

УДК 316.77:[004.738.5+004.8]

DOI <https://doi.org/10.32782/2707-9147.2023.100.5>

**Р. Я. ЛЮБЕНКО**

аспірант Докторської школи імені родини Юхименків  
спеціальності «Соціологія»

Національний університет «Києво-Могилянська академія»

**О. Г. ПУГАЧОВА**

кандидат економічних наук,  
доцент кафедри соціології

Національний університет «Києво-Могилянська академія»

## **РИЗИКИ ПОПУЛЯРИЗАЦІЇ ПОРОДЖУВАЛЬНОГО ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ДОСЛІДЖЕННЯ САЙТІВ СОЦІАЛЬНИХ МЕРЕЖ**

*У статті розкриваються причини та ранні наслідки поточної кризи вільного машинного доступу до агрегованих даних сайтів соціальних мереж за допомогою прикладних програмних інтерфейсів, або API. Зокрема, розглянуто кейси таких популярних вебсайтів, як Twitter (нещодавно перейменованний на X) та Reddit. Проаналізовано основні фактори, що зумовили режим відносно безперешкодного використання масових онлайн-даних, який мав місце до розповсюдження так званих ШІ-чатботів, найпопулярнішим серед яких наразі є ChatGPT від американської компанії OpenAI. Пояснено, як саме значний прорив у сфері породжувального штучного інтелекту, ознаменований публічним запуском ChatGPT наприкінці 2022-го року, є пов'язаним зі змінами політик доступу до API із боку одних із найбільших сайтів соціальних мереж. Систематизовано дії соціальної мережі X та сайту тематичних обговорень Reddit навколо доступу до своїх даних через API для членів дослідницької спільноти протягом 2023-го року, які дозволяють особливо чітко простежити причинно-наслідковий зв'язок між «гонкою озброєнь» у сфері штучного інтелекту та кризою у сфері досліджень, що покладаються на дані сайтів соціальних мереж. Зокрема, акцентовано увагу на патерні обмежень X відносно своїх публікацій та їхніх намагань встановити контроль за використанням своїх даних у судовому порядку, а також обґрунтовано, чому нові правила API Reddit, незважаючи на їхні більш обмежені наслідки, усе ще є несприятливими для роботи вчених. Описано, якої саме шкоди дослідникам завдають або можуть завдати нові тенденції в тому, як соціальні мережі схильні управляти своїми даними. Висвітлено важливість та релевантність подальшого розвитку досліджень на основі даних сайтів соціальних мереж, зокрема через забезпечення легкого доступу до API, як для наукової спільноти, так і для аудиторії соціальних мереж та їхнього керівництва. Проведений аналіз дає підстави стверджувати, що зараз, у час переосмислення комерційної цінності користувачьких даних, захист*

*інтересів дослідників соціальних мереж та дослідників інтернету загалом є особливо актуальним.*

**Ключові слова:** сайти соціальних мереж, породжувальний штучний інтелект, Twitter (X), Reddit, прикладний програмний інтерфейс (API).

**Постановка проблеми.** Сайти соціальних мереж (далі також ССМ) або, використовуючи повсякденну лексику, соціальні мережі, всього за пару десятиліть стали чи не наймасштабнішим та найбільш впливовим публічним простором – як онлайн, так і офлайн. Глобально успішні ССМ, як-от Facebook, Twitter, Instagram та подібні, поступово ввібрали в себе більшість знайомих із реального світу проявів міжперсональної взаємодії, зайнявши провідну роль у фасилітації обміну соціальною підтримкою [1, с. 675]. Відповідно, публічно видима частина соцмережевого «контенту» (поширений спосіб покликатися на всю сукупність онлайн-публікацій на тій чи тій вебплатформі) швидко стала для вчених у сфері соціальних наук улюбленим об'єктом аналізу, а масовий збір соцмережевих даних – основою численних досліджень. Останні суттєво вплинули на забезпечення здорового розвитку онлайн-спільнот та розуміння людських відносин загалом.

