

Шекеряк-Кушка Я. В.

Львівський національний університет імені Івана Франка

ШТУЧНИЙ ІНТЕЛЕКТ У ЖУРНАЛІСТИЦІ ДАНИХ: МОЖЛИВОСТІ, ЗАГРОЗИ ТА ЕТИЧНІ ДИЛЕМИ

Метою наукової розвідки є дослідження використання елементів штучного інтелекту у роботі журналіста. Розглянуто, як змінюється професійна діяльність медійників під впливом технологій ШІ, і проаналізовано, у яких саме формах ШІ інтегрується у журналістську практику – від автоматизації рутинних процесів до складного аналізу та генерації контенту. У статті аналізується зростаюча роль ШІ у медіа загалом та тих аспектів журналістської діяльності, які охоплюють роботу з даними зокрема визначаючи журналісту даних як одну із медіапрактик, де потенціал ШІ розкривається найповніше. Особливу увагу приділено розвитку цих процесів у українському медійному контексті, під час повномасштабної війни.

Проаналізовано як журналісти використовують готові інструменти та розробляють власні ШІ-рішення, описано приклади застосування у проєктах з OSIN (розвідка на основі відкритих даних), обробки супутникових знімків, ідентифікації воєнних злочинців та боротьби з дезінформацією. Наведено приклади ефективного використання ШІ українськими та міжнародними медіа. Визначено, що українські журналісти даних мають досвід як генеративного використання ШІ, так і аналітичного, пов'язаного з виявленням трендів, аналізом та візуалізацією даних.

Окреслено етичні виклики, пов'язані з інтегруванням штучного інтелекту у роботу журналістів, такі як прозорість, маніпулювання аудиторією, проблема авторства, «галюцинації», точність. Розкрито специфічні безпекові аспекти використання ШІ у журналістиці даних в умовах війни, такі як захист джерел інформації, анонімізація чутливих даних та уникнення загроз для цивільного населення. Визначено основні переваги та недоліки використання ШІ для інтеграції даних у журналістику, окреслено ключові проблеми та можливі шляхи їх вирішення. Проаналізовано успішні кейси використання ШІ журналістами даних та на їх основі сформульовано рекомендації щодо етичного використання ШІ у журналістиці та забезпечення безпеки в умовах сучасної медійної дійсності (прозорість, фактчек, маркування матеріалів, відкритість методології та захист приватності).

Ключові слова: журналістика даних, штучний інтелект, ШІ, персональні дані, алгоритми, етика, безпека, війна, дезінформація, Україна, нові медіа, медіа.

Постановка проблеми. Сучасний медіаландшафт стрімко трансформується під впливом широкого впровадження штучного інтелекту на всіх етапах виробництва контенту: від генерування ідей, пошуку та обробки інформації до створення текстів, візуальних і мультимедійних матеріалів, аналізу аудиторії та вибору найефективніших каналів поширення, а часом і прямого зв'язку з потенційними слухачами/глядачами. У зв'язку з цим у багатьох професійних колах, зокрема журналістських, з'явилися дискусії з побоюваннями втрати роботи, адже людську працю (повністю або частково) замінить ШІ. Водночас піднімаються питання загрози втрати креативності, експертності, цінності людської роботи.

Якщо спочатку дискусії точилися навколо питання, чи замінить ШІ людей, якісь професії та напрямки, зокрема журналістику, то згодом

фокус змінився на щось, що могло зустрітися під заголовком «топ 10 нових професій в еру ШІ» чи «7 перевірених способів заробляти \$5,000 на місяць за допомогою ШІ інструментів». Зрештою, видається логічніше зосередитися на тому як саме змінюється професійна діяльність медійників під впливом розвитку нових технологій загалом і штучного інтелекту зокрема, аніж заперечувати або ігнорувати ці трансформації. У цьому контексті також особливої актуальності набуває обговорення нових професійних ролей і компетентностей.

У цьому контексті особливої уваги заслуговує журналістика даних як одна з тих медійних практик, що найбільш органічно інтегрує штучний інтелект у професійну діяльність, і де його подальше застосування видається ледь не найбільш обґрунтованим.

У суспільстві, де обсяг інформації зростає щодня, а усі наші дії тим чи іншим чином записані у даних, ця медійна практика набуває особливого значення. Вона дозволяє не просто передавати новини, а структурувати складні процеси й виявляти зв'язки, які неочевидні без аналізу великих масивів інформації. А також розповідати історії з того, що здавалося б є просто масивом нулів і одиниць. Це особливо актуально в умовах війни, коли одночасно циркулюють сотні інформаційних потоків – від офіційних зведень до свідчень очевидців, і кожне з цих джерел потребує верифікації, порівняння й контексту.

Інтеграція штучного інтелекту у цей процес не є чимось абсолютно новим. Саме в журналістиці даних використання багатьох автоматизованих алгоритмів було звичною (з урахуванням особливостей регіону і медіаландшафту) практикою ще до стрімкого розвитку генеративного ШІ. Скажімо для збору даних, їх структурування, аналізу (наприклад пошуку трендів чи винятків), очищення, побудови графіків й інших видів візуалізації. Це дає змогу, з одного боку, спиратися на напрацьований досвід роботи з алгоритмами в межах цієї медіапрактики, а з іншого – проаналізувати, як нові можливості, які відкриває сучасний ШІ, уже сьогодні трансформують підходи до створення журналістських матеріалів на основі даних.

Особливою уваги також заслуговує досвід використання штучного інтелекту журналістами даних під час висвітлення війни в Україні: для збору даних із відкритих джерел, ідентифікації військової техніки та російських воєнних злочинців, картографування подій чи роботи з супутниковими знімками. Адже тут є як і переваги: швидкість, масштабованість, аналітична точність, аргументованість, можливість використання матеріалів надалі як доказів у судових справах, так і ризики – етичні дилеми, питання приватності та безпеки (зокрема відповідальності за помилки алгоритмів).

