

3. Fowler M. Microservices: a definition of this new architectural term. 2014. URL: <https://martinfowler.com/articles/microservices.html>.
4. Richardson C. Microservices Patterns: With examples in Java. Manning Publications, 2018. 520 p.
5. Dragoni N., Giallorenzo S., Lafuente A. L. [et al.] Microservices: Yesterday, Today, and Tomorrow. **Present and Ulterior Software Engineering**. Springer, Cham, 2017. P. 195–216.
6. Kalske M., Mäkitalo N., Mikkonen T. Adopting Microservices: A Systematic Mapping Study. **International Conference on Web Engineering**. Springer, 2017. P. 248–263.
7. Apache JMeter. User's Guide: Building a Load Test Plan. 2023. URL: <https://jmeter.apache.org/usermanual/>.
8. PostgreSQL Global Development Group. PostgreSQL 15.0 Documentation. 2022. URL: <https://www.postgresql.org/docs/15/index.html>.
9. Indrasiri K., Kuruppu P. gRPC: Up and Running: High Performance RPC Framework for Pervasive Computing. O'Reilly Media, 2020. 210 p.
10. React Documentation. Managing State in Distributed Web Applications. 2024. URL: <https://react.dev/learn>.

References:

1. Microsoft Learn. .NET Microservices. Architecture for Containerized .NET Applications : Microsoft Guide. 2023. URL: <https://learn.microsoft.com/en-us/dotnet/architecture/microservices/>.
2. Newman S. Building Microservices: Designing Fine-Grained Systems. 2nd Edition. O'Reilly Media, 2021. 614 p.
3. Fowler M. Microservices: a definition of this new architectural term. 2014. URL: <https://martinfowler.com/articles/microservices.html>.
4. Richardson C. Microservices Patterns: With examples in Java. Manning Publications, 2018. 520 p.
5. Dragoni N., Giallorenzo S., Lafuente A. L. [et al.] Microservices: Yesterday, Today, and Tomorrow. **Present and Ulterior Software Engineering**. Springer, Cham, 2017. P. 195–216.
6. Kalske M., Mäkitalo N., Mikkonen T. Adopting Microservices: A Systematic Mapping Study. **International Conference on Web Engineering**. Springer, 2017. P. 248–263.
7. Apache JMeter. User's Guide: Building a Load Test Plan. 2023. URL: <https://jmeter.apache.org/usermanual/>.
8. PostgreSQL Global Development Group. PostgreSQL 15.0 Documentation. 2022. URL: <https://www.postgresql.org/docs/15/index.html>.
9. Indrasiri K., Kuruppu P. gRPC: Up and Running: High Performance RPC Framework for Pervasive Computing. O'Reilly Media, 2020. 210 p.
10. React Documentation. Managing State in Distributed Web Applications. 2024. URL: <https://react.dev/learn>.

Дата першого надходження статті до видання: 03.05.2026

Дата прийняття статті до друку після рецензування: 19.05.2026

ISSN 2786-6025 Online

УДК 009:004.8:004.65

[https://doi.org/10.52058/2786-6025-2026-5\(59\)-5572-5587](https://doi.org/10.52058/2786-6025-2026-5(59)-5572-5587)

Побіженко Ірина Олександрівна кандидат технічних наук, доцент, доцент кафедри цифрових комунікацій та інформаційних технологій, Харківська державна академія культури, м. Харків, доцент кафедри програмної інженерії, Харківський національний університет радіоелектроніки, м. Харків, <https://orcid.org/0000-0002-0723-1878>

Брусенцев Віталій Олександрович кандидат технічних наук, доцент, доцент кафедри цифрових комунікацій та інформаційних технологій, Харківська державна академія культури, м. Харків, <https://orcid.org/0000-0002-0020-4025>

НОВІ ПІДХОДИ ДО ГУМАНІТАРНИХ ДОСЛІДЖЕНЬ ІЗ ЗАСТОСУВАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ ТА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Анотація. У роботі розглянуто актуальні методиками гуманітарних досліджень на базі штучного інтелекту (ШІ) та Data Mining у межах еволюції Digital Humanities. Досліджено специфіку впровадження інструментів ШІ, машинного навчання, комп'ютерної лінгвістики (NLP), текстового аналізу (Text Mining) та великих даних (Big Data) у соціогуманітарний простір.

Акцент зроблено на міждисциплінарну взаємодію спеціальностей В13 (029) «Бібліотечна, інформаційна та архівна справа» та F6 (126) «Інформаційні системи та технології». Визначено, що вектор розвитку для фахівців галузі В13 (029) полягає в цифровізації архівів і бібліотек, залученні OCR-інструментів, а також упровадженні смарт-каталогізації, семантичного пошуку та інформаційно-аналітичного супроводу користувачів. Водночас для спеціальності F6 (126) пріоритетом є проектування аналітичних платформ, систем прийняття рішень та алгоритмів машинного навчання, адаптованих для опрацювання гуманітарних даних.