Однак цей статус-кво не був сталим та історично покладався на сприятливий збіг факторів, за якого ліберальний доступ до агрегованих даних сайтів соціальних мереж був індустрійною нормою. Незважаючи на те, що поступовий відхід від цієї філософії (продиктований переважно фінансовими та безпековими міркуваннями) тривав уже багато років, переломним для сфери досліджень соціальних мереж можна вважати саме 2023-ий рік. Всього за 12 місяців використання штучного інтелекту через медіум так званих чатботів стало всюдисущим та призвело до перегляду цінності соцмережевих даних із далекосяжними наслідками для дослідницької спільноти.

**Метою статті** є аналіз витоків поточної кризи доступу до соцмережевих даних для членів дослідницької спільноти та її перебігу протягом 2023-го року, а також огляд її короткострокових наслідків.

#### **Виклад основного матеріалу.**

##### *Минуле соцмережевих досліджень: культура відкритості*

Потужність соціальних мереж як поля для включених спостережень та збору емпіричних даних (не лише про соціальну підтримку, але й про людську взаємодію загалом) майже не потребує обґрунтування: структуруючим ефектом соцмережевих платформ для комунікації користуються не лише науковці, а і звичайні користувачі, які вводять ключові слова в поле пошуку чи переходять по так званих «хештегах». Додатковим бонусом соцмереж як повністю цифрових сутностей є значне спрощення масового збору й обробки інформації, порівняно з даними на аналогових носіях, що дозволяє великою мірою автоматизувати емпіричний етап дослідження.

З огляду на це, бум досліджень, заснованих на аналізі публікацій у соціальних мережах, який послідував за популяризацією останніх,

був очікуваним, хоч етичний статус масового збереження персональних публікацій, потенційно асоційованих із різними соціодемографічними чи персональними маркерами, без відома авторів цих публікацій залишається суперечливим [2, с. 45].

Основними методами для пошуку та збереження вебконтенту серед індивідуальних користувачів є так званий скрейпінг даних (від англ. *scrape* – зішкрябувати) та прикладні програмні інтерфейси, або API (від англ. *application programming interface*), які публікуються самими соціальними мережами. Хоча скрейпінг є більш гнучким та універсальним методом, він вимагає суттєво вищого рівня технічних навичок, ніж використання API, та в багатьох випадках порушує офіційні умови користування того чи іншого соцмережевого сервісу. Тож за умови, коли API пропонує доступ до необхідних родів даних у належному обсязі, дослідники схильні покладатися саме на нього.

Реакція сайтів соціальних мереж на зростання цікавості до користувачького контенту на своїх серверах із боку дослідницької спільноти різнилась, однак консенсус серед найбільших гравців ринку тривалий час полягав у толеруванні цього сценарію використання своїх даних та ставленні до нього з «пасивною апатією» [3, с. 3]. Можна сказати, що принцип відкритості даних до машинної обробки з боку третіх сторін – як компаній так і індивідів – був закладений у ССМ на системному рівні, адже на ранніх етапах існування незалежних соцмережевих проєктів від того, наскільки багато інших систем чи платформ зможе взаємодіяти з ними, залежало те, наскільки поширеними вони зможуть стати.

Водночас навіть зважаючи на рамкову політику відкритості, довгостроковий тренд полягав у поступовому накопиченні обмежень. Зростання аудиторій соціальних мереж було прямо пропорційним зростанню популярності їхніх API, витрат на їхню підтримку й контроль за недобросовісним використанням. Що більше людей отримували доступ до API-інструментів, то частішали випадки зловживання автоматичним зберіганням великих обсягів користувацьких даних, зокрема й ті, що набували широкого розголосу.

Своєрідною кульмінацією цього довготривалого процесу стала історія дослідницької фірми Cambridge Analytica, яка у 2018-му році перетворилася на об'єкт міжнародного скандалу через заволодіння персональними даними десятків мільйонів користувачів Facebook та їхнє неправомірне використання для таргетування політичної реклами. Користувацькі дані Facebook, що опинилися в центрі скандалу, були зібрані за допомогою API соцмережі Александром Коганом, академіком, який на той час працював у Кембриджському університеті. Інцидент із Facebook та Cambridge Analytica потрапив на шпальти світових ЗМІ та призвів до суттєвих обмежень у роботі API як для Facebook, так і для Instagram (обидві соцмережі мають спільну компанію-власницю) [4].

