

ТЕМА 4. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В НАУКОВИХ ДОСЛІДЖЕННЯХ

План

1. Види наукової інформації та її обробка.
2. Типи експериментальних даних, підготовка їх до обробки.
3. Цифрові інструменти текстового, табличного, графічного та математичного аналізу в наукових дослідженнях
4. Прикладне програмне забезпечення для візуалізації, аналізу і публікації даних.
5. Спеціалізовані пакети обробки даних в наукових дослідженнях.
6. Використання штучного інтелекту для автоматизації аналізу великих даних



Наукова інформація – це систематизовані, перевірені та інтерпретовані відомості про об’єктивну реальність, отримані в межах наукової методології та призначені для формування нового знання.

Ключові ознаки:

- доказовість;
- відтворюваність;
- об’єктивність (наскільки це можливо);
- структурованість;
- джерельна верифікованість;
- логічна узгодженість.

Властивості наукової інформації:

- має чітко визначену методику отримання;
- підлягає перевірці іншими дослідниками;
- інтегрується в систему наукового знання.

Класифікація видів наукової інформації

за рівнем узагальнення	<ul style="list-style-type: none">– емпірична (результати спостережень, експериментів, опитувань, вимірювань);– теоретична (узагальнення емпіричних даних: моделі, гіпотези, концепції, закономірності);метатеоретична (методологічні підходи, філософські засади науки, парадигми)
за формою подання	<ul style="list-style-type: none">– текстова (монографії, статті);– числова (статистика);– графічна (схеми, карти, інфографіка) ;– аудіовізуальна;машиночитана (бази даних, коди, алгоритми)
за джерелом походження	<ul style="list-style-type: none">– первинна (отримана безпосередньо дослідником);– вторинна (аналіз уже існуючих джерел);третинна (аналітичні огляди, систематизація досліджень (meta-analysis))
за характером даних	<ul style="list-style-type: none">– кількісна (числові показники, що підлягають математичній обробці);– якісна (сміслові, описові дані)
за правовим режимом	<ul style="list-style-type: none">– відкрита– конфіденційна– комерційна таємниця– державна таємниця– персональні дані

Життєвий цикл наукової інформації

1. **Формування запиту** – це визначення наукової проблеми, постановка дослідницького питання та формулювання гіпотези. Саме на цьому етапі закладається логіка всього дослідження.
2. **Збір даних** – отримання первинної або вторинної інформації відповідно до обраної методології.
3. **Верифікація** – перевірка достовірності та надійності отриманих даних.
4. **Очищення (Data Cleaning)** – підготовка даних до аналізу шляхом усунення помилок.
5. **Кодування** – перетворення даних у форму, придатну для аналізу.
6. **Аналіз** – виявлення закономірностей, зв'язків та структур у даних.
7. **Інтерпретація** – пояснення отриманих результатів у контексті наукової теорії та практики.
8. **Публікація** – оприлюднення результатів дослідження в науковому середовищі.
9. **Архівування** – довгострокове збереження дослідницьких матеріалів.

МЕТОДИ ОБРОБКИ НАУКОВОЇ ІНФОРМАЦІЇ

Логічні методи	<ul style="list-style-type: none">– аналіз і синтез– індукція та дедукція– абстрагування– узагальнення– моделювання
Статистичні методи	<ul style="list-style-type: none">– описова статистика– кореляційний аналіз– регресійні моделі– дисперсійний аналіз– факторний аналіз– кластеризація
Методи якісного аналізу	<ul style="list-style-type: none">– контент-аналіз– дискурс-аналіз– case-study– компаративний аналіз– експертне оцінювання
Цифрові методи	<ul style="list-style-type: none">– Big Data-аналітика– машинне навчання– NLP– data mining– візуалізація

ПРОБЛЕМИ В РОБОТІ З НАУКОВОЮ ІНФОРМАЦІЄЮ

- 1) інформаційне перевантаження (проблема надлишку джерел);
- 2) низька якість даних (помилки вимірювання, упередженість вибірки);
- 3) реплікаційна криза (неможливість повторити результати);
- 4) етичні ризики (фабрикація, фальсифікація, плагіат, маніпуляція статистикою)

ВИМОГИ ДО ОБРОБКИ НАУКОВОЇ ІНФОРМАЦІЇ

1. Об'єктивність
2. Системність
3. Відтворюваність
4. Прозорість методології
5. Документування процедур
6. Захист персональних даних

Експериментальні дані – це інформація, отримана внаслідок контрольованого втручання дослідника у досліджуваний процес з метою перевірки гіпотези та встановлення причинно-наслідкових зв'язків. Їх ключовою відмінністю від спостережних даних є активна роль дослідника: він не лише фіксує реальність, а й змінює умови функціонування об'єкта, контролює змінні та вимірює ефект цього впливу.

Експериментальна методологія передбачає:

- наявність незалежної змінної (фактору впливу);
- залежної змінної (результату);
- контрольних змінних;
- можливість повторення (реплікації);
- стандартизовані умови вимірювання.

Класифікація експериментальних даних

За природою інформації	<ul style="list-style-type: none">– кількісні дані: отримуються шляхом вимірювання, тестування, обчислення або статистичного фіксування показників;– якісні дані: мають описовий характер і виражаються у формі текстів, спостережень, інтерв'ю, поведінкових реакцій або експертних оцінок;– змішані дані поєднують кількісні та якісні компоненти
За шкалою вимірювання	<ul style="list-style-type: none">– номінальні дані відображають категорії без внутрішнього порядку (тип організації, вид правопорушення).– порядкові (ординальні) мають ранжування, але інтервали між значеннями не рівні (рівень довіри, ступінь ризику).– інтервальні характеризуються рівними інтервалами, але не мають абсолютного нуля (температура за Цельсієм).– дані шкали відношень мають абсолютний нуль і дозволяють виконувати всі арифметичні операції (дохід, кількість подій, час)
За часовою організацією	<ul style="list-style-type: none">– крос-секційні дані відображають стан об'єкта в конкретний момент часу;– поздовжні (longitudinal) фіксують динаміку змін;– панельні дозволяють спостерігати одні й ті самі об'єкти протягом кількох періодів
За структурою	<ul style="list-style-type: none">– структуровані – організовані у таблиці або бази даних.– напівструктуровані – мають частково визначену структуру (JSON, XML).– неструктуровані – тексти, аудіо, відео, зображення.

Підготовка даних – це системний процес перетворення сирової інформації на аналітично придатний масив. Вона є ключовою умовою валідності дослідження.

Етапи підготовки даних:

- **верифікація** (спрямована на перевірку достовірності отриманих даних та включає логічний контроль; перевірку меж допустимих значень; виявлення аномалій; повторні вимірювання; перехресну перевірку джерел;
- **очищення** (Data Cleaning) полягає в усуненні дублікатів; обробці пропущених значень; усуненні статистичних викидів; корекції форматів; узгодженні одиниць вимірювання;
- **кодування** – це перетворення даних у форму, придатну для аналізу;
- **стандартизація та нормалізація** застосовуються для забезпечення коректності математичного моделювання;
- **перевірка статистичних припущень**;
- **документування**

Цифрові інструменти текстового, табличного, графічного та математичного аналізу стали невід'ємною складовою сучасних наукових досліджень. Вони формують технологічну основу дослідницького процесу, забезпечують обробку великих масивів інформації, підвищують точність аналітики та створюють умови для відтворюваності результатів. У сучасній академічній практиці володіння цифровими інструментами є не технічною навичкою, а методологічною компетентністю дослідника.

Інструменти текстового аналізу

- **Текстові процесори та системи структурованого письма:** забезпечують створення, редагування та форматування наукових текстів відповідно до академічних стандартів.
- **Менеджери бібліографії** дозволяють: збирати джерела з наукових баз; автоматично генерувати цитування у різних стилях (APA, Chicago, MLA); формувати бібліографію; синхронізувати дані між пристроями.
- **Інструменти контент-аналізу:** дозволяють кодувати тексти; виділяти теми та категорії; аналізувати частоту понять; виявляти семантичні зв'язки; будувати концептуальні карти. Вони використовуються у соціології, праві, психології, політології.

Табличні інструменти аналізу

- **Електронні таблиці** дозволяють: організовувати бази даних; виконувати формульні розрахунки; будувати зведені таблиці; застосовувати логічні та статистичні функції; створювати базові діаграми.
- **Системи управління базами даних** дозволяють: працювати з мільйонами записів; здійснювати складні запити; інтегрувати різні джерела інформації.

Графічні інструменти та візуалізація даних

- **Статистичні пакети з вбудованою графікою** дозволяють: будувати гістограми; кореляційні матриці; регресійні графіки; теплові карти; інтерактивні панелі.
- **ВІ-платформи** використовуються для: інтерактивних дашбордів; візуалізації великих даних; порівняльного аналізу; стратегічного моделювання;
- Геоінформаційні системи (GIS)** дозволяють: аналізувати просторові дані; будувати карти; моделювати територіальні ризики; досліджувати регіональні процеси.

Математичні та статистичні інструменти

- **Статистичні пакети дозволяють:** виконувати описову статистику; тестувати гіпотези; будувати регресійні моделі; здійснювати факторний і кластерний аналіз; працювати з панельними даними.
- **Математичне моделювання** використовується для: розв'язання диференціальних рівнянь; моделювання процесів; чисельних методів; оптимізаційних задач.

Основні можливості штучного інтелекту у наукових дослідженнях

Автоматизоване очищення даних – алгоритми ШІ виявляють аномалії, пропущені або некоректні значення, усувають дублікати та стандартизують інформацію;

Збір та інтеграція даних з різних джерел – ШІ дозволяє автоматично збирати дані з відкритих баз, соціальних мереж, наукових репозитаріїв;

Класифікація та категоризація – алгоритми машинного навчання автоматично сортують дані за категоріями або темами, що значно прискорює обробку текстів, зображень, відео чи аудіо.

Виявлення закономірностей та аномалій – ШІ застосовує алгоритми кластеризації, регресійного аналізу, нейронних мереж та дерев рішень для виявлення прихованих зв'язків у даних, які неможливо помітити при ручній обробці.

Прогнозування – моделі глибокого навчання дозволяють будувати прогнози на основі історичних даних.

Переваги використання штучного інтелекту

Автоматизація рутинних процесів – очищення, сортування, класифікація даних.

Масштабування обробки – обробка мільйонів записів за лічені хвилини.

Виявлення складних закономірностей – алгоритми ШІ здатні знаходити приховані взаємозв'язки, недоступні для традиційного аналізу.

Прогнозування та моделювання сценаріїв – від економіки до медицини та екології.

Інтерактивність і інтеграція – інтеграція з візуалізацією, веб-додатками, інтерактивними панелями.

Штучний інтелект у сфері аналізу великих даних – це насамперед здатність системи самостійно виявляти закономірності, формувати узагальнення, прогнозувати ризики та підтримувати прийняття управлінських рішень через навчання, узагальнення, класифікацію, виявлення відхилень, оцінювання ймовірностей.

Автоматизація аналізу через використання штучного інтелекту є інструментом підтримки прийняття рішень, дозволяючи:

- формувати оцінку ризику;
- ранжувати об'єкти за рівнем небезпеки;
- пропонувати прогноз розвитку ситуації.

Використання штучного інтелекту трансформує культуру аналізу, оскільки він дозволяє:

- перейти від ретроспективного аналізу до прогнозного;
- зменшити роль випадковості;
- забезпечити системність контролю;
- підвищити прозорість процесів.

Фактично відбувається перехід до нової моделі управління – управління на основі даних

Етичні та управлінські виклики щодо використання штучного інтелекту:

- проблема відповідальності за рішення, прийняті алгоритмом;
- необхідність забезпечення прозорості моделей;
- ризик алгоритмічної упередженості;
- захист персональних даних.

Використання штучного інтелекту потребує не лише технологічного, а й нормативного, етичного та інституційного регулювання.

Використання штучного інтелекту в наукових дослідженнях

1. ШІ як інструмент наукового аналізу

Обробка великих масивів даних (Big Data)

- Автоматичне виявлення закономірностей у числових, текстових, геномних, кліматичних та соціальних даних.
- Застосування алгоритмів для швидкого порівняння тисяч і навіть мільйонів досліджень та результатів.

Прогнозування та моделювання

- Моделювання складних явищ, які неможливо дослідити емпірично (наприклад, кліматичні сценарії або еволюцію епідемій).
- Прогнозування результатів експериментів на основі попередніх даних, що економить ресурси та час.

Автоматизована класифікація та систематизація знань

- Систематизація літератури, аналітичне групування досліджень за темами, ключовими словами або методологічними підходами.
- Пошук релевантних джерел, що значно скорочує час на літературний огляд.

2. Використання ШІ для забезпечення академічної доброчесності

+	-
<p>Виявлення плагіату та академічних порушень</p> <ul style="list-style-type: none">- Алгоритми ШІ здатні перевіряти текст на збіги з науковими джерелами, навіть якщо використані перефразування.- Виявляють частковий або структурний плагіат у статтях та дисертаціях. <p>Контроль даних та прозорість експериментів</p> <ul style="list-style-type: none">- ШІ дозволяє автоматично відстежувати джерела даних, кроки аналізу та метадані експериментів.- Це підвищує відтворюваність результатів, що є ключовим принципом академічної доброчесності. <p>Аудит результатів досліджень</p> <ul style="list-style-type: none">- Алгоритми можуть перевіряти статистичну коректність, відповідність методології та наявність потенційних помилок у розрахунках.- Допомагають уникнути випадкових або системних помилок, які можуть поставити під сумнів достовірність публікації.	<p>Автоматичне створення текстів та фальсифікація результатів</p> <ul style="list-style-type: none">- Генеративні моделі ШІ можуть створювати наукові тексти або результати без проведення реальних експериментів.- Використання таких результатів без перевірки порушує академічну доброчесність. <p>Сліпа довіра до алгоритмів</p> <ul style="list-style-type: none">- Якщо дослідник повністю покладається на висновки ШІ, можна пропустити методологічні або логічні помилки.- Важливо, щоб людина залишалася відповідальною за інтерпретацію та перевірку результатів. <p>Ризик упередженості моделей</p> <ul style="list-style-type: none">- Алгоритми навчаються на історичних даних. Якщо дані мають системні упередження, це може відобразитися в результатах дослідження.- Наприклад, у соціальних науках ШІ може відтворювати гендерні або соціальні стереотипи.

ЕТИЧНІ ПРИНЦИПИ ЗАСТОСУВАННЯ ШІ В НАУЦІ

Прозорість алгоритмів – пояснювати, як ШІ обробляє дані та приймає рішення.

Верифікація результатів – результати ШІ завжди перевіряти експертами.

Використання достовірних даних – уникати навчання на сумнівних або спотворених масивах.

Документування процесу – всі кроки аналізу повинні бути зафіксовані, щоб забезпечити відтворюваність.

Використання ШІ як допоміжного інструменту – людина залишається відповідальною за інтерпретацію результатів.

ДЯКУЮ ЗА УВАГУ!!!