Аналіз останніх досліджень і публікацій. Попри те, що поняття «журналістика даних» відоме в науковому дискурсі уже останні 10–20 років, усе ще більшість досліджень, на які можна спиратися, є закордонними. Часто дослідження журналістики даних формуються на основі аналізу та систематизації практичного досвіду, як, наприклад, зробили автори двох класичних посібників по роботі з даними («Журналістика даних: Посібник» (Data journalism handbook) і «Посібник із журналістики даних: до критичної практики роботи з даними» (Data journalism handbook: towards a critical data practice), Ліліана Бонегру (Liliana Bounegru),

Люсі Чемберс (Lucy Chambers), Джонатан Грей (Jonathan Gray) [1; 2]), розвинувши потім теми критичного ставлення до даних, відповідальності журналістики даних, організації роботи відділу даних у медіа та знаходження шляхів впливу на розв'язання реальних проблем через матеріали на основі даних. В Україні схожі інструкції та посібники для роботи з даними частіше роблять не так для медійників, як для громадських організацій та органів місцевого самоврядування (Газін, А., Горбаль, А., Кубай, Д., Шульга, Є., Шаповаленко, Г.; Шадська У.) [3]. Водночас з'являються праці, присвячені дослідженню ролі даних у висвітленні російсько-української війни (зокрема Мерабет Д. [4], Алекса ван Сікл (Alexa van Sickle), Юніс Оу (Eunice Au) [5]). Однак проблематика безпеки та етичних викликів у цих роботах досліджена лише частково й потребує подальшого осмислення у науковому полі.

Використання штучного інтелекту у медіа – також не нова тема для науковців, так, наприклад журналіст і аналітик Франческо Марконі у книзі «Newsmakers: штучний інтелект і майбутнє журналістики» (Newsmakers: Artificial Intelligence and the Future of Journalism) виокремлює три хвилі розвитку ШІ у журналістиці: автоматизація стандартних новинних форматів (фінансові звіти, спортивні результати); розширення можливостей ШІ для аналізу великих обсягів даних і виявлення закономірностей; генерація текстів за допомогою мовних моделей, здатних створювати матеріали у заданому стилі (наприклад, ChatGPT) [6]. Якраз з появою останніх почало з'являтися ще більше наукових розвідок на тему взаємодії таких типів ШІ й медіа [7].

Взаємодія так званих big data і штучного інтелекту також становить інтерес дослідників, зокрема у контексті того, що використання ШІ значно оптимізовує процеси обробки та збору даних (Ені Браун) [8; 9]. Так само як і використання даних та ШІ медійниками у контексті російсько-української війни [10; 11]. Проблематику використання ШІ у даних також розглядають у розрізі етики та безпеки.

З огляду на окреслені дослідницькі контексти, запропоноване дослідження підкреслює важливість розуміння принципів використання ШІ у журналістиці даних, як і з огляду на тяглість процесу та можливості використання кращих практик в інших галузях медіа, так і з боку етичних і безпекових викликів у роботі. Урахування принципів захисту приватних даних та упередженості алгоритмів ШІ також є необхідним для мінімізації

ризиків маніпуляцій та порушення прав суб'єктів інформації й особливо актуалізується в умовах війни.

Постановка завдання. Метою статті є аналіз практики використання штучного інтелекту у журналістиці даних у контексті її потенціалу впливати на суспільну думку, сприяти взаємодії з соціальними інститутами та формувати довіру з аудиторією, а також можливість використання штучного інтелекту у поєднанні з аналізом даних як інструменту для розв'язання глобальних проблем. Для цього зокрема потрібно окреслити роль журналістики даних як форми суспільно значущої комунікації у цифрову епоху; дослідити, яким чином штучний інтелект інтегрується у практику журналістики даних; проаналізувати переваги та ризики використання ШІ у медіа загалом та журналістиці даних зокрема, у тому числі в умовах війни; визначити етичні та безпекові аспекти співпраці журналістів даних з алгоритмами ШІ; запропонувати підходи до відповідального й ефективного використання ШІ в журналістиці даних та медіа загалом.

Виклад основного матеріалу. Дані, Big Data та їхня роль у журналістиці. Через конвергентність, змішування жанрів та типів медіапродукції пересічні читачі (та й часто і самі журналісти) не постійно споживаючи той чи інший контент задумуються про якого класифікацію чи особливості створення. Серед багатьох трендів та тенденцій присутніх у медіасфері зараз (подкасти, короткі відео, переважання коротких форм у багатьох видах сторітелінгу) ми зосередимося на використанні даних. Зрештою, їх наскільки багато у кожній продукції журналістики (стаття про ДТП містить вулицю, кількість постраждалих; культурний оглядач аналізує дані про кількість та відвідуваність вистав, а карти повітряних тривог, якими користуються українці кожного дня, ні що інакше як класичний приклад журналістики даних: зібрати, проаналізувати й візуалізувати), що науковці й практики часто замість журналістики даних (на позначення медійної практики, що обробляє, аналізує та представляє дані у журналістських цілях) вживають термін «датафікація журналістики».

Зрештою, це доцільно, адже підходи й практики журналістики даних є важливим інструментом для будь-якого сучасного журналіста – для успішної діяльності треба знати, що таке дані, де їх знайти, як очистити та проаналізувати й подати аудиторії, а ще, що інколи навіть важливіше за попередні «хрестоматійні» кроки – звідки вони,

хто і яким чином їх збирав і які справді тенденції та явища вони ілюструють.

Для користувачів та журналістів, які не надто часто працюють з даними, вони часто видаються чимось масивним, масштабним і не до кінця зрозумілим, і обов'язково асоціюються з терміном *big data*. Якщо поглянути на це з термінологічного погляду, то розрізнення між так званими *big data* і *small data* не наскільки чітке. Наприклад ірландський науковець Роберт Кітчі наводить критерії для визначення «величини» даних, такі як: об'єм, вичерпність, ідентифікація, реляційність, швидкість приросту, різноманіття, гнучкість, розширюваність. За допомогою них він розрізняє так звані «small data» і «big data» [12, с. 46–67]. Дослідники також терміном *big data* часто називають не лише сам феномен наявності великих наборів інформації, які не можна проаналізувати традиційним способом. Хоча найлегше побачити відмінності у самій роботі журналістів даних.