*ШІ-чатботи та новий розділ в історії досліджень соцмереж*

Скандал навколо Facebook та Cambridge Analytica призвів не лише до перегляду правил доступу API, а й до широких політичних та соціальних наслідків, ставши в популярному дискурсі своєрідним символом безвідповідального ставлення великих соцмережових компаній до персональних даних своїх користувачів. Однак можна посперечатися, що навіть більш значущий зсув в узагальненому сприйнятті даних соціальних мереж стався значно більш недавно.

ChatGPT, так званий ШІ-чатбот, доступ до якого широка онлайн-спільнота отримала наприкінці 2022-го року, практично одразу став першою глобально успішною спробою поєднати демонстраційну версію так званого породжувального штучного інтелекту (англ. *generative artificial intelligence*) зі зручним та інтуїтивним користувацьким інтерфейсом. Згідно з деякими оцінками, всього за два місяці ChatGPT досягнув понад ста мільйонів щомісячних активних користувачів (стандартна метрика для визначення популярності онлайн-сервісів), оновивши рекорд із найшвидше зростаючого застосування в історії інтернету [5].

Феномен ChatGPT ознаменував досягнення сейсмічного зрушення в розвитку штучного інтелекту та, згідно з журналістськими репортажами [6], призвів до кризових реакцій усередині найбільших технологічних компаній, таких як Google та Meta, які були на різних стадіях розробки власних ШІ-проектів (уже навесні 2023-го року Google випустила власний ШІ-чатбот під назвою Bard). Однак найбільш релевантним для цієї статті наслідком переосмислення рівня й темпу прогресу штучного інтелекту стало безпрецедентно активне обговорення субстрату, в якому «вирощують» великі мовні моделі, такі як GPT-3 та GPT-4 (що лежать в основі ChatGPT), а саме контенту публічних вебсторінок – у тому числі сторінок сайтів соціальних мереж [7].

У результаті, як буде показано нижче, керівництво принаймні деяких із цих сайтів прийшло до висновку, що той величезний масив онлайн-взаємодій живих людей, який останні багато років накопичувався в їхніх дата-центрах, є надзвичайно цінним не лише для точного таргетування рекламних оголошень, але і як матеріал для навчання (чи «тренування») ШІ-моделей. Із цього випливало, що актуальна філософія оперування соцмережових сервісів із її історично зумовленою орієнтацією в бік відкритості та інтероперабельності (можливості співпрацювати з іншими сервісами) даних призвела до втрати потенційно значної частки цієї нововідкритої цінності.

Нарешті, перегляд цієї філософії безпосередньо призвів до значних викликів для принаймні частини дослідницької спільноти, яка покладається в своїх проєктах на дані соціальних мереж. На мою думку, найбільш ілюстративними в цьому плані наразі є кейси Twitter (переіменованого на X у липні 2023-го року), сервісу так званих мікроблогів, який здобув світову популярність своєю імплементацією коротких

повідомлень (або «твітів»), та Reddit, сайту тематичних спільнот та обговорень, який покладається на агрегацію та ранжування покликань на інші вебсторінки та містить багато елементів соціальної мережі.

Сумарна аудиторія кожного з цих сервісів зокрема вимірюється сотнями мільйонів користувачів, хоча жоден із цих сервісів наразі не є популярним в Україні. Незважаючи на невелике проникнення Twitter в укрнеті, особливо відносно інших соцмереж, таких як TikTok чи Facebook, кількість активних користувачів з України на платформі помітно зросла після початку російського вторгнення 2022-го року, досягши, за однією з оцінок, майже 600 тисяч [8]. Reddit в Україні залишається ще більш нішевим сервісом, хоча там, наприклад, поступово з'являються українські ЗМІ, пробуючи капіталізуватися на збільшеному зацікавленні в українських темах із боку міжнародної аудиторії, що стежить за світовими подіями [9].

