробленого десятиліттям раніше. Отже, велика кількість пристроїв буде незайнятою в кожний момент часу, що дозволить збирати великі обчислювальні ресурси і ресурси зберігання даних, наявні на периферії мережі. Цих ресурсів буде достатньо для забезпечення повсюдних (ubiquitous) мобільних обчислень.

Основною метою бездротових систем від 1G до 4G є досягнення дедалі вищих швидкостей бездротового зв'язку для підтримки переходу від голосоцентричного (voice-centric) трафіку до трафіку, спрямованого на мультимедійні технології (multimedia-centric). Оскільки швидкості бездротового зв'язку наближаються до швидкостей аналогічної дротової мережі (wireline), то метою 5G є підтримка вибухового розвитку інформаційно-комунікаційних технологій (ІКТ) та Інтернету загалом: системи 5G підтримуватимуть комунікації, обчислення, управління (communications, computing, control), надання контенту (content), або набір функцій 4С. Для 5G постає широкий спектр нових застосунків і послуг – онлайн-ігри в режимі реального часу, віртуальна реальність, потокове відео (video streaming) ультрависокої роздільної здатності, які вимагають надзвичайно високої швидкості доступу й низької затримки. Хоча виявилися дещо завищеними прогнози компанії Cisco (заснована у 1984 р. з назвою, що походить від міста San Francisco; CSCO у лістингу біржі NASDAQ) про підключення до Інтернету близько 50 млрд. гаджетів IoT (датчиків, носних (wearable) пристроїв) на 2020 р., більшість таких гаджетів має обмежені ресурси для обчислень, зв'язку, зберігання даних, а тому потребуватиме хмарних або периферійних пристроїв для підвищення їхніх спроможностей. Втілення амбіційних планів мілісекундних затримок для обчислень і комунікацій в 5G потребуватиме не лише ХО. Крім того, обмін даними між кінцевими користувачами та віддаленими хмарами збільшуватиме обсяги передачі даних і зменшуватиме ефективність транспортних мереж зв'язку (backhaul networks). Тому ХО слід доповнювати системою МЕС, яка спрямовує трафік, обчислювальні та мережеві функції на краї мережі. Це відповідає ключовій характеристиці мереж наступного покоління про те, що інформація все більше генерується і споживається локально внаслідок розвитку застосунків IoT, соціальних мереж і доставки контенту.

Поняття МЕС вперше запропонував у 2014 р. Європейський інститут телекомунікаційних стан-

дартів (European Telecommunications Standard Institute, ETSI) як нову платформу, що надає спроможності ІКТ і ХО у мережі радіодоступу (Radio Access Network, RAN) в безпосередній близькості до абонентів мобільного зв'язку.

Вчені також звернули увагу на форми нагляду за даними, що впливають на догляд за дітьми поза освітніми закладами, – застосунки для батьків (parenting apps) і технології відстеження дітей (child-tracking). Розповсюдження датафікації вказує на роль поширеної ідеології датафікації [5], за якою об'єктивну істину можна знаходити шляхом відображення будь-якої соціальної проблеми як цифрових даних для обчислювального аналізу.

Логіка датафікації суттєво критикувалася науковцями соціальних наук. Фахівці з конфіденційності (privacy) та стеження висвітлювали широко поширені питання навколо способів, якими датафікація пасивно сприяє збору персональної інформації і про які користувачі платформи можуть не знати, коли дані зберігаються для широкого кола застосувань у майбутньому, формально неузгоджених з користувачами. Оскільки датафікація поширюється на все більше сфер соціального життя, поняття згоди (consent) стає менш важливим, бо користувачі цифрових платформ і сервісів, пов'язаних з даними (datafied services), часто відчують свою неспроможність відхилити цифрові пропозиції. Крім того, масштабні витіки (leaks) даних і хакерські атаки (hacks) на платформи соціальних медіа демонструють крихкість систем захисту даних високих стандартів.

Крім створення проблем (concerns) приватності, датафікація може відтворювати і загострювати існуючі соціальні нерівності. Керовані великими даними системи оцінки ризиків, які тепер регулярно вживаються провайдерами фінансових послуг і страховими компаніями, можуть увічнювати (perpetuate) дискримінацію щодо маргіналізованих спільнот [12]. Така дискримінація маскується міфологією об'єктивності, прозорливості (insight), точності навколо цих систем, незважаючи на їх часто неясні практичні дії (opaque workings). Хоча різні форми дискримінації існували завжди, а датафікація й згадані системи не є першопричинами дискримінації, ці системи сприяють такій дискримінації, яку складно фіксувати і яка небезпечно легітимізує соціальні нерівності як підтвержене даними природне явище, а не продукт неявного суб'єктивного упередження, притаманного людині [13–15].