Так, наприклад *Texty.org.ua*, українське видання, яке спеціалізується на журналістиці даних, підготувало матеріал «Вогняний вал. Як і які снаряди прилітають на українські позиції», де у тому числі завдяки даним від військовослужбовців та з інших джерел разом з автоматичними алгоритмами вони зробили графіку, яка ілюструє одну хвилину в окопі, по якому йдуть обстріли [13]. Попри те, що тут показано багато видів озброєння, це приклад використання *small data*. Є конкретний окоп. Матеріал побудований на конкретному запитанні: як «відчувається» хвилина у цій точці? з якою частотою і що «прилітає»? Подана конкретна візуалізація цього процесу

Small data можуть бути насправді досить великі за кількістю рядків і стовпчиків, але вони відповідають на конкретне питання, підтверджують (чи спростовують) гіпотезу, і їх можна помістити в електронні таблиці [14].

Натомість це саме видання також працювало над публікацією «Врожай з окупованих територій», де за допомогою ШІ (хоча більш правильно буде сказати машинного навчання) та аналізу супутникових знімків визначалися площі полів та межі культур, які на них вирощують [15]. Таким чином автори показали яких збитків завдала конкретно у цій галузі російська окупація, та яким чином вони продовжують використовувати поля і продавати крадений врожай (у тому числі якими шляхами). Тут уже журналісти мали справу з *big data* – вони показували взаємозв'язки й шукали причинно-наслідкові сполучення, виокремлювали тенденції та залежності.

Описаний вище поділ досить умовний, також не варто забувати про ще один термін, важливий для журналістики – відкриті дані (тобто такі, до яких є цілодобовий вільний доступ та подача у машиночитному вигляді). Об'єднує всі ці речі, окрім власне захованих всередині історій та можливостей «копнути глибше» для журналістів питання обробки, очищення та аналізу. І не завжди їх можна (чи доцільно) робити «вручну». Саме тут журналістам і допомагає автоматизація, у тому числі алгоритми машинного навчання і ШІ.

ШІ, медіа, журналістика даних. Штучний інтелект, його наявне та потенційне застосування нині є предметом міждисциплінарних досліджень – від прикладної математики до гуманітаристики. Відтак немає потреби доводити, що його інтеграція у сферу медіа є, з одного боку, процесом не новим, з іншого – незворотним, а також таким, що викликає зростаючий інтерес у науковій спільноті.

У цьому ж контексті варто зазначити, що у науковому середовищі досі не існує єдиного усталеного визначення поняття «штучний інтелект». Наприклад у черговому перевиданні однієї з фундаментальних праць на тему ШІ «Штучний інтелект: сучасний підхід» («Artificial Intelligence: A Modern Approach»), Стюарт Рассел та Пітер Норвіг розглядають штучний інтелект як галузь, що вивчає створення програмних або технічних агентів, здатних орієнтуватися в зовнішньому середовищі та діяти таким чином, щоб ефективно досягати визначених цілей [16, с. 4]. А у звіті Звіту JournalismAI (The Journalism AI Report, проєкт Journalism AI Лондонської школи економіки у співпраці з Google News Initiative) ШІ визначають як сукупність концепцій, технологій та методів, що дозволяють машинам виконувати завдання, які зазвичай потребують людського інтелекту [17]. По суті, це напрям комп'ютерних наук, спрямований на створення систем, здатних приймати рішення або розв'язувати проблеми самостійно. Попри те, варто пам'ятати, що зараз, зокрема у контексті впливу на медіа, спільнота розглядає здебільшого «обмежений» ШІ, той який функціонує лише в межах заздалегідь визначених алгоритмів і не здатен до самостійного узагальнення чи виходу за межі заданого сценарію.

Використання таких алгоритмів у медіа було доволі до появи чату GPT, Gemini чи інших застосунків, які тепер асоціюються з цією абревіатурою. Наприклад, як уже зазначалося, Франческо Марконі виокремлює три етапи інтеграції ШІ в журналістику: автоматизацію, аналітику

та генерацію [6]. Також існує уже більш-менш усталений перелік того, для чого медійники переважно використовують штучний інтелект: автоматизації (спрощення повторюваних процесів, створення шаблонів), аналізу текстів, персоналізації та роботи з аудиторією (боти, голосові помічники), фактчекінгу, обробки даних [18]. Попри те, що використання таких інструментів більш ресурсозатратне (зокрема у контексті часу та необхідних коштів), навіть у не надто сприятливих умовах його все одно використовують.

Так, скажімо, відповідно до дослідження, здійсненого у червні 2024 року Інститутом масової інформації (ІМІ) 22 % українських журналістів використовують ШІ для роботи на постійній основі (30 % – використовують час від часу, 20 % – взагалі не використовують). Що ще цікаво, переважна більшість (62 %) тих, хто відповіли ствердно на питання про постійне використання ШІ, задіюють його для спрощення рутинних завдань – розшифрування інтерв'ю (62 %), створення зображень (31 %), виправлення одруківки та інших хиб у текстах (21 %). Ще частина застосовує їх для генерування текстів (21 % – статті та інші текстові матеріали, 19 % – підбір заголовків та врізок), аналізу та обробки даних (14 %) [19].

Попри те, що у відповідях загалом згадувалося і використання ШІ для аналізу даних чи персоналізації контенту (на основі того ж аналізу), лише 14 % серед респондентів, які постійно користуються ШІ, вказали, що роблять це для аналізу та обробки даних. Що видається доволі невеликим показником. Однак у застосуванні ШІ у галузі журналістики даних є особливості. Зазвичай у медіа ШІ використовуються для рутинних завдань, а у журналістиці даних розкривається глибший потенціал.

Попри згадки у відповідях про використання ШІ для аналізу даних чи персоналізації контенту (зокрема на основі цих аналітичних процесів), показник у 14 % серед респондентів, які постійно користуються ШІ, вказує на порівняно обмежене застосування таких інструментів саме для роботи з даними. Водночас саме в журналістиці даних потенціал штучного інтелекту розкривається ширше, ніж у випадку його використання виключно для рутинних завдань, характерних для традиційного редакційного процесу.

Так у медіа загалом алгоритми переважно застосовуються для автоматизації процесів, які до цього надаються: генерування текстів, складання новин чи «звітів» за шаблонами, автоматична модерація чи персоналізація новинних стрічок.