#### *Приклад X (Twitter)*

Соціальна мережа, що наразі називається X, існувала під назвою Twitter із 2006-го року та тривалий час служила полем для соціологічних досліджень на широкий спектр тем завдяки своєму безплатному API та відносній ресурсоощадності аналізу індивідуальних повідомлень, які аж до 2017-го року мали ліміт у всього 140 символів. Через легкість доступу до свого API Twitter донедавна був найпопулярнішим об'єктом для дослідників онлайн-платформ [10]. Так, за десятиліття, що передувало 2021-му року, за однією з оцінок, дані Twitter були використані в понад десять тисячах статей [11, с. 1].

Нова ера в існуванні Twitter, і як компанії, і як соцмережі почалася в жовтні 2022-го року, коли технологічний підприємець Ілон Маск викупив Twitter Inc. Він одразу звільнив майже все вище керівництво колишньої публічної компанії, яка стала приватною. Всього за рік із моменту покупки компанії, невдовзі перейменованої на X Corp., Маск із групою лоялістів-керівників запровадили довгий перелік радикально нових політик у функціонуванні схожим чином перехрещеної соціальної мережі. Однією з найпомітніших серед цих політик був великий наголос на просуванні платних підписок різної вартості (відомих як X Premium) як найкращого способу користування соцмережею [12; 13].

На початку 2023-го року соціальна мережа анонсувала рішення про скасування безплатного доступу до свого API, що зрештою було швидко відкликане після критики на користь кількох рівнів доступу – «Безплатний», «Базовий» та «Ентерпрайз». У рамках безплатного рівня дозволялося лише автоматичне публікування твітів, але не їхнє читання (останнє є необхідним для агрегації публікацій інших користувачів Twitter/X). Для можливості відсилати читальні запити через API тепер потрібно сплачувати принаймні 100 доларів щомісяця, хоча обсяг доступних даних при цьому залишається на кілька порядків меншим відносно того, що раніше можна було отримати безплатно [14]. Так, незважаючи на те, що Twitter пропонував платні версії API

і до ери Маска, безплатний варіант усе ще давав доступ до мільйонів публікацій щомісяця [15]. Однак згідно з новими тарифами, залежно від рівня користування, ціна доступу до API X могла сягати десятків тисяч доларів щомісяця [16].

Ці нові правила на момент своєї першої публікації не робили жодних винятків для сприяння дослідницькій діяльності та не містили конкретних майбутніх кроків у цьому напрямку, окрім загальної декларації про «пошук нових шляхів продовження служіння науці» [17]. Це спричинило занепокоєння як із боку окремих учених соціальних мереж, так і адвокаційних груп, наприклад «Коаліції незалежних технологічних досліджень». Остання опублікувала відкритого листа, привертаючи увагу до негативного впливу, якого нові правила завдавали дослідницькій спільноті [14]. Станом на кінець 2023-го року спеціальної програми доступу API для дослідницьких цілей так і не з'явилося; розробницька платформа X пропонує дослідникам ті самі платні плани, що й для всіх інших [18].

Офіційним поясненням для API-нововведень із боку компанії був високий рівень зловживання безплатним API, що призводив до великої кількості неавтентичних акаунтів (ботів), які займалися шкідливою діяльністю на платформі [19]. Однак пізніші дії Twitter/X дають вагомі підстави припустити, що обмеження API були принаймні частково мотивовані також бажанням припинити неконтрольоване автоматичне експортування даних соцмережі. Влітку Twitter тимчасово відключив можливість перегляду публікацій на сайті для незареєстрованих користувачів. Коментуючи цю зміну, Маск поскаржився на те, що «кілька сотень організації (можливо, більше) надзвичайно агресивно скреїпили дані Twitter» [20]. Наступного дня мільярдер оголосив чергову несподівану зміну в функціонуванні своєї соціальної мережі: тимчасові обмеження на кількість публікацій, які могли переглянути зареєстровані користувачі в межах доби. Ці обмеження становили 6 тисяч публікацій для «підтверджених» облікових записів (відколи Маск перейняв керування Twitter, підтвердження акаунта можна отримати автоматично як одну з функцій підписки Premium), 600 публікацій для «непідтверджених» акаунтів (без Premium-підписки, яка є необхідною умовою верифікації) та лише 300 публікацій для «нових непідтверджених» акаунтів. Уже наступного дня ліміти були підвищені до 10 тисяч, 1000 й 500 публікацій відповідно. Пояснюючи причини обмежень, Маск заявив, що вони є реакцією на «надзвичайні обсяги скреїпінгу даних та маніпуляції систем» (не надавши будь-яких подальших деталей чи доказів своїх слів) [21].