Вчені також висловлювали занепокоєння з приводу датафікації соціальних наук і шляхів, якими

обчислювальні методи впливали на соціальні дослідження. Комп'ютерним соціальним наукам приписують той недолік, що вони представляють великі дані, які кажуть самі за себе (*speaking for itself*) та своєрідно захоплюють соціальні відносини: великі дані виникають з нових форм цифрової соціальності та розгортаються комерційними силами. Наприклад, використання даних соціальних медіа як спосіб калібрувати (*gauge*) громадську думку часто неналежним чином видає такі дані за репрезентативні для суспільства в цілому, нехтуючи важливими відмінностями між демографічними характеристиками різних платформ і рівнями специфічної доступності (*affordance*) цифрових просторів, що може породжувати відмінності відгуків. Аналогічно ці великі дані можуть встановлювати кореляції між різнорідними (*disparate*) змінними, які виявляються оманливими у процесі доведення причинно-наслідкового зв'язку [4]. Можлива омана не означає, що вченим треба нехтувати цими даними: треба дотримуватися обережності в інтерпретації цих даних для гарантування того, що ці дані належним чином контекстуалізовані з теоретично обґрунтованим соціальним науковим аналізом. Якісні відмінності слід ретельно вивчати, а не обходити.

Одним з прикладів проблем аналізу даних у соціальному середовищі є застосування економіко-математичних методів у колишньому СРСР [16]. Леонід Канторович (народився у м.Санкт-Петербург Російської імперії у 1912 р., здобув наукові ступені магістра математики Ленінградського державного університету (ЛДУ) (заснованого у 1819 р.) у 1930 р. та доктора фізико-математичних наук (без захисту дисертації) у 1935 р., професор ЛДУ імені А.С.Бубнова з 1934 р., лауреат Нобелівської премії з економіки 1975 р., помер у м.Москва СРСР у 1986 р.) у 1930-х роках під час індустріалізації СРСР працював у Ленінградському інституті інженерів залізничного транспорту (заснованому у 1809 р.) і прагнув застосовувати математичні дослідження на практиці. У 1938 р. Канторович консультував фанерний трест з проблеми ефективного використання луцильних верстатів і формулював проблему як задачу максимізації лінійної функції багатьох змінних при великій кількості обмежень у формі лінійних рівностей і нерівностей. Для розв'язання цієї задачі Канторович запропонував модифікований метод невизначених множників Лагранжа (1736–1813) і опублікував книгу [17], де описав задачі математичної економіки, які можна чисельно розв'язувати згаданим методом, створивши засади лінійного програму-

вання. Коли у 1939 р. був заснований Військовий інженерно-технічний університет (ВІТУ), Канторович став завідувачем кафедри математики ВІТУ. Під час Другої світової війни у 1942 р. ВІТУ був евакуйований до м.Ярославль, а Канторович став звертатися зі своїми пропозиціями до Держплану (1923–1991) СРСР. У 1943 р. доповідь Канторовича була обговорена на нараді у голови Держплану СРСР, але економіко-математичні методи Канторовича були відхилені як такі, що запозичують положення так званих буржуазних економічних теорій і не відповідають трудовій теорії вартості Маркса (1818–1883). З 1948 р. підполковник Канторович очолював відділ в Інституті математики і механіки ЛДУ імені А.О.Жданова, який за розпорядженням Й.В.Сталіна був підключений до розрахунків, пов'язаних з розробкою ядерної зброї.

Аналогічно до оптимального розпилу фанерного листа, методи Канторовича були запропоновані до розрізання сталевих листів. У 1949 р. впровадження цих методів оптимізації на Ленінградському вагонобудівному заводі імені І.Є.Сторова дало такі наслідки: за теоретичними розрахунками, при даному обсягу сировини, на одному з видів продукції випуск мав збільшитися на 4%; цей розрахований показник був включений у план; на практиці продуктивність зросла на 2,5%, що означало зрив плану; завод також не виконав план з металобрухту, значна частина якого складалася з відходів сталевих листів; керівництво заводу отримало догану, а його працівники залишилися без премій. Однак після скарги в партійні органи СРСР працівники отримали премію як виняток, а інноваційний метод крою металу був завпроваджений у широку практику.

