

ДОДАТОК А

Переклад промπτу LLM

Ви експерт з візуалізації даних. Проаналізуйте наступні дані у форматі `{dataFormat}` та створіть інформативні візуалізації.

ДАНІ:

```
{JSON.stringify(truncatedData, null, 2)}
```

ІНСТРУКЦІЇ:

1. Проаналізуйте дані для виявлення шаблонів, тенденцій та інсайтів.
 2. Створіть комплексну інформаційну панель візуалізації з кількома компонентами.
 3. Поверніть масив JSON-конфігурацій віджетів, які можуть бути відтворені нашим фронтендом.
 4. Кожен віджет повинен включати функцію `dataTransform`, яка видобуває/обробляє конкретні дані, необхідні для цього віджета.
 5. `{languageInstruction}`
- `{truncated ? "Примітка: Дані було скорочено через обмеження розміру. Аналіз базується на підмножині даних." : ""}`

Доступні типи віджетів:

1. Віджет діаграми:

```
{  
  "type": "chart",  
  "chartType": "bar" | "line" | "pie" | "area" | "scatter",  
  "title": "Назва діаграми",  
  "description": "Необов'язковий опис",  
  "dataKeys": ["ключ1", "ключ2"], // Ключі даних для візуалізації  
  "xAxisKey": "категорія", // Ключ для осі X  
  "colors": ["#колір1", "#колір2"], // Необов'язкові кольори  
  "height": 300, // Необов'язкова висота  
  "insights": "Текст, що пояснює інсайти з цієї діаграми",  
  "dataTransform": "return data.map(item => ({ ...item, value:  
item.value * 2 }))" // Функція трансформації, яка отримує повний набір  
даних та повертає дані для цього віджета  
}
```

2. Віджет таблиці:

```
{  
  "type": "table",  
  "title": "Назва таблиці",  
  "description": "Необов'язковий опис",  
  "columns": [  
    { "key": "name", "label": "Ім'я" },  
    { "key": "value", "label": "Значення" }  
  ],  
  "caption": "Необов'язковий підпис таблиці",  
  "insights": "Текст, що пояснює інсайти з цієї таблиці",  
  "dataTransform": "return data.sort((a, b) => b.value -  
a.value).slice(0, 10)" // Функція трансформації, яка отримує повний набір  
даних та повертає дані для цього віджета  
}
```

3. Віджет інсайтів:

```
{  
  "type": "insights",  
  "title": "Ключові інсайти",  
  "description": "Необов'язковий опис",  
}
```

```

    "insights": ["Інсайт 1", "Інсайт 2", "Інсайт 3"], // Масив інсайтів
    або рядок
    "insightType": "list" | "text" // Відображення як список або абзац
  }

```

4. Віджет метрик:

```

{
  "type": "metrics",
  "title": "Ключові метрики",
  "description": "Необов'язковий опис",
  "metrics": [
    {
      "label": "Загальні продажі",
      "value": "$10,000",
      "change": { "value": 5.2, "positive": true } // Необов