Але це все працює лише з тими процесами, де варіативність готових результатів не настільки велика, щоб не можна було її прорахувати наперед. Наприклад можна автоматизувати щоденний прогноз погоди: абзац тексту завжди однаковий, автор залишає порожні місця для інформації, що змінюється (температура, ризик опадів, швидкість вітру тощо) і ШІ кожного дня «підтягує» ці нові дані у шаблон, створюючи готові новини для стрічки. А от репортаж чи інтерв'ю з метеорологом він не проведе (хоча може допомогти у формулюванні питань, особливо, якщо перед тим натренується на попередніх схожих інтерв'ю).

Журналістика даних зосереджена на аналізі великих обсягів інформації, очищенні даних, виявленні патернів, трендів чи аномалій, створення інтерактивних графіків і візуалізацій, що вже вимагає складних алгоритмів машинного навчання та обробки даних. Попри те, не можемо стверджувати, що в інших галузях медіа складні алгоритми не використовуються. Вони присутні, зазвичай, у контексті аналізу, персоналізації та журналістських розслідувань, просто для журналістики даних (зокрема розслідувальної) це більш пи́томо.

Окрім технологічно ємких алгоритмів у взаємодії ШІ та журналістики даних існує потреба й у глибокому критичному мисленні та поєднанні результатів роботи алгоритмів з людською експертизою. Такий підхід доречно переймати іншим медійним галузям.

Це не означає, що журналісти даних не використовують «звичайний/вже традиційний ШІ», який підкреслить граматичні помилки у тексті чи заповнить за зразком шаблон новини. Допоміжні алгоритми також мають місце у цій структурі. Медійники, що працюють з даними, є прикладом того, що якщо на постійній основі використовувати алгоритми – їх людську роботу ніхто не замінить. Якщо у традиційних редакціях є побоювання, що використання ШІ зведе їх роль з автора, до куратора чи редактора, то журналісти даних навпаки розвивають свої аналітичні навички водночас маючи змогу працювати з більшими й складнішими масивами інформації.

Етапи застосування ШІ у журналістиці даних накладаються зі звичайним розвитком цієї технології: спершу це автоматизація простих завдань, як уже напевно хрестоматійна алгоритмізація спочатку спортивних новин, а потім і роботизована стрічка в *Associated Press* [20]. З розвитком машинного навчання і великих мовних моделей (LLM), ШІ почав застосовуватися для складні-

шого аналізу даних і полегшив створення інтерактивних візуалізацій, що дало поштовх також і журналістиці даних. Розвиток великих мовних моделей в останні роки надало можливість глибокого аналізу великих масивів даних, користь з чого так само припала на журналістику даних.

Застосування готових і розробка власних моделей ШІ. Український контекст. Медіа переважно використовують готові інструменти ШІ для роботи, і в журналістиці даних така тенденція також спостерігається.

Так, наприклад застосунки на кшталт Tableau або Google Data Studio, інтегровані з ШІ, використовуються для побудови інтерактивних графіків і карт [21]. Більшість інструментів для очищення даних (Google Tablets, Microsoft Exel, Open Refine) також мають «підказки» та функції на основі ШІ, які полегшують роботу журналістам.

Також часто редакції вдаються до послуг компаній, що займаються технічною розробкою чи аналітикою даних у своїх сферах, щоб використати це у розробці. Так, наприклад у згаданій публікації «Врожай з окупованих територій» українська команда дата журналістів використала дані на основі супутникових знімків та машинного навчання агротехнологічної компанії OneSoil, щоб визначити меж полів і культур, які на них вирощують та на цій підставі оцінювали обсяг збитків нанесених російською окупацією [22]. А журналісти Bellingcat співпрацюють з компаніями, що спеціалізуються на обробці супутникових знімків і AI-аналітиці (Planet Labs, Maxar Technologies), отримуючи від них готові аналітичні матеріали для розслідувань [23]. Такими послугами від різних сторонніх компаній користуються великі закордонні видання, як *The New York Times* [24].

Натомість деякі медіа мають достатньо ресурсів, щоб створити свої ресурси. Наприклад *Reuters* створив платформу Lynx Insight, яка використовує машинне навчання для аналізу великих масивів даних і допомагає журналістам швидко ідентифікувати важливі теми та аномалії [25]. Ця система поєднує ШІ-аналітику з людським контролем. *Bloomberg* розробив власну мовну модель штучного інтелекту BloombergGPT, спеціально адаптовану для фінансової сфери, яка навчена на широкому спектрі фінансових даних, включно з фінансовими документами, новинами, прес релізами, а також архівами *Bloomberg* і загальнодоступними джерелами та покращує завдання обробки природної мови (NLP) у фінансовій журналістиці, такі як аналіз настроїв, розпізнавання іменованих об'єктів, класифікація текстів та

автоматичне створення описів фінансових подій. Зокрема допомагає журналістам даних швидко і точно аналізувати великі обсяги фінансової інформації, що суттєво підвищує якість і оперативність новин та аналітичних матеріалів [26].

Уся світова медіасфера розвиває ШІ, думаючи про зручність та нові можливості. Можна використовувати дуже простий інструмент на основі нейромереж – це звільнить час журналістів для важливіших досліджень. Можна застосовувати складні алгоритми для аналізу великих масивів даних – це допоможе у розслідуваннях чи інших матеріалах. Проте в українському просторі є своя специфіка. Якщо з простими інструментами усе зрозуміло – дослідження ІМІ показує, що для рутинних завдань його використовують – то складніші алгоритми переважно застосовуються в інших темах – для покращення озброєння, наведення, аналізу відео з дронів чи роботи оповіщення повітряних тривог [27].

Журналістика даних загалом видозмінилася під час війни, це включає взаємодію держави з громадськістю через інструменти, такі як чат-боти для повідомлень про повітряні тривоги, збір даних про вибухонебезпечні предмети чи пошкодження житла, а також офіційні бази даних, такі як списки укриттів або реєстри російських військових, що відповідальні за воєнні злочини. Журналістські зусилля також зосереджуються на тому, як ідентифікувати озброєння, відстежувати переміщення військ та доводити воєнні злочини, використовуючи супутникові знімки, дані про масові поховання та викриття фейкових новин у соціальних мережах [28].