У конкурентному соціальному середовищі експертні брокери даних збирають інформацію про підписку на медіа (наприклад, газети та журнали), роздрібних продавців замовлень поштою (*mail-order*), опитування (*polls*), огляди (*surveys*), туристичні агенції, симпозиуми, конкурси, реєстрацію продуктів і гарантії, компанії з обробки платежів (*payment handling*), державні записи (*government records*) та інше [18]. З появою Інтернету галузь брокерів даних швидко розширювалася завдяки диверсифікації методів збору (*capture*) та консолідації даних, пропонуючи розмаїття продуктів і послуг [19]. Крім того, у повсякденному житті люди беруть участь у різноманітних онлайн-діяльностях, які розкривають їхню персональну інформацію. Такі онлайн-діяльності включають використання мобільних застосунків, купівлю житла чи автомобіля, підписку на видання, проведення

транзакцій кредитною карткою в універмагах або через онлайн-каталог, участь в опитуваннях, веб-серфінг (surfing the web), спілкування (chatting) з друзями на платформі соціальних медіа, участь у розіграшах призів (sweepstakes), підписку на новинні веб-сайти. Такі щоденні діяльності генерують різноманітні інформації про цих людей, яка, в свою чергу, доставляється чи продається брокерам даним [20].

Брокери даних агрегують дані з різноманітних джерел. На додаток до існуючих відкритих джерел даних, ці брокери також купують або орендують дані про окремих осіб у сторонніх (third-party) компаній. Зібрані дані можуть містити дії веб-перегляду (web browsing), інформацію про банкрутство, гарантію реєстрації, інформацію про голосування, дані про покупки споживачів, інформацію про іншу щоденну діяльність веб-взаємодії. Як правило, брокери даних не дістають дані безпосередньо від окремих осіб, а тому більшість людей не знають, що їхні дані збираються і споживаються брокерами даних. Отже, можливо, що подробиці життя окремих людей конструюватимуться та упакуватимуться в якості кінцевого продукту шляхом обробки й аналізу компонентів даних, які постачаються від різних джерел брокерів даних [21].

Брокери даних придбають і зберігають дані окремих осіб як продукти в інфраструктурі конфіденційних даних, яка зберігає, поширює і споживає дані через мережеві технології [22]. Ці дані здаватимуться в оренду чи продаватимуться з метою отримання прибутку. Продукти даних містять списки перспективних осіб, які задовольняють певним умовам, включаючи такі деталі, як імена, номери телефонів, адреси, адреси електронної пошти, а також такі елементи даних, як вік, стать, дохід, наявність дітей, етнічна приналежність, кредитний статус, володіння кредитною карткою, вартість житла, хобі, купівельна звичка та сімейний стан (background status). Ці колекції похідних продуктів даних, де брокери даних мають додану вартість через методи аналізу даних та інтеграції даних, використовуються для цільового маркетингу і рекламних промоцій, соціальної класифікації осіб, оцінювання кредитного рейтингу та послуг відстеження (tracing) [18].

Інтеграція та перепродаж даних, супроводжувані відповідними послугами з доданою вартістю, – це велика прибуткова галузь [23]. Ця галузь щогодини торгує великими обсягами даних і похідної інформації на ряді ринків, що спеціалізуються на фінансах, роздрібній торгівлі, логістиці, туризмі, нерухомості,

охороні здоров'я, політичному голосуванні, бізнес-аналітиці (business intelligence), приватній безпеці тощо. Ці дані охоплюють всі аспекти повсякденного життя і включають громадське адміністрування (public administration), споживання товарів і медіа, відпочинок, комунікації, подорожі, злочини та взаємодії в соціальних медіа [19].