Окрім того, X Corp. почала переслідувати звинувачення про скреїпінг своїх даних у судовому порядку. Зокрема, компанія подала позов на одну з ізраїльських фірм за «нелегальний скреїпінг даних технологічних компаній, як-от X Corp.» [22], а також на неприбуткову організацію «Центр протидії цифровій ненависті» (англ. Center

for Countering Digital Hate) за «незаконний скрейпінг даних Twitter та їхнє використання для створення хибних досліджень про платформу», що, на думку X Согр., спричинило відтік рекламодавців із вебсайту та втрату десятків мільйонів доларів доходу. Незадовго перед цим «Центр протидії цифровій ненависті» опублікував дослідження, в якому наводив докази недостатньої реакції X на випадки мови ненависті в твітах платних підписників [23].

#### *Приклад Reddit*

Іншим кейс-стаді тренду із занепаду культури відкритих онлайн-даних у 2023-му році став Reddit. Цей сервіс тематичних обговорень був заснований 2005-го року, на рік раніше Twitter. Одним із його засновників, а з 2015-го року головним виконавчим директором, є Стів Гаффмен. За кілька місяців після публічного релізу ChatGPT він заявив, що «масив даних Reddit є дуже цінним. Але ми не повинні віддавати цю цінність найбільшим у світі компаніям безоплатно» [24] (OpenAI, компанія-розробниця ChatGPT, має домовленість про стратегічне партнерство з Microsoft, яка зробила мільярдні інвестиції в компанію, переважно у вигляді комп'ютерних потужностей для навчання її ШІ-моделі [25]; звичайно, Гаффмен, найімовірніше, мав на увазі не лише Microsoft, а й інші глобальні технологічні компанії).

Заява гендиректора була зроблена в контексті анонсу Reddit про впровадження платного доступу до свого API. Цей анонс, як і у випадку Twitter/X, зіткнувся з гучною критикою, особливо з боку тих, хто користувався платформою через сторонні програми-клієнти. За кілька тижнів до вступу в силу API-обмежень модератори тисяч Reddit-спільнот у знак протесту зробили свої сторінки недоступними для перегляду сторонніми користувачами. Однак, незважаючи на безпрецедентний масштаб протестів, компанія не змінила своїх планів, увівши в дію платний API першого липня [26].

На відміну від X, нові API-тарифи Reddit були спрямовані першочергово на розробників сторонніх клієнтів Reddit – найбільш інтенсивних користувачів API. Також Reddit, на відміну від X, публічно заявив, що доступ для «некомерційних дослідників та академіків залишається безплатним у межах опублікованого ліміту користування» [27] та опублікував форму для прийняття запитів на надання своїх даних для академічних цілей. Однак через місяць після цієї заяви вищезгадана «Коаліція незалежних технологічних досліджень» випустила звіт про своє опитування користувачів API Reddit (як для дослідницьких, так і недослідницьких цілей), у якому підсумовувала, що «розмиті обіцянки» Reddit «не повністю задовольняють потреби дослідників... і дають можливість Reddit блокувати використання API для досліджень» [28]. Справді, для всіх досліджень на основі своїх даних Reddit вимагає отримувати копію перед публікацією [29].

**Висновки і пропозиції.** Протягом понад десятиліття використання даних сайтів соціальних мереж у соціальних дослідженнях було

історією успіху як для дослідницької спільноти, яка могла втілювати раніше нереалістичні за масштабом проекти та заповнювати прогалини в науковому знанні, так і для самих соціальних мереж, які мали нагоду використовувати результати досліджень, щоб зробити свої платформи безпечнішими та більш ефективними для своєї аудиторії.