Наприклад, *Bellingcat*, міжнародна група незалежних журналістів, використовує OSINT (відкриті джерела інформації) для документування шкоди, завданої цивільним в Україні, такої як жертви, руйнування інфраструктури, а також для спростування тверджень про фальсифікацію російських воєнних злочинів у Бучі та Краматорську, використовуючи онлайн-фотографії, відео, новинні звіти та супутникові знімки. *The New York Times* використовував супутникові знімки компанії Махаг Technologies для перевірки доказів російських злочинів у Бучі. Так само команда «Схеми» *Radio Свобода* виявила четверту масову могилу в окупованому Маріуполі, використовуючи знімки Махаг.

Ще одна велика сфера – картографія. Вона включає як карти повітряних тривог, так і (наприклад проєкт *Deep State Live*) – інтерактивну онлайн-карту військових операцій в Україні, яка

дозволяє користувачам відстежувати зміни на фронті та хід російсько-української війни [29].

До перерахованих форм взаємодій з даними вже інтегрований штучний інтелект, наприклад, для аналізу супутникових знімків чи виявлення загроз. Ще одна ключова сфера, у якій він застосовується у контексті журналістики даних, – дезінформація.

Факультет журналістики Львівського національного університету імені Івана Франка співпрацював з факультетом електроніки над проєктом з виявлення російської дезінформації. Проєкт полягав у розробці системи для автоматичного аналізу медіа та соціальних мереж з метою виявлення джерел фейкових новин, шкідливих даних, агресивної риторики та закликів до несанкціонованих дій. Журналісти моніторили та перевіряли новини українською та російською мовами з різних платформ. Новини категоризувалися за надійністю (Фейк, Можливо правда, Правда) та емоційним тоном (Агресивний, Нейтральний, Позитивний), також була категорія «Маніпулятивний». Було проаналізовано понад 2000 україномовних та таку ж кількість російськомовних повідомлень. Ось узагальнені висновки: російська пропаганда використовує індивідуалізовані повідомлення, з відмінностями в тональності на різних платформах соціальних мереж та новинних ресурсах. Після вторгнення фейкові новини на українських сайтах зменшились, поступившись витонченішими маніпуляціями.

Цей проєкт з розробки ШІ-алгоритму на основі аналізу даних наразі призупинено через технічні затримки та фінансові труднощі, однак схожі проєкти допомагають учасникам та оснащують громадян для боротьби з пропагандою. Водночас в Україні існують ініціативи, що вже ефективно застосовують штучний інтелект для протидії російській дезінформації. Серед них – стартапи *Osavul* та *Mantis Analytics*, створені в перші дні повномасштабного вторгнення [30; 31]. Їхня діяльність полягає у виявленні кампаній дезінформації за допомогою великих мовних моделей (LLM) та технологій обробки природної мови (NLP). Ці системи аналізують великі обсяги цифрового контенту та ідентифікують повторювані наративи, що дозволяє виявляти інформаційні атаки ще на ранньому етапі та оперативно реагувати на них, поширюючи перевірену інформацію. Зокрема, *Osavul* здатен не лише ідентифікувати фейки, а й визначати канали поширення пропагандистських повідомлень.

Обидва проєкти також співпрацюють з державними структурами, у тому числі з Радою наці-

ональної безпеки та оборони України, а також пропонують свої інструменти з моніторингу та аналізу дезінформації за допомогою ШІ для медіа, бізнесу й освітніх організацій.

Поруч з такими великими стартапами, існують і локальні – як, наприклад, описаний вище університетський проєкт або проєкт «Деза війни» від *Texty.org.ua* [32]. Журналісти ще з 2019 року застосовували нейромережі для збору повідомлень, їх кластеризації та аналізу у контексті боротьби з російською дезінформацією, моніторячи 72 російські вебсайти, що фокусуються на українських темах. З 2022 року проєкт отримав назву «Деза війни» Журналісти зокрема використовували автоматичне моделювання тем (BERTopic) для групування абзаців новин за темами. Неактуальні теми (наприклад, COVID-19, спорт, погода, попірки, нещасні випадки) були виключені, а теми, пов'язані з війною та її впливом на Росію, були виокремлені. Кожна стаття могла містити кілька тем.

Ризики, закони й етика. ШІ й журналістика даних взаємопов'язані, що дає унікальну можливість для журналістів використовувати складні алгоритми, а для розробників нейромереж мати базу для навчання та тренування. Зі збільшенням переваг, однак підвищується і ризик використання та впровадження таких алгоритмів.

Зрозуміло, є загальні побоювання щодо використання ШІ. Вчені Медіалабораторії Массачусетського технологічного інституту припускають, що використання методів навчання з обмеженим доступом (LLM) може фактично зашкодити навчанню, особливо для молодших користувачів [33]. Вибірка була невелика, про те вона показала, що ті респонденти, які використовували чат GPT, мали схильність до зниження когнітивних навичок протягом експерименту. Суть дослідження не у тому, що показати який чат GPT «поганий», а у підкресленні шкоди саме від використання генеративних нейромереж у навчанні. Це цікаво ще і в контексті того, що компанія Open AI сама випускала рекомендації про використання генеративного штучного інтелекту у навчанні.

Так само в українському контексті тривожним і важливим був звіт аналітичної групи News Guard, у якому досліджувалося, як російська пропагандистська машина тренувала алгоритми ШІ на своєму матеріалі, що призвело до видачі прокремлівської інформації в основних чат-ботах (повторення російських наративів у 33 % випадків) [34]. Враховуючи як багато людей використовують той самий чат GPT як пошуковик, це дуже важливо –

у таких генеративних моделей немає «світогляду» чи понять «добрих» і «поганих» сторі. Якої інформації було більше у процесі їх навчання – ту вони й генерують. А протидії таким російським спробам «перепрограмувати» популярні боти на свій лад поки не було.

Є звичні загальні етичні та безпекові виклики, наприклад підтримка довіри читачів – необхідне чітке маркування у яких матеріалах і яким чином використовувався ШІ. Генерування текстів, на відміну від аналізу інформації та виділення головного, часто потребує ретельного фактчекінгу. Це пов'язано з поширеним явищем «галюцинацій», коли ШІ видає правдоподібні, але фактично невідповідні відповіді.

Серед інших значних проблем – авторське право, коли виникають складнощі з визначенням авторства згенерованих зображень чи текстів.