У межах дослідження висвітлено ключові інструменти Data Mining у соціогуманітарному середовищі, зокрема Text Mining, NLP, машинне навчання, семантичний моніторинг та методи інфографіки. Окреслено прикладний потенціал штучного інтелекту під час атрибуції історичних пам'яток, дослідження інформаційного поля, соціальних зв'язків та протидії маніпулятивним медіавпливам. Окрему увагу приділено використанню генеративного штучного інтелекту у науковій та освітній діяльності.

У результаті дослідження зроблено висновок, що впровадження штучного інтелекту та інструментів Data Mining у соціогуманітарний простір виступає рушійною силою для розширення меж Digital Humanities. Цей процес модернізує інформаційно-аналітичні практики та закладає підґрунтя для створення інноваційних міждисциплінарних моделей наукових досліджень у цифровому суспільстві.

Ключові слова: штучний інтелект, цифрова гуманітаристика, Digital Humanities, інтелектуальний аналіз даних, Data Mining, машинне навчання, обробка природної мови, NLP, Big Data, генеративний штучний інтелект, інформаційно-аналітична діяльність, цифрові архіви, електронні бібліотеки.

Pobizhenko Iryna PhD in Technical Sciences, Associate Professor, Associate Professor at the Department of Digital Communications and Information Technologies, Kharkiv State Academy of Culture, Kharkiv, Associate Professor at the Department of Software Engineering, Kharkiv National University of Radioelectronics, Kharkiv, (<https://orcid.org/0000-0002-0723-1878>)

Brusentsev Vitalii PhD in Technical Sciences, Associate Professor, Associate Professor at the Department of Digital Communications and Information Technologies, Kharkiv State Academy of Culture, Kharkiv, (<https://orcid.org/0000-0002-0020-4025>)

NEW APPROACHES TO HUMANITIES RESEARCH USING ARTIFICIAL INTELLIGENCE AND INTELLIGENT DATA ANALYSIS

Abstract. This paper examines current methodologies in humanities research based on artificial intelligence (AI) and Data Mining within the context of the evolution of the Digital Humanities. The specifics of implementing AI tools, machine learning, natural language processing (NLP), Text Mining, and Big Data in the social sciences and humanities are examined.

The focus is on interdisciplinary collaboration between the fields of study B13 (029) “Library, Information, and Archival Studies” and F6 (126) “Information Systems and Technologies.” It has been determined that the development path for specialists in the B13 (029) field lies in the digitization of archives and libraries, the use of OCR tools, as well as the implementation of smart cataloging, semantic search, and information-analytical support for users. At the same time, for the F6 (126) specialty, the priority is the design of analytical platforms, decision-making systems, and machine learning algorithms adapted for processing humanities data.

This study highlights key Data Mining tools in the social sciences and humanities, including Text Mining, NLP, machine learning, semantic monitoring,

ISSN 2786-6025 Online

and infographics. It outlines the practical potential of artificial intelligence in the attribution of historical monuments, the study of the information field and social connections, and countering manipulative media influences. Particular attention is paid to the use of generative artificial intelligence in scientific and educational activities.

The study concluded that the integration of artificial intelligence and Data Mining tools into the social sciences and humanities serves as a driving force for expanding the boundaries of the Digital Humanities. This process modernizes information and analytical practices and lays the groundwork for the creation of innovative interdisciplinary models of scientific research in the digital society.

Keywords: artificial intelligence, digital humanities, Data Mining, machine learning, natural language processing (NLP), Big Data, generative artificial intelligence, information and analytical activities, digital archives, electronic libraries.

Постановка проблеми. Еволюція сучасного інформаційного соціуму характеризується експоненційним накопиченням цифрового контенту, глобальною цифровізацією соціогуманітарного простору та масовою інтеграцією інструментів штучного інтелекту в освітньо-наукове середовище.

Функціонування електронних бібліотечних фондів, віртуальних архівів, мас-медійних платформ і соцмереж генерує масштабні масиви гетерогенних даних, що потребують прогресивних інструментів для їхнього структурування та контент-аналізу. Оскільки класичний гуманітарний інструментарій уже не здатний оперативно й релевантно опрацювати такі обсяги інформації, виникає гостра потреба в залученні методів ШІ та Data Mining.

Зазначене питання стає критично важливим в умовах розширення горизонтів цифрової гуманітаристики (Digital Humanities), що репрезентує синергію соціогуманітарного знання та сучасного ІТ-інструментарію: машинного навчання, комп'ютерної лінгвістики (NLP), Big Data та Data Mining.

Завдяки впровадженню інтелектуальних технологій стає можливим автоматичне опрацювання текстових масивів, моніторинг інформаційного простору, цифровізація надбань культурної спадщини та експлікація прихованих трендів у суспільних комунікаціях. Проте імплементація ШІ в гуманітарну практику породжує комплекс методологічних, інженерно-технологічних та етичних викликів.

Однією з ключових проблем є недостатній рівень міждисциплінарної інтеграції гуманітарних наук та інформаційних технологій. Специфіка сучасних соціогуманітарних розвідок вимагає залучення високотехнологічних смарт-систем, проте значна частина гуманітарних фахівців не володіє необхідними цифровими компетентностями. З іншого боку, представники ІТ-

індустрії часто не враховують специфіку гуманітарної методології. Зазначена диспропорція зумовлює об'єктивну необхідність у формуванні нових моделей підготовки фахівців, чиї компетенції охоплюють стик інформаційних технологій та гуманітарної сфери.