Однак початок ери породжувального ШІ та запуск низки чатботів, орієнтованих на користування неспеціалістами, спонукали до зміщення сприйняття даних сайтів соціальних мереж – та Web 2.0-контенту загалом – із боку їхніх компаній-власниць. Чергове оновлення політики приватності Google у липні 2023-го року, згідно з яким компанія залишала за собою право використовувати будь-який публічний текст в інтернеті для тренування своїх ШІ-моделей [30], послужило втіленням нового статусу-кво, де цифрові дані розглядаються передусім як ресурс у гонці з побудови найкращої та найбільш комерційно успішної версії штучного інтелекту.

Історії X та Reddit у 2023-му році демонструють, що побічною втрапою в цій технологічній трансформації може стати вільний доступ до соцмережових даних із боку значної частини дослідницької спільноти, особливо незалежних дослідників без університетської афіліації. Впровадження нових підходів до монетизації API становить серйозну загрозу проведенню нових досліджень за допомогою соцмережового контенту та відтворенню результатів уже опублікованих. Крім того, непевності щодо майбутнього додає те, що дії X та Reddit є частиною ширшого тренду, до якого можна включити запровадження надзвичайно обмежуючих умов використання своїх даних іншими великими соціальними платформами, як-от TikTok та LinkedIn [31, с. 6–7] та побоювання щодо закриття CrowdTangle, інструмента, що дозволяв дослідникам та журналістам аналізувати дані Facebook та Instagram [32].

Додаткову небезпеку становлять судові позови, до яких віднедавна вдається X Corp. Зокрема її вищезгадані звинувачення проти «Центру протидії цифровій ненависті» є серйозним викликом, що може призвести до популяризації сприйняття досліджень сайтів соціальних мереж не як діяльності задля кращого розуміння того, що на них відбувається (в інтересах не лише вузького наукового кола, а й глобальної інтернет-аудиторії), але як ворожої стратегії з експліцитною метою заподіяти фінансової шкоди конкретній соцмережовій корпорації. Така тенденція має потенціал стати навіть більш руйнівною для досліджень у сфері сайтів соцмереж, ніж збільшення вартості машинного доступу до публікацій.

Станом на сьогодні вже можна стверджувати, що екосистемі онлайн-досліджень було завдано значної шкоди, оскільки багатьом членам академічної сфери довелося відмовитися від планів збору даних Twitter/X та переорієнтувати дизайни досліджень чи навіть переглянути теми своїх проектів. Як зазначають представники спільноти, навіть оплата «базового» тарифу доступу до API соцмережі є суттєвим фінансовим тягарем [19].



Отже, у поточному технологічному кліматі, де боротьба за лідерство в сфері штучного інтелекту залишається активною, для дослідників є особливо важливим, по-перше, об'єднання в адвокаційні групи; по-друге, акти колективної дії, що мають на меті привернути увагу ширшої онлайн-спільноти, в тому числі інтернет-регуляторів, до потреби захисту своїх інтересів (які, знову-таки, узгоджуються із суспільним інтересом); і по-третє, на більш загальному рівні, випрацювання нового, справедливого суспільного договору щодо використання користувачьких даних для розвитку ШІ між інтернет-компаніями та їхніми користувачами.