Ще одна проблема – витік конфіденційних даних. Існує ризик несанкціонованого розголошення інформації, коли під час нарад використовують алгоритми для запису та виокремлення важливих моментів, або ж завантажують конфіденційні звіти у програми. Зокрема, Reuters дедалі обґрунтованіше пише про передачу даних користувачів та аналітики розвідувальним та військовим структурам КНР розробниками чат-бота DeepSeek.

Також можливе використання неякісного контенту. Існує спокуса видавати згенеровану ШІ роботу низької якості за власну, авторську.

Водночас журналістика даних стикається також з унікальними викликами, пов'язаними з масштабом і складністю роботи з даними, прозорістю алгоритмів, захистом приватності та відповідальністю за результати. Зокрема необхідна особлива увага до прозорості алгоритмів, якості даних і методів верифікації, оскільки помилки у даних можуть призвести до поширення дезінформації та хибних упереджень. Це зокрема аналізує Кеті О'Ніл у своїй праці «Big data: Зброя математичного знищення», показуючи через аналіз алгоритмів оцінки кредитоспроможності, які використовують непрямі показники (наприклад, поштовий індекс), як він дискримінує мешканців бідних районів. Така структура створює замкнуте коло, де люди з низьким рейтингом не отримують кредитів, мають через це нижчу платоспроможність, гіршу фінансову історію, яка знову не дає їм шансу на позику, і повторення цих кроків поглиблює їхню фінансову виключеність [35, с. 150–160]. Подібна логіка діяла й у випадку з алгоритмічними системами патрулювання, такими як PredPol: алго-

ритм визначав «зони підвищеної злочинності» на основі даних, спрямовуючи туди більше поліції. У результаті зростала кількість арештів саме у цих районах, що лише посилювало уявлення про їхню «небезпечність» і знову приводило до більшої концентрації правоохоронців. Такі приклади ілюструють небезпеку самопідсилюваних систем, якщо їх не супроводжує критичний аналіз і ретельна верифікація даних.

Тому якщо в інших галузях медіавіклик більше пов'язані з автоматизацією контенту, ризиком маніпуляції та зниженням якості медіапродукту, то у журналістиці даних усі вони залишаються, але особлива увага надається прозорості та аналізу роботи самого алгоритму. Якщо він навчається на некоректних чи неповних даних, результати також будуть неправильні. Наприклад системи автопілота, зокрема Tesla, можуть некоректно розпізнавати об'єкти, які не були достатньо представлені в навчальних даних, що створює потенційні ризики для безпеки пішоходів і породило навіть своєрідний мем про те, як ці автомобілі становили більшу загрозу для жінок із візочками, людей старшого віку та різної расової приналежності.

Так само алгоритми для аналізу даних (наприклад, кластеризації соціальних мереж) часто є «чорними скриньками». Журналісти не можуть пояснити, як система прийшла до висновку. А погана анонімізація даних часто може призвести до ідентифікації осіб із наборів «анонімних» даних.

Журналісти також обізнані з цими проблемами. Так перевірка джерел і додавання «шуму» до даних допомагає попередити деанонімізацію чи неправильні висновки. Для убезпечення від незрозумілої роботи алгоритмів використовуються інструменти з відкритим кодом (наприклад, Python-бібліотеки *scikit-learn*). Також допомагає практика публікування усієї методології роботи над проектом, як і наборів даних, які використовувалися у матеріалі (так робить, наприклад *Bellingcat*, *Texty.org.ua* і більшість якісних видань, що спеціалізуються на журналістиці даних).

Окрім цього є і загальні етичні та законодавчі регулятори, хоч і не вповні впроваджені. Так набув чинності у 2024-му і впроваджується у 2025-му Закон ЄС про штучний інтелект (AI Act), який ділить алгоритми ШІ відповідно до ризику застосування. Згідно з цими нормами високоризикові застосування ШІ (біометрія, прийняття рішень у кредитах, освіті) підлягають суворому контролю, реєстрації й дотриманню вимог з управління

ризиками; а у випадку користування алгоритмами, які переважно застосовують у медіа (чатботи, генерація текстів) зазначена необхідність дотримуватися підзвітності та прозорості, маркування [36]. OpenAI та інші великі компанії співпрацюють з ЄС для впровадження цього регламенту.

В Україні запроваджена концепція розвитку штучного інтелекту та затверджений План заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2025–2026 роки [37]. Рекомендації щодо впровадження штучного інтелекту також надають Міністерства освіти й науки України, Мінцифри, Мінекономіки, Експертно-консультаційний комітет з розвитку штучного інтелекту при Мінцифрі, Центр демократії та верховенства права (ЦЕДЕМ), НДІ Інтелектуальної власності Національної академії правових наук України, навчальні заклади [38; 39].

Для медійників особливо цікавими у цьому контексті також є Рекомендації з відповідального використання штучного інтелекту у сфері медіа, розроблені Мінцифрою разом з партнерами та профільними експертами у рамках реалізації дорожньої карти регулювання ШІ в Україні [40]. Вони мають на меті посприяти етичному та відповідальному використанню штучного інтелекту в журналістиці, забезпечуючи дотримання прав людини, демократії та верховенства права. Основними її принципами є прозорість (чітко маркувати коли й де саме використовується який конкретний ШІ); відповідальність (встановлювати механізми для притягнення медіаорганізацій до відповідальності за хибні факти чи твердження, що ґрунтуються на ШІ й подаються медіа як правильні та аргументовані); інклюзивність (перевіряти ШІ у контексті різноманітності контенту та унеможливлення упередженості); точність (використання ШІ для покращення перевірки фактів і запобігання поширенню дезінформації).

Висновки. Журналістика даних є однією з найпитоміших сфер застосування ШІ у медіа, оскільки вона поєднує великі обсяги даних, складний аналіз і потребу у глибокому критичному мисленні. На відміну від інших галузей, де ШІ часто використовується для автоматизації рутинних завдань, у журналістиці даних ШІ є інструментом, що розширює аналітичні можливості журналістів і підвищує якість інформації.