У межах спеціальності В13 (029) «Бібліотечна, інформаційна та архівна справа» пріоритетного значення набувають питання цифровізації фондів, автоматизованого опису та каталогізації документів, проектування інтелектуальних систем інформаційного пошуку та забезпечення ефективного управління цифровими фондами. Натомість вектор спеціальності F6 (126) «Інформаційні системи та технології» орієнтований на розробку сучасних інформаційних систем, моделей машинного навчання та аналітичних середовищ, що адаптовані для релевантного опрацювання великих масивів гуманітарних даних.

Аналіз останніх досліджень і публікацій. Використання штучного інтелекту та методів Data Mining у гуманітарній площині перетворилося на один із головних векторів сучасної Digital Humanities [1, 2]. Стрімке зростання кількості досліджень, пов'язаних із впровадженням алгоритмів машинного навчання, інструментів NLP, генеративного ШІ та великих даних, яскраво демонструє становлення новітньої міждисциплінарної наукової парадигми.

Сучасні дослідження дедалі частіше фокусуються на потенціалі великих мовних моделей (LLM) у межах соціогуманітарного знання. Зокрема, А. Karjus у своєму дослідженні [1] аналізує роль ШІ в автоматизації складних якісних завдань та впровадженні масштабованого аналізу даних. Дослідник акцентує, що інструменти штучного інтелекту забезпечують ефективний синтез класичної гуманітарної методології та кількісних методів обробки інформації.

Застосування технологій NLP у цифровій гуманітаристиці наразі перебуває в центрі уваги наукової спільноти. Зокрема, у фокусі міжнародної ініціативи «Natural Language Processing for Digital Humanities (NLP4DH)» перебуває створення інструментів для автоматизованого текстового аналізу, семантичного пошуку, оцифрування архівних фондів та тематичного моделювання.

Головний акцент дослідники роблять на оптимізації процесів обробки історичних першоджерел, електронних бібліотечних каталогів та об'єктів культурної спадщини.

Останні наукові публікації дедалі частіше акцентують на зростаючому потенціалі генеративного ШІ в гуманітарній сфері. Зокрема, у праці [3] обґрунтовано, що впровадження Text Mining, машинного навчання та генеративних моделей оптимізує процеси обробки соціогуманітарних даних, стимулює цифровізацію кураторства та посилює міждисциплінарну інтеграцію. Разом із тим дослідники застерігають про наявність супутніх викликів: етичних

ISSN 2786-6025 Online

дилем, алгоритмічної суб'єктивності та загрози дегуманізації через надмірну автоматизацію дослідницьких процесів.

Модернізація процесів опрацювання цифрових архівних масивів та формування метаданих наразі є одним із найбільш затребуваних дослідницьких векторів. Зокрема, у дослідженні [4] проаналізовано потенціал великих мовних моделей у контексті автоматизованої генерації метаданих та інтелектуального описування цифрових гуманітарних фондів. Зазначений підхід є стратегічно важливим для модернізації електронних бібліотек, сучасних архівних систем і профільних цифрових платформ у межах спеціальності В13 (029) «Бібліотечна, інформаційна та архівна справа».

Особливе місце в розвитку цифрової гуманітаристики посідають дослідження, що базуються на машинному навчанні та кількісному опрацюванні текстових масивів. У науковому огляді [5] систематизовано новітні підходи до виявлення інтертекстуальних зв'язків за допомогою статистичного моделювання, нейромереж та алгоритмів глибокого навчання. Дослідники наголошують, що впровадження цих технологій забезпечує фундаментальну трансформацію класичного якісного аналізу в масштабні гуманітарні проєкти на основі концепції data-driven.

Самостійним вектором сучасного наукового дискурсу є вивчення етичного виміру впровадження генеративного ШІ в академічне та освітнє середовище. У публікації [6] акцентовано на імперативності розробки чітких регламентів прозорого опрацювання даних за допомогою ШІ та безумовного дотримання вимог академічної доброчесності. Дослідники констатують, що на сьогодні у просторі цифрової гуманітаристики все ще бракує уніфікованих стандартів і консенсусних правил щодо інтеграції генеративних моделей у дослідницьку практику. Паралельно з цим у фокусі сучасного наукового дискурсу перебувають проблеми інформаційної безпеки та деструктивні ризики, пов'язані з експлуатацією штучного інтелекту. Дослідники акцентують на загрозах маніпулятивного застосування мовних моделей, проявах алгоритмічної суб'єктивності та поширенні дезінформації. Особливий інтерес становлять праці, де проаналізовано специфіку обходу захисних бар'єрів ШІ за допомогою художніх або гуманітарних контекстів, що безпосередньо вказує на гостру потребу в модернізації систем моніторингу та етико-правового регулювання технологій.

Важливою тенденцією сучасних досліджень є посилення міждисциплінарної взаємодії гуманітарних наук та інформаційних технологій. У працях, присвячених Computational Humanities та Digital Humanities, наголошується на необхідності підготовки фахівців нового покоління, які поєднуюватимуть гуманітарну методологію з навичками роботи з даними, алгоритмами машинного навчання та інтелектуальними системами.

Метою статті є дослідження сучасних підходів до гуманітарних досліджень із застосуванням технологій штучного інтелекту та інтелектуального аналізу даних, визначення особливостей їх використання у цифровій гуманітаристиці, а також обґрунтування перспектив міждисциплінарної інтеграції спеціальностей В13 (029) «Бібліотечна, інформаційна та архівна справа» та F6 (126) «Інформаційні системи та технології» в умовах цифрової трансформації науки й освіти.

Виклад основного матеріалу. Нинішня фаза еволюції інформаційного соціуму відзначена лавиноподібним збільшенням масивів електронних даних, масовим розгортанням інструментів штучного інтелекту та докорінною перебудовою класичної дослідницької методології. Зазначені тенденції яскраво виявляються в соціогуманітарному секторі, де модернізація архівних фондів, веббібліотек, медіаплатформ та каналів комунікації відкриває якісно нові горизонти для обробки, декодування й структурування відомостей. За таких обставин гуманітарні студії все чіткіше демонструють міждисциплінарну природу, інтегруючи традиційний аналітичний інструментарій із новітніми досягненнями ІТ-індустрії.

Провідне місце в зазначених трансформаціях посідають інструменти штучного інтелекту, машинного навчання та Data Mining, що дозволяють автоматизувати процеси обробки масштабних текстових, мультимедійних і структурованих масивів. Використання методів інтелектуального аналізу даних, обробки природної мови (NLP), семантичного моделювання та прогнозу аналітики відкриває перед науковцями можливості для експлікації прихованих трендів, моніторингу інформаційних потоків, реконструкції соціокультурних явищ і розробки інноваційних парадигм декодування гуманітарного знання.

Максимальну актуальність ці технологічні рішення мають для підготовки здобувачів вищої освіти за спеціальностями В13 (029) та F6 (126). У межах спеціальності В13 (029) інструменти ШІ створюють передумови для оптимізації та автоматизації рутинних бібліотечно-архівних процедур, інтелектуалізації пошукових алгоритмів, оцифрування історико-документальної спадщини та когнітивного аналізу інформаційних масивів. Натомість спеціальність F6 (126) фокусується на проектуванні та розгортанні інтелектуальних систем, архітектурі машинного навчання, методології Big Data та сервісах підтримки прийняття рішень, які формують інженерно-технологічний базис сучасної Digital Humanities.

Стрімка еволюція цифрового середовища та глобальна інформатизація соціуму зумовили появу інноваційних парадигм у сфері організації та методології гуманітарного знання. Класичні підходи до дослідження текстових масивів, історичних першоджерел, об'єктів культурної спадщини та

ISSN 2786-6025 Online

соціокультурних явищ наразі активно збагачуються алгоритмами штучного інтелекту, машинного навчання та Data Mining. На стику цих дисциплін формується самостійний міждисциплінарний простір – цифрова гуманітаристика (Digital Humanities), що забезпечує синергію соціогуманітарних наук із комп'ютерною аналітикою та інформаційними технологіями.

Штучний інтелект розглядається як комплекс методологій та технологічних рішень, орієнтованих на моделювання систем, що спроможні реалізувати когнітивні функції людини. До спектра таких функцій належать деструктуризація текстів, ідентифікація образів, систематизація відомостей, прогностичне моделювання, обґрунтування рішень та генерація рекомендацій. У межах соціогуманітарного знання ШІ затребуваний передусім як інструмент автоматизованої обробки масштабних масивів інформації, що дає змогу суттєво інтенсифікувати та оптимізувати дослідницький процес.

Серед пріоритетних напрямів застосування ШІ в гуманітарному просторі виокремлюється Data Mining, спрямований на експлікацію латентних закономірностей, кореляцій та трендів у масивах структурованої та неструктурованої інформації. Зазначений інструментарій уможливорює комплексне опрацювання масштабних документальних колекцій, контенту соціальних мереж, цифрових архівів, медіатекстів та статистичних показників, суттєво розширюючи методологічні межі гуманітаристики. Імплементация таких технологій стимулює трансформацію класичного дескриптивного аналізу в наукові розвідки на основі концепції data-driven, що базуються на кількісному та алгоритмічному моделюванні.

Серед методів Data Mining, що знайшли найширше застосування у гуманітаристиці та суміжних інформаційних галузях, виділяють такі:

– корпусний аналіз і дистантне читання (distant reading) – статистичне вивчення великих текстових масивів через частотність, колокації та n-грами для виявлення дискурсивних тенденцій, недоступних при традиційному прочитанні окремих текстів. Може бути базовим інструментом роботи з електронними документальними фондами;

– темп-моделювання (topic modeling) на основі алгоритмів LDA (Latent Dirichlet Allocation), NMF і BERTopic – автоматичне виявлення прихованих тематичних структур у колекціях документів. Може забезпечувати автоматизацію тематичної каталогізації й класифікації – ключових функцій бібліотечно-архівної справи;

– мережевий аналіз (SNA, Social Network Analysis) – дослідження структури зв'язків між акторами (особами, установами, документами) для виявлення прихованих зв'язків у великих документальних масивах. Може бути незамінним при реконструкції корпоративних, родинних і наукових мереж на основі архівних матеріалів;

– розпізнавання іменованих сутностей (NER) – автоматична ідентифікація власних назв, дат, географічних об'єктів у неструктурованих текстах. Може бути основою для автоматичного складання покажчиків і метаданих у бібліотечних і архівних системах;

– кластеризація і класифікація документів – автоматичне групування за схожістю ознак як без попереднього маркування (unsupervised), так і на основі навчальної вибірки (supervised). Критично важливе для впорядкування великих електронних архівів.

Поява технологій штучного інтелекту нового покоління суттєво розширила та частково змінила цей арсенал. Принципово важливою тенденцією є те, що системи ШІ та інструменти Data Mining не залишаються статичними – вони постійно й інтенсивно вдосконалюються, що є ключовою характеристикою сучасного стану галузі.

По-перше, стрімке зростання можливостей великих мовних моделей (LLM). З 2022 р. до 2024 р. розміри контекстних вікон зросли більш ніж у 100 разів: від 4 тис. токенів у перших версій GPT-3.5 до 200 тис. токенів у Claude 3.5 і 2 млн токенів у Gemini 1.5 Pro. Це означає, що сучасна модель здатна проаналізувати цілу архівну справу або кілька сотень сторінок документів в межах одного запиту. Одночасно якість розуміння тексту, точність перекладу і здатність до складних міркувань покращуються з кожним новим поколінням моделей.

По-друге, перехід від спеціалізованих моделей до мультимодальних агентних архітектур. Якщо ще кілька років тому системи розпізнавання тексту, аналізу зображень і обробки мови існували окремо, сучасні платформи поєднують усі ці функції в єдиному робочому середовищі. GPT-4o, Claude 3 Opus і Gemini 1.5 Pro здатні одночасно обробляти текст, зображення, аудіо та структуровані дані, що відкриває якісно нові можливості для роботи з мультимедійними архівними фондами.

ШІ-агенти нового покоління можуть самостійно виконувати послідовності дослідницьких операцій: знаходити документи, транскрибувати їх, перекладати, аналізувати і формувати звіт без участі людини на проміжних кроках.

По-третє, поява та розвиток спеціалізованих україномовних ресурсів. Ця тенденція особливо актуальна у контексті реалізації Стратегії розвитку штучного інтелекту в Україні на 2024–2030 роки [7], яка визначає розбудову вітчизняного ШІ як державний пріоритет. В рамках цієї стратегії активізується робота зі створення україномовних NLP-корпусів та навчання мовних моделей на вітчизняних текстових даних. Серед актуальних ініціатив — проекти Ukrainer та Vox Ukraine зі створення анотованих українських корпусів, пілотні проекти Національного банку кінофотофонодокументів з автоматизованої

ISSN 2786-6025 Online

каталогізації та розвиток україномовних адаптацій Transkribus для кирилических рукописів.

По-четверте, демократизація доступу до потужних ШІ-інструментів. Якщо кілька років тому впровадження систем глибокого навчання вимагало значних обчислювальних ресурсів і вузькоспеціалізованих технічних знань, сьогодні більшість провідних ШІ-систем доступні через вебінтерфейс або просте API. Це означає, що фахівці зі спеціальності В13 (029), які не мають глибокої технічної підготовки, вже сьогодні можуть використовувати потужні інструменти аналізу даних у своїй щоденній практиці.

Наочна порівняльна характеристика сучасного інструментарію обох класів технологій подана у таблиці 1.

Таблиця 1

Порівняльна характеристика ШІ-інструментів та засобів data mining для гуманітарних досліджень (станом на 2024–2025 рр.)

| Інструмент | Клас / тип | Ключові функції | Застосування в 029 / 126 | Доступ | Тенденція розвитку |
|-------------------|------------------|--|---------------------------------|--------------------|-------------------------|
| spaCy / NLTK | NLP, Python | NER, морфологія, синтаксис | Автоіндексація документів | Open source | Нові укр. моделі |
| Gensim / BERTopic | Topic modeling | LDA, NMF, BERTopic, word2vec | Тематична каталогізація фондів | Open source | Нейромережеві моделі |
| Voyant Tools | Корпусний аналіз | Частоти, конкорданси, хмари слів | Аналіз документальних колекцій | Безкоштовно, веб | Стабільний |
| Gephi | Мережевий аналіз | SNA, центральності, спільноти | Зв'язки між архівними фондами | Open source | Інтеграція з LLM |
| Transkribus | HTR / ШІ | Розпізнавання рукописів, 300+ моделей | Оцифровка архівних документів | Freemium, веб | Укр. кирилическі моделі |
| Whisper (OpenAI) | ASR / ШІ | Транскрипція усного мовлення | Фонодокументи, усна історія | Open source | Мультимовна підтримка |
| GPT-4o | LLM / ШІ | Текст + зображення + аудіо, 128k токенів | Реферування, переклад, метадані | API / ChatGPT Plus | Агентні можливості |
| Claude 3.5 Sonnet | LLM / ШІ | 200k токенів, аналіз великих документів | Масивні архівні справи | API / Claude.ai | Зростання контексту |
| Gemini 1.5 Pro | LLM / ШІ | 2М токенів, мультимодальний | Відео-, аудіофонди | API / Google AI | Агентні системи |

| Інструмент | Клас / тип | Ключові функції | Застосування в 029 / 126 | Доступ | Тенденція розвитку |
|----------------------------------|-------------------|---------------------------------|--------------------------|---------------------|--------------------|
| Elasticsearch + векторний індекс | Семантичний пошук | Пошук за змістом без збігу слів | Пошук у цифрових архівах | Open source / хмара | Інтеграція з LLM |
| Elicit | LLM-пошук | Скринінг, огляди літератури | Бібліографічний пошук | Freemium, веб | Активний розвиток |

Разом із тим застосування розглянутих технологій пов'язане з низкою серйозних методологічних і практичних викликів, про які необхідно знати як дослідникам, так і розробникам інформаційних систем. Найбільш принциповим є проблема галюцинацій LLM – генерування правдоподібних, але фактично хибних тверджень, включно з вигаданими цитатами та датами. Дослідження фіксують від 3 до 8% фактичних помилок у відповідях навіть найкращих моделей станом на 2024 р. Для архівно-бібліотечної справи, де точність метаданих і атрибуцій є критичною, це вимагає обов'язкової верифікації результатів ШІ-аналізу компетентним фахівцем.

Фундаментальним теоретичним базисом цифрової гуманітаристики виступає концепція великих даних (Big Data), орієнтована на опрацювання масштабних інформаційних масивів, що відзначаються значним обсягом, гетерогенністю та високою динамікою оновлення. Сучасний соціогуманітарний простір дедалі частіше оперує цифровими ресурсами соціальних платформ, вебкаталогами бібліотек, онлайн-медіа, відкритими реєстрами та мультимедійними даними. Своєю чергою, це актуалізує потребу в залученні інноваційних інформаційних систем та алгоритмів машинного навчання, спроможних гарантувати високу ефективність аналітичної обробки такої інформації.

У межах спеціальності F6 (126) пріоритетного значення набуває проєктування інтелектуальних інформаційних платформ, адаптованих для супроводу соціогуманітарних розвідок. Подібні системи забезпечують інструментальну підтримку прийняття рішень, автоматизовану рубрикацію та класифікацію джерел, прогностичне моделювання інформаційних потоків і семантичний аналіз. Такий інструментарій формує фундамент для принципово нових дослідницьких парадигм, що базуються на синтезі прикладних технологічних та аналітичних підходів.

Серед найбільш затребуваних методологічних інструментів у соціогуманітарному просторі особливе місце посідає Text Mining – технологія інтелектуального аналізу текстових масивів, орієнтована на експлікацію латентних закономірностей, ключових дескрипторів та тематичних кореляцій. Цей підхід активно залучається для деструктуризації історичних першоджерел,

ISSN 2786-6025 Online

наукових розвідок, контенту мас-медіа, соціальних комунікативних каналів та архівних фондів. Імплементация Text Mining уможлиблює автоматизовану рубрикацію документів, ідентифікацію рекурентних інформаційних конструкцій та реалізацію масштабного контент-аналізу значних обсягів вербальної інформації.

Важливе місце в методологічному арсеналі сучасних гуманітарних розвідок належить технологіям NLP. Вони створюють умови для автоматизованої деструктуризації текстових масивів, експлікації семантичних кореляцій, тематичного моделювання, дистрибутивного аналізу тональності (сентимент-аналізу) та маркування ключових дескрипторів. Впровадження NLP уможлиблює лінгвокомп'ютерне вивчення масштабних текстових корпусів, автоматизоване опрацювання архівних джерел, моніторинг медіасфери та моделювання суспільних комунікативних процесів. У межах спеціальності В13 (029) зазначений інструментарій є стратегічно важливим для модернізації та цифровізації бібліотечно-архівних фондів, проектування електронних каталогів нового покоління та розгортання систем інтелектуального пошуку знань.

NLP забезпечують автоматичне розпізнавання, аналіз та інтерпретацію людської мови за допомогою алгоритмів штучного інтелекту. Серед основних завдань NLP можна виділити:

- автоматичний переклад текстів;
- аналіз тональності повідомлень;
- визначення ключових слів та понять;
- тематичне моделювання;
- розпізнавання іменованих сутностей;
- автоматичне реферування документів та ін.

Технології NLP мають особливе значення у процесах цифровізації бібліотечних і архівних ресурсів, створення інтелектуальних каталогів, електронних інформаційних систем та сервісів пошуку документів. Автоматизований аналіз текстів значно підвищує ефективність інформаційно-аналітичної діяльності та спрощує роботу з великими документальними фондами.

Одним із перспективних напрямів є використання машинного навчання (Machine Learning) для класифікації та прогнозування гуманітарних процесів. Алгоритми машинного навчання дозволяють аналізувати поведінкові моделі користувачів, прогнозувати інформаційні тенденції, здійснювати автоматичне групування документів та визначати приховані закономірності у даних. У гуманітарних дослідженнях це дає змогу:

- аналізувати інформаційні потоки у медіасередовищі;
- виявляти дезінформаційні кампанії;
- прогнозувати соціальні та культурні тенденції;

– моделювати процеси поширення інформації у суспільстві.

Методи машинного навчання є основою створення інтелектуальних інформаційних систем, здатних автоматично аналізувати та обробляти гуманітарні дані. Особливу роль відіграють нейронні мережі та алгоритми глибокого навчання (Deep Learning), які забезпечують високу точність обробки текстової, аудіовізуальної та графічної інформації.

Суттєве значення у сучасних гуманітарних дослідженнях має семантичний аналіз інформації. Його метою є встановлення змістових зв'язків між поняттями, документами та інформаційними об'єктами. Семантичні технології використовуються для побудови онтологій, знанневих графів та інтелектуальних систем підтримки прийняття рішень. Це дозволяє формувати структуровані моделі знань та забезпечувати більш точний пошук інформації у цифрових ресурсах.

Окремим напрямом є технології візуалізації даних, які забезпечують графічне представлення результатів аналізу інформації. Використання діаграм, інтерактивних карт, мережеских графів та аналітичних панелей дозволяє дослідникам ефективніше інтерпретувати великі обсяги даних та виявляти приховані закономірності. У гуманітарній сфері візуалізація застосовується для аналізу соціальних мереж, інформаційних зв'язків, культурних процесів та історичних подій.

Важливим сучасним напрямом є використання генеративного штучного інтелекту, який здатний створювати тексти, узагальнення, аналітичні матеріали та мультимедійний контент. Такі технології активно використовуються у сфері інформаційно-аналітичної діяльності, освітніх процесах, медіакommunікаціях та цифровій гуманітаристиці. Водночас застосування генеративних моделей потребує контролю достовірності результатів і дотримання принципів академічної доброчесності.

Одним із найпоширеніших напрямів використання штучного інтелекту є автоматизований аналіз історичних джерел та архівних документів. Технології OCR (Optical Character Recognition) дозволяють розпізнавати друковані та рукописні тексти, переводити їх у цифровий формат і здійснювати подальше інтелектуальне опрацювання. Це значно спрощує процес цифровізації архівних фондів, створення електронних бібліотек та збереження культурної спадщини. Використання алгоритмів NLP дає змогу автоматично класифікувати документи, визначати ключові поняття та здійснювати семантичний пошук інформації у великих текстових масивах.

Створення інтелектуальних бібліотечних та архівних систем дозволить забезпечити автоматизовану каталогізацію документів, формування електронних метаданих, рекомендацію інформаційних ресурсів користувачам та аналіз читацьких запитів. Використання штучного інтелекту у бібліотечній сфері

ISSN 2786-6025 Online

сприяє підвищенню ефективності інформаційного обслуговування та оптимізації роботи з цифровими ресурсами.

Значного поширення набуло застосування ШІ для аналізу соціальних комунікацій та медіапростору. Соціальні мережі, новинні платформи та онлайн-медіа генерують великі обсяги інформації, аналіз яких дозволяє досліджувати суспільні настрої, інформаційні тенденції та комунікаційні процеси. Методи аналізу тональності текстів (Sentiment Analysis) допомагають визначати емоційне забарвлення повідомлень, виявляти реакцію суспільства на певні події та оцінювати інформаційний вплив медіаконтенту.

Активне впровадження технологій штучного інтелекту у гуманітарну сферу суттєво змінює традиційні підходи до наукових досліджень, інформаційно-аналітичної діяльності та роботи з цифровими ресурсами. Використання алгоритмів машинного навчання, обробки природної мови, інтелектуального аналізу даних та генеративних моделей відкриває нові можливості для гуманітарних наук, водночас створюючи низку технологічних, етичних і методологічних викликів.

У сучасному цифровому середовищі гуманітарні дослідження пов'язані з аналізом значних масивів текстових, аудіовізуальних та мультимедійних даних, які неможливо ефективно опрацювати виключно традиційними методами. Алгоритми штучного інтелекту дозволяють швидко здійснювати пошук, класифікацію, систематизацію та аналіз інформації, що значно підвищує продуктивність наукової діяльності.

Важливою перевагою є також можливість аналізу неструктурованих даних. Технології NLP, Text Mining та машинного навчання забезпечують автоматичне опрацювання текстів, визначення тематичних зв'язків, аналіз тональності повідомлень і виявлення прихованих закономірностей у документах та медіаресурсах. Це створює нові можливості для дослідження історичних джерел, соціальних комунікацій, культурних процесів та інформаційного простору.

Алгоритми штучного інтелекту працюють на основі наявних даних і моделей навчання, тому результати їхньої роботи значною мірою залежать від якості інформації та методів обробки. Помилки у вихідних даних або некоректне налаштування моделей можуть призводити до викривлення результатів дослідження.

Суттєвою проблемою є алгоритмічна упередженість (algorithmic bias). Штучний інтелект може відтворювати або навіть посилювати соціальні, культурні та інформаційні упередження, присутні у навчальних даних. У гуманітарній сфері це може призводити до спотворення результатів аналізу текстів, неправильної інтерпретації історичних подій чи формування необ'єктивних інформаційних висновків.

Гуманітарні дослідження часто передбачають роботу з персональними даними, архівними документами та конфіденційною інформацією. Використання інтелектуальних систем потребує дотримання принципів кібербезпеки, захисту інформаційних ресурсів та контролю доступу до даних.

Також існує ризик надмірної автоматизації гуманітарних досліджень. Незважаючи на значні можливості штучного інтелекту, він не здатний повністю замінити гуманітарну інтерпретацію, критичне мислення та творчий підхід дослідника. Гуманітарні науки значною мірою базуються на контекстному аналізі, культурному розумінні та суб'єктивній інтерпретації, що потребує участі людини у процесі дослідження.

Стрімкий розвиток інформаційних технологій, штучного інтелекту та інтелектуального аналізу даних формує новий етап еволюції гуманітарних наук, у межах якого цифрові гуманітарні дослідження стають одним із ключових напрямів наукової діяльності. Цифрова гуманітаристика (Digital Humanities) поступово трансформується з допоміжного інструменту аналізу інформації у самостійну міждисциплінарну сферу, що поєднує гуманітарні науки, інформаційні технології, аналітику даних та методи штучного інтелекту. Це відкриває широкі перспективи для модернізації наукових досліджень, освітнього процесу та інформаційно-аналітичної діяльності.

Однією з основних тенденцій розвитку цифрової гуманітаристики є подальша інтеграція штучного інтелекту у процеси аналізу гуманітарних даних. Використання алгоритмів машинного навчання, NLP, генеративного ШІ та семантичних технологій дозволяє автоматизувати складні процеси опрацювання текстів, історичних джерел, медіаконтенту та соціальних комунікацій. У перспективі такі системи зможуть забезпечувати не лише аналіз інформації, а й формування аналітичних прогнозів, моделювання суспільних процесів та підтримку прийняття рішень у гуманітарній сфері.

Перспективним напрямом є створення цифрових гуманітарних лабораторій та міждисциплінарних дослідницьких центрів, які об'єднуюватимуть фахівців гуманітарного та ІТ-профілю. Така взаємодія сприятиме розробці нових методів аналізу інформації, інтеграції сучасних технологій у гуманітарні дослідження та підготовці фахівців нового покоління. У сучасних умовах особливої актуальності набуває концепція data-driven research, за якої наукові висновки формуються на основі комплексного аналізу великих обсягів даних.

Генеративні моделі можуть використовуватися для автоматичного створення аналітичних матеріалів, узагальнення інформації, перекладу текстів, формування рекомендацій та підтримки освітнього процесу. У гуманітарних дослідженнях це сприятиме пришвидшенню опрацювання інформації та підвищенню ефективності наукової роботи з дотриманням принципів академічної доброчесності.

ISSN 2786-6025 Online

Важливою перспективою є розвиток цифрових компетентностей фахівців гуманітарної сфери. Сучасний фахівець повинен володіти навичками роботи з інформаційними системами, аналітичними платформами, технологіями обробки даних та інструментами штучного інтелекту. Це потребує оновлення освітніх програм, інтеграції міждисциплінарних курсів та формування нових моделей професійної підготовки фахівців.

В Україні цифрова гуманітаристика перебуває на етапі активного становлення. Розвиток електронних архівів, цифрових бібліотек, національних інформаційних платформ та систем відкритих даних створює основу для подальшої інтеграції штучного інтелекту у гуманітарні дослідження. Водночас важливими завданнями залишаються розвиток цифрової інфраструктури, підготовка кваліфікованих фахівців та формування нормативно-етичних засад використання ШІ у науковій діяльності.

Висновки. У результаті дослідження було визначено, що сучасний розвиток цифрових технологій, штучного інтелекту та інтелектуального аналізу даних докорінно змінює методологічний простір соціогуманітарного знання, започатковуючи інноваційні формати дослідницької та інформаційно-аналітичної практики.

У цьому контексті Digital Humanities утворюється як ключова міждисциплінарна платформа, що інтегрує класичну гуманітарну парадигму, комп'ютерні науки, аналітику даних та інтелектуальні системи. Використання сучасних цифрових інструментів гарантує автоматизацію рутинних операцій з опрацювання джерел, оптимізує аналіз масштабних інформаційних масивів і створює додаткові перспективи для вивчення трансформацій у соціокультурній та комунікаційній сферах.

Інструменти штучного інтелекту, зокрема алгоритми машинного навчання, обробки природної мови (NLP), Text Mining, семантичного моделювання та Big Data, наразі активно імплементуються в соціогуманітарний простір. Стратегічну цінність зазначені технологічні рішення мають для модернізації підготовки фахівців за спеціальностями В13 (029) та F6 (126), оскільки забезпечують інтеграцію гуманітарної методології з сучасними інформаційними системами та аналітичними платформами.

Застосування штучного інтелекту та інтелектуального аналізу даних у гуманітарних дослідженнях є важливим чинником модернізації сучасної науки та освіти. Інтеграція гуманітарних підходів із можливостями цифрових технологій створює передумови для розвитку нових методів дослідження, підвищення ефективності інформаційно-аналітичної діяльності та формування інноваційного освітнього й наукового середовища відповідно до вимог цифрового суспільства.