Наприклад, Facebook використовує мережі користувачів, завантажений користувачами контент і профілі мільйонів своїх активних користувачів. Зібрані дані (коментарі, відео, фотографії та лайки (likes; вподобання) користувачів) застосовуються для формування набору таких рекламних продуктів, як Lookalike Audiences, Partner Categories, Managed Custom Audiences. Facebook у 2003 р. заснував як Facemash Марк Цукерберг (народився у м.Нью-Йорк у 1984 р., у середній школі збудував програму, що дозволила всім комп'ютерам між його будинком і стоматологічним кабінетом (dental office) його батька комунікувати один з одним, розробляв музичний плеєр Synapse Media Player з елементами машинного навчання про звички користувача (згаданий у квітні 2003 р. на веб-сайті Slashdot (заснували у 1997 р. Роб Малда (народився у 1976 р., здобув науковий ступінь бакалавра коледжу Хоуп (заснованого у 1865 р.) у м.Холланд штату Мічиган) та його однокурсник Джеф Бейтс (здобув наукові ступені бакалавра з історії та біології коледжу Хоуп у 1998 р.) та оцінений PC Magazine (заснували у 1982 р. Девід Баннелл (народився і помер у м.Берклі штату Каліфорнія у 1947 р. і 2016 р. відповідно, здобув науковий ступінь бакалавра з історії Університету Небраска (заснованого у 1869 р.) у 1969 р.), Джим Едлін (здобув науковий ступінь бакалавра Університету штату Мічиган (заснованого у 1855 р. і де навчався один з авторів даної роботи) з відзнакою), Шеріл Вудард). Сьогодні Facebook працює на 112 мовах світу і належить Meta Platforms (META у лістингу біржі NASDAQ).

Висновки. Продаж даних брокерам став головним потоком доходів для багатьох компаній. Наприклад, роздрібні торговці регулярно продають дані про такі транзакції клієнтів, які деталі кредитної картки, інформація про покупки клієнтів, програми лояльності клієнтів, менеджмент стосунків з клієнтами, інформація про підписку. Інтернет-магазини продають дані потоків кліків (clickstream) стосовно того, як людина переміщувалася (navigated) через веб-сайт і скільки часу витратила на різних сторінках. Аналогічно медіакомпанії (газети, радіо, телевізійні станції) збирають дані, що містяться в межах їхнього контенту (наприклад, новинні історії та рекламні

оголошення), а компанії соціальних медіа агрегують метадані та контенти своїх користувачів, щоб обробляти таку інформацію для побудови профілів осіб і вироблення своїх власних продуктів даних на продаж брокерам даних. Оскільки володіння циф-

ровими образами матеріальних і нематеріальних активів створює додаткові можливості маніпуляції, спекуляції та капіталізації, то процеси генерації та застосування цифрових даних потребують належного вивчення, висвітлення і регулювання.

Список літератури:

1. Southerton C. Datafication. *Encyclopedia of Big Data*. L.A.Schintler, C.L.McNeely (eds.) Cham, Switzerland: Springer Nature Switzerland, 2022. P. 358–361.
2. Горбачук В.М., Лупей М.І., Сулейманов С.-Б., Батіг Л.О., Симонов Д.І. До системної інженерії великих даних. *Збірник наукових праць університету кораблебудування імені адмірала Макарова*. 2022. № 2 (489). С. 66–75.
3. Boyd D., Crawford K. Critical questions for big data. *Information, Communication and Society*. 2012. 15 (5). P. 662–679.
4. Kitchin R. Big data, new epistemologies and paradigm shifts. *Big Data & Society*. 2014. 1 (1). P. 1–12.
5. van Dijck J. Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveillance & Society*. 2014. 12 (2). P. 197–208.
6. Birchall C. *Shareveillance: the dangers of openly sharing and covertly collecting data*. Minneapolis, MN: University of Minnesota Press, 2017. 86 p.
7. van Dijck J. *The culture of connectivity: a critical history of social media*. Oxford, UK: Oxford University Press, 2013. 240 p.
8. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology Impact*. 2015. 30 (1). P. 75–89.
9. Lupton D. *Data selves: more-than-human perspectives*. Cambridge, UK: Polity, 2019. 80 p.
10. Ruckenstein M., Schüll N.D. The Datafication of health. *Annual Review of Anthropology*. 2017. 46 (1). P. 261–278.
11. Taylor E. *Surveillance schools: security, discipline and control in contemporary education*. London, UK: Palgrave Pivot, 2013. 149 p.
12. Leurs K., Shepherd T. Datafication & discrimination. *The datafied society: studying culture through data*. M.T. Schäfer, K. van Es (eds.) Amsterdam, Netherlands: Amsterdam University Press, 2017. P. 211–231.
13. Горбачук В.М., Сирку А.А., Сулейманов С.-Б. Ірраціональність і раціональність у прийнятті рішень. *Економічний вісник Запорізької державної інженерної академії*. 2017. Вип. 2 (08). Ч. 2. С. 132–136
14. Горбачук В.М., Макаренко О.С. Нейрофізіологія прийняття рішень людиною. *VII Конгрес Українського товариства нейронаук*. Київ: Інститут фізіології ім. О.О.Богомольця НАН України, 2017. С. 82.
15. Горбачук В.М., Макаренко О.С. Особливості прийняття рішень людиною для розв'язання складних міждисциплінарних проблем. *Системні дослідження та інформаційні технології*. 2017. № 3. С. 73–87.
16. Великий А.П., Горбачук В.М., Єрмольєв Ю.М., Кнопов П.С. Взаємодія централізації і децентралізації у міждисциплінарній кібернетиці академіка Глушкова. *Сучасна інформатика: проблеми, досягнення та перспективи розвитку*. Київ: Інститут кібернетики імені В. М. Глушкова НАН України, 2013. С. 241–243.
17. Леонид Витальевич Канторович: человек и ученый. Т. 1. В.Л. Канторович, С.С.Кутателадзе, Я.И. Фет (ред.-сост.) Новосибирск: Ин-т математики им. С.Л.Соболева, Ин-т вычисл. математики и мат. геофизики СО РАН, 2002. 542 с.
18. *On the data trail: how detailed information about you gets into the hands of organizations with whom you have no relationship. A report on the Canadian data brokerage industry*. Ottawa, Ontario: Canadian Internet Policy and Public Interest Clinic (CIPPIC), 2006. 55 p.
19. Kitchin R. The data revolution: big data, open data, data infrastructures and their consequences. *Small data, data infrastructures and data brokers*. Revised edition. R.Kitchin (ed.). London, UK: Sage, 2014. P. 27–47.
20. Ramirez E., Brill J., Ohlhausen M.K., Wright J.D., McSweeney T. *Data brokers. A call for transparency and accountability*. Washington, DC: Federal Trade Commission, 2014. 110 p.
21. Anthes G. Data brokers are watching you. *Communications of the ACM*. 2015. 58 (1). P. 28–30.
22. Kitchin R., Lauriault T.P. Small data in the era of big data. *GeoJournal*. 2015. 80. P. 463–475.
23. Mayer-Schönberger V., Cukier K. *Big data: a revolution that will transform how we live, work, and think*. Boston, MA, New York, NY: An Eaman Dolan Book, Houghton Mifflin Harcourt Publishing Company, 2013. 158 p.

Gorbachuk V.M., Lupei M.I., Nikolenko D.I., Dunaievskyi M.S., Suleimanov S.-B., Batih L.O.
DIGITAL DATA GENERATION PROCESSES AND THEIR APPLICATION

The article is devoted to datafication processes to transform a wide range of phenomena into digitalized (digitized) quantifiable units of information for their analysis. Due to the massive infiltration of smart technologies into everyday life and increased filtering of social interaction through social media platforms and other online services, data is generated and collected from a diverse range of practices today. It is defined that datafication and computational social science can offer essential insights into life with embedded digital technologies. At the same time, many scientists from the social sciences claim that the datafication process would narrow the complexity of the research object and the richness of the social context in which the datafication itself takes place. Datafication refers to the process by which subjects, objects and practices are transformed into digital data. In connection with the development of digital technologies, digitalization and big data, many experts claim that datafication is increasing with the deployment of new dimensions of social life in digital spaces. Datafication renders various types of information into quantifiable and machine-readable data for the purpose of aggregation and analysis. Datafication is also used as a term to describe the logic that things in the world are data sources to be mined for refinement or sale, which can be used to gain insights into human behavior and social phenomena. The paper reveals this term is often used by those who seek to criticize such logic and the corresponding information transformation processes. It has been clarified the concept of datafication was originally applied by scholars seeking to examine how the digital world is changing with the development of big data and data economies. As datafication itself becomes more widespread, the scientific approach in a number of disciplines and subdisciplines needs a corresponding concept to understand the broader shifts towards rendering information as data for image analysis, not limited to online platforms. Data scientists seek to draw conclusions about a wide range of issues by examining the cyberspace data traces that people leave behind.

Key words: *digital platforms, social networks, infrastructures, confidential information, measurable indicators, datafication.*