### Список використаної літератури

1. Liu C.-Y., Yu C.-P. Can Facebook use induce well-being?. *Cyberpsychology, Behavior, and Social Networking*. 2013. Т. 16, № 9. С. 674–678. URL: <https://doi.org/10.1089/cyber.2012.0301> (дата звернення: 24.12.2023).
2. Chapter 2: Users' views of ethics in social media research: informed consent, anonymity, and harm / M. L. Williams та ін. *The Ethics of Online Research*. 2017. С. 27–52. URL: <https://doi.org/10.1108/s2398-601820180000002002> (дата звернення: 24.12.2023).
3. Tromble R. Where have all the data gone? A critical reflection on academic digital research in the post-api age. *Social Media + Society*. 2021. Т. 7, № 1. URL: <https://doi.org/10.1177/2056305121988929> (дата звернення: 24.12.2023).
4. Schroepfer M. An update on our plans to restrict data access on Facebook. *Facebook Newsroom*. URL: <https://about.fb.com/news/2018/04/restricting-data-access/> (дата звернення: 24.12.2023).
5. Hu K. ChatGPT sets record for fastest-growing user base – analyst note. *Reuters*. URL: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> (дата звернення: 24.12.2023).
6. Inside the A.I. arms race that changed Silicon Valley forever / K. Weise та ін. *New York Times*. URL: <https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html> (дата звернення: 24.12.2023).
7. Paul K. Meta's new AI assistant trained on public Facebook and Instagram posts. *Reuters*. URL: <https://www.reuters.com/technology/metas-new-ai-chatbot-trained-public-facebook-instagram-posts-2023-09-28/> (дата звернення 24.12.2023).
8. Kemp S. Digital 2023: Ukraine. *DataReportal*. URL: <https://datareportal.com/reports/digital-2023-ukraine> (дата звернення: 24.12.2023).
9. Таргачний О. Reddit для медіа: як, навіщо та кому це треба. *Медіамейкер*. URL: <https://mediamaker.me/reddit-dlya-media-yak-navishho-ta-komu-cze-treba-5811/> (дата звернення: 24.12.2023).
10. Kupferschmidt K. Twitter's plan to cut off free data access evokes 'fair amount of panic' among scientists. *Science*. URL: <https://www.science.org/content/article/twitters-plan-cut-free-data-access-evokes-fair-amount-panic-among-scientists> (дата звернення: 24.12.2023).
11. Antonakaki D., Fragopoulou P., Ioannidis S. A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*.

2021. Т. 164. С. 114006. URL: <https://doi.org/10.1016/j.eswa.2020.114006> (дата звернення: 24.12.2023).

12. Elon Musk's Twitter is a Shakespearean psychodrama set in Silicon Valley / K. Wagner та ін. *Bloomberg*. URL: <https://www.bloomberg.com/news/features/2022-12-14/elon-musk-twitter-ownership-full-of-firings-ad-cuts-chaos> (дата звернення: 24.12.2023).

13. Milmo, D. 'Musk destroyed all that': Twitter's business is flailing after a year of Elon. *The Guardian*. URL: <https://www.theguardian.com/technology/2023/oct/27/elon-musk-x-twitter-takeover-revenue-users-advertising> (дата звернення: 24.12.2023).

14. Executive Board. Letter: Twitter's new API plans will devastate public interest research. *The Coalition for Independent Technology Research*. URL: <https://independenttechresearch.org/letter-twitters-new-api-plans-will-devastate-public-interest-research/> (дата звернення: 24.12.2023).

15. Mehta I. Twitter's restrictive API may leave researchers out in the cold. *TechCrunch*. URL: <https://techcrunch.com/2023/02/14/twitters-restrictive-api-may-leave-researchers-out-in-the-cold/> (дата звернення: 24.12.2023).

16. Stokel-Walker C. Twitter's \$42,000-per-month API prices out nearly everyone. *WIRED UK*. URL: <https://www.wired.co.uk/article/twitter-data-api-prices-out-nearly-everyone> (дата звернення: 24.12.2023).

17. Розробники X. [Служіння академічній спільноті]. X. URL: <https://x.com/XDevelopers/status/1641222788911624192> (дата звернення: 24.12.2023).

18. Academic research. X Developer Platform. URL: <https://developer.x.com/en/use-cases/do-research/academic-research> (дата звернення: 24.12.2023).

19. Fung B. Elon Musk pledged transparency at Twitter. But he's walling off researchers. *CNN Business*. URL: <https://edition.cnn.com/2023/02/09/tech/musk-twitter-transparency-researchers/index.html> (дата звернення: 24.12.2023).

20. Musk E. [Тимчасові ліміти]. X. URL: <https://x.com/elonmusk/status/1675187969420828672> (дата звернення: 24.12.2023).

21. Saligrama, A. Musk says Twitter will limit how many tweets users can read. *Reuters*. URL: <https://www.reuters.com/technology/musk-says-twitter-applies-temporary-limit-address-data-scraping-system-2023-07-01/> (дата звернення: 24.12.2023).

22. Wrobel S. Musk's X Corp sues Israel's Bright Data for scraping data. *The Times of Israel*. URL: <https://www.timesofisrael.com/musks-x-corp-sues-israels-bright-data-for-illegally-scraping-data/> (дата звернення: 24.12.2023).

23. Roth E. Elon Musk's X sues anti-hate researchers for allegedly scraping data from Twitter. *The Verge*. URL: <https://www.theverge.com/2023/8/1/23815515/twitter-cddh-anti-hate-research-group-lawsuit> (дата звернення: 24.12.2023).

24. Isaac M. Reddit wants to get paid for helping to teach big A.I. systems. *New York Times*. URL: <https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html> (дата звернення: 24.12.2023).

25. Albergotti R. OpenAI has received just a fraction of Microsoft's \$10 billion investment. *Semafor*. URL: <https://www.semafor.com/article/11/18/2023/openai-has-received-just-a-fraction-of-microsofts-10-billion-investment> (дата звернення: 24.12.2023).

26. Peters J. How Reddit crushed the biggest protest in its history. *The Verge*. URL: <https://www.theverge.com/23779477/reddit-protest-blackouts-crushed> (дата звернення: 24.12.2023).

27. Key facts to understanding Reddit's recent API updates. *Reddit Blog*. URL: <https://www.redditinc.com/blog/apifacts> (дата звернення: 24.12.2023).

28. Survey report: Reddit's actions continue to undermine moderation & research / S. Gilbert та ін. *The Coalition for Independent Technology Research*. URL: <https://independenttechresearch.org/reddit-survey-results/> (дата звернення: 24.12.2023).

29. Reddit developer services: Research. *Reddit*. URL: [https://support.reddithelp.com/hc/en-us/articles/14945211791892#h\\_01H69EJ3EFY7G7HNV17ASH24KS](https://support.reddithelp.com/hc/en-us/articles/14945211791892#h_01H69EJ3EFY7G7HNV17ASH24KS) (дата звернення: 24.12.2023).

30. Germain T. Google says it'll scrape everything you post online for AI. *Gizmodo*. URL: <https://gizmodo.com/google-says-itll-scrape-everything-you-post-online-for-1850601486> (дата звернення: 24.12.2023).

31. Platform-controlled social media APIs threaten open science / B. I. Davidson та ін. *Nature Human Behaviour*. 2023. URL: <https://doi.org/10.1038/s41562-023-01750-2> (дата звернення: 24.12.2023).

32. Facebook's gutting of CrowdTangle: a step backward for platform transparency. *Algorithm Watch*. URL: <https://algorithmwatch.org/en/crowdtangle-platform-transparency/> (дата звернення: 24.12.2023).

### **Lyubenko R. Ya., Puhachova O. H. Risks of mainstream generative AI for social networking sites' research**

*The article discusses the causes and early effects of the current crisis in free machine access to aggregated social networking sites' data using application programming interfaces, or API. In particular, the cases of such popular websites as Twitter (recently renamed to X) and Reddit are considered. The main factors leading to the regime of relatively unfettered access to mass online data before the surge of so-called AI chatbots, the most popular among them currently being ChatGPT from the American company OpenAI, are analyzed. The text explains, how exactly the breakthrough in the space of generative AI, heralded by the public launch of ChatGPT at the end of 2022, is related to the changes in API access policies on the part of the biggest social networking sites. The actions of social network X and topical discussions site Reddit around the research community's access to their data through API during 2023 are summarized; they show the causal link between the AI "arms race" and the crisis of research that relies on social networks' data especially clearly. Particular attention is paid to the pattern of limitations imposed by X on its posts as well as the company's litigious tendency in trying to control the usage of its data. Also, the article demonstrates how the new Reddit API rules, despite their more narrow impact, are still unfavorable to the researchers. The harmful consequences, which the researchers either have suffered or might suffer due to the new trends in how social networking sites tend to manage their data are described. The importance and relevance of further developing the research based on social networking sites data, in particular through easy API access, both to the academic community and social networks' users and leadership, is highlighted. The analysis conducted gives grounds to suggest that right now, during the major rethinking of what the value of user data is, the protection of the interests of researchers who study social networks and the Internet in general is especially pertinent.*

**Key words:** social networking sites, generative AI, Twitter (X), Reddit, application programming interface (API).