В українському медіапросторі штучний інтелект у журналістиці даних також активно використовується для боротьби з російською дезінформацією. Присутні є досить чіткі й потрібні рекомендації розроблені урядовими діячами. Це

з одного боку. З іншого – журналістика даних також більш чутлива до «підводних каменів» ШІ, зокрема помилок в алгоритмах, упереджених даних, технологій без чіткого прозорого механізму дії. Окрім того, в умовах війни більшість саме технологічних розробок зосереджені у галузі оборони. І попри наявність багатьох рекомендацій законодавство все ще не чітко ні в контексті авторського права, ні у рамках дозволеного використання ШІ, ані у зв'язку з притягненням відповідальності через помилки алгоритмів (введення в оману читачів – це вина розробників чи все-таки автора, який скористався нейромережею?). Поруч з доцільними законодавчими ініціативами українська медіасфера також має досвід обмежити доступ до даних про публічні закупівлі та іншої інформації під виглядом захисту чутливих даних; чи спроба криміналізувати журналістів, які поширюють інформацію з публічних реєстрів, зокрема дані про використання коштів, що поставило під загрозу журналістів-розслідувачів, які працюють з відкритими даними. Те ж саме цілком ймовірно може відбуватися у сфері регулювання використання медіа ШІ – або недоцільні обмеження, або використання цих алгоритмів як механізму накопичення несанкціонованих персональних даних.

Отже, варто використовувати потенціал ШІ у медіа, зокрема в журналістиці даних. При цьому слід враховувати як європейське, так і вітчизняне законодавство та рекомендації, постійно моніторити та реагувати на зміни, які уряд може ввести й реагувати на їх доцільність. Також доцільним буде впровадження власних регулятивних механізмів на рівні редакцій (включення положень про використання ШІ до етичних кодексів та визначення редакційних політик у цьому контексті).

Така діяльність неможлива без відповідних навичок і розуміння самого механізму роботи алгоритму, адже через те, що багато хто сприймає генерування від великих мовних моделей за надання відповіді на основі моніторингу і не проводить додатковий фактчек, і стаються багато найбанальніших помилок. Те саме відбувається і на глибших етапах – з алгоритмами, які підтверджують упередження.

Не варто також забувати про контекст медіасередовища. Так в умовах повномасштабного вторгнення розвивати ШІ для того, щоб моніторити тренди для музичних медіа буде не зовсім доречно, натомість використовувати алгоритми машинного навчання, нейромережі й дані для боротьби з російською пропагандою – і доцільніше, і ефективніше.

Список літератури:

1. Chambers L., Gray J., Bounegru L. Data journalism handbook. O'Reilly Media, Incorporated, 2012, 242 p. URL: <https://datajournalism.com/read/handbook/one> (date of access: 31.10.2025)..
2. Gray J., Bounegru L. Data journalism handbook: towards a critical data practice. European Journalism Centre, 2019. 270 p. URL: <https://datajournalism.com/read/handbook/two>.
3. Кубай, Д., Горбаль, А., Газін, А., Шульга, Є. та Шаповаленко, Г. (2016), Відкритий посібник з відкритих даних для громадських організацій, журналістів, і не тільки, Київ : Український центр суспільних даних, URL: <https://socialdata.org.ua/manual/> (дата перегляду 24.09.2025).
4. Мерабет, Д. (2024), Як журналістика даних допомагає висвітлювати відновлення, Центр практичної освіти «Накипіло», URL: <https://surl.li/eqzpgk> (дата перегляду 20.06.2025).
5. Sickel A. V., Au E. Data journalism top 10: A year of war in ukraine, russian casualties, trans identities, unexpected NBA performances. *Global investigative journalism network*. 2023. URL: <https://surl.li/pwmxur> (date of access: 20.06.2025).
6. Marina Adami. Is ChatGPT a threat or an opportunity for journalism? Five AI experts weigh in. Reuters Institute for the Study of Journalism. *Reuters Institute*. 2023 URL: <https://surl.li/cc/tskagj> (date of access: 20.06.2025),
7. Muhammed Enes Calli. What impact could ChatGPT have on journalism. *Anadolu Ajansı*. 2023. URL: <https://surl.li/cc/pvyurx> (date of access: 20.06.2025).
8. Annie Brown. Utilizing AI And Big Data To Reduce Costs And Increase Profits In Departments Across An Organization. *Forbes*. 2021. URL: <https://surl.li/exohbx> (date of access: 20.06.2025).
9. How Do Big Data and AI Work Together? *QlikTech International AB*. 2023. URL: <https://www.qlik.com/us/augmented-analytics/big-data-ai> (date of access: 20.06.2025).
10. Мерабет Д. Як журналістика даних допомагає висвітлювати відновлення. *Центр практичної освіти «Накипіло»*. 2024. URL: <https://surl.li/fouxsj> (дата звернення: 21.06.2024).
11. Sickel A. V., Au E. Data journalism top 10: A year of war in ukraine, russian casualties, trans identities, unexpected NBA performances. *Global investigative journalism network*. 2023. URL: <https://surl.li/oujnhh> (date of access: 08.06.2025).

12. Kitchin R. Data revolution: big data, open data, data infrastructures and their consequences. SAGE Publications, Limited, 2014. 240 p.
13. Вогняний вал. Як і які снаряди прилітають на українські позиції / Є. Дроздова та ін. *Texty.org.ua*. URL: <https://surl.lu/qtqwts> (дата звернення: 28.05.2025).
14. Ліченко А. Д. Варіативність тлумачення терміна «журналістика даних» у сучасних медіадослідженнях. *Наукові записки Таврійського національного університету ім. В. І. Вернадського. Серія: Філологія. Соціальні комунікації*. 2020. Т. 4, № 2. С. 137–143. URL: <https://doi.org/10.32838/2663-6069/2020.2-4/21> (дата звернення: 25.06.2025).
15. Дроздова Є., Кельм Н. Врожай з окупованих територій. Скільки Росія заробляє на українській землі. *Texty.org.ua*. URL: <https://texty.org.ua/projects/111036/> (дата звернення: 04.06.2025).
16. Russell S. J., Norvig P. Artificial intelligence: a modern approach. Pearson Education, Limited, 2010. 1151 p.
17. Charlie Beckett. New powers, new responsibilities. A global survey of journalism and artificial intelligence. *LSE*. 2019. URL: <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/> (дата звернення: 04.06.2025).
18. Штучний інтелект у медіа: програми пишуть новини, але не замінюють журналістів. *MEDIASAPIENS*. 2021. URL: <https://surl.lt/axzwaw> (дата звернення: 04.06.2025).
19. Машкова Я. Опитування ІМІ: 22 % українських редакцій використовують штучний інтелект на постійній основі. *Інститут масової інформації*. URL: <https://surli.cc/xsoahb> (дата звернення: 01.06.2025).
20. Дмитрик Х. Робот писатиме новини для Associated Press. *ipress.ua*. URL: <https://surl.lt/wgzdqq> (дата звернення: 28.05.2025).
21. AI in Tableau. *Online Help | Tableau Software*. URL: https://help.tableau.com/current/tableau/en-gb/about_tableau_gai.htm (date of access: 01.06.2025).
22. Дроздова Є., Кельм Н. Врожай з окупованих територій. Скільки Росія заробляє на українській землі. *Texty.org.ua*. URL: <https://texty.org.ua/projects/111036/> (дата звернення: 04.06.2025).
23. Bellingcat can now access specialised satellite imagery. Tell us where we should look. *Bellingcat*. URL: <https://surl.li/vzuvd> (date of access: 30.05.2025).
24. Philp R. Reporter's tipsheet for acquiring free satellite images. *Global Investigative Journalism Network*. URL: <https://surl.li/teklss> (date of access: 03.06.2025).
25. Albeanu C. New automated tool at Reuters helps its 'cybernetic newsrooms' find stories in data. *Journalism jobs (media, editorial), news for journalists*. URL: <https://surl.li/zuhvdo> (date of access: 12.06.2025).
26. Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance. *Bloomberg*. URL: <https://surl.li/tguyvb> (date of access: 11.06.2025).
27. Дар'я Кустанович. The NY Times про американські стартапи, які можуть допомогти Україні у війні. *ГЛАВКОМ*. URL: <https://surl.li/cgikxa> (дата звернення: 18.06.2025).
28. Шекеряк Я. Карти, боти й пошук людей – як журналісти й урядові структури використовують відкриті дані під час воєнного стану. *Медіакритика*. 2022. URL: <https://surl.li/bvwffx> (дата звернення: 16.05.2025).
29. DeepStateMAP.live. *DeepStateMap*. URL: <https://deepstatemap.live/#6/49.4383200/32.0526800> (date of access: 17.06.2025).
30. Mantis Analytics. URL: <https://election.mantisanalytics.com/2/> (дата звернення: 02.03.2025).
31. Russian disinformation detection and fake news detection. *Osavul: AI-Powered Security Against Information Threats*. URL: <https://www.osavul.cloud/russian-disinformation-detection> (date of access: 03.03.2025).
32. Деца війни. *Texty.org.ua*. URL: <https://topic-radar.texty.org.ua/#/> (дата звернення: 20.02.2025).
33. Chow A. R. ChatGPT's impact on our brains according to an MIT study. *TIME*. URL: <https://time.com/7295195/ai-chatgpt-google-learning-school/> (date of access: 03.06.2025).
34. NewsGuard. A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda. *NewsGuard's Reality Check*. URL: <https://surl.li/qstfpo> (date of access: 09.03.2025).
35. О'Ніл, К. Зброя математичного знищення: як великі дані збільшують нерівність і загрожують демократії / Кеті О'Ніл ; пер. з англ. – Лондон : Penguin, 2017. – 259 с.
36. З 2 лютого 2025 року діють статті 1-5 Закону про штучний інтелект, невиконання яких може призвести до великих штрафів. *Kancelaria LBKP Legal*. URL: <https://surli.cc/unucue> (дата звернення: 18.06.2025).
37. Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2025–2026 роки. *Офіційний вебпортал парламенту України*. URL: <https://surl.li/hcbpju> (дата звернення: 23.06.2025).
38. Штучний інтелект у закладах вищої освіти: рекомендації для викладачів, студентів і працівників ЗВО. *Міністерство освіти науки*. URL: <https://surl.lt/mdvfft> (дата звернення: 19.06.2025).

39. Як використовувати штучний інтелект без порушення ІР права: рекомендації команди експертів. *Укрпатент*. URL: <https://ukrpatent.org/uk/news/main/shi-bez-porushennia-ip-06112024> (дата звернення: 23.06.2025).

40. Як відповідально використовувати штучний інтелект: розробили рекомендації для медіа. *Міністерство цифрової трансформації України*. URL: <https://surl.li/qfhrpk> (дата звернення: 23.06.2025).

Shekeriak-Kushka Ya. V. USE OF ARTIFICIAL INTELLIGENCE IN THE WORK OF DATA JOURNALISTS

This study aims to explore how elements of artificial intelligence (AI) are used in journalistic work. The paper examines how professional media practices are being transformed by AI technologies and analyzes the specific forms in which AI is integrated into journalism, from automating routine tasks to advanced data analysis and content generation. The growing role of AI in media in general is considered, with a focus on the aspects of journalistic activity related to data. Data journalism is defined as a media practice in which the potential of AI is most fully realized. Special attention is given to the development of these processes within the Ukrainian media landscape during the full-scale war.

The article analyzes how journalists use existing AI tools and develop their own solutions. It presents examples of the use of AI in open-source intelligence (OSINT) projects, satellite image analysis, identification of war criminals, and combating disinformation. Cases of effective AI usage by both Ukrainian and international media outlets are discussed. It is shown that Ukrainian data journalists have experience with both generative and analytical AI applications, including trend identification, data analysis, and visualization.

The study outlines ethical challenges associated with the integration of AI into journalistic work, such as transparency, audience manipulation, authorship concerns, hallucinations, and accuracy. It also highlights the specific security risks of using AI in data journalism during wartime, including source protection, anonymization of sensitive data, and minimizing risks to civilians. The key benefits and drawbacks of AI in journalism are identified, and core problems and potential solutions are outlined. Successful case studies of AI use by data journalists are analyzed, and based on these, recommendations are offered for ethical and secure AI integration into journalism (including transparency, fact-checking, content labeling, methodological openness, and privacy protection).

Key words: data journalism, artificial intelligence, AI, personal data, algorithms, ethics, security, war, disinformation, Ukraine, new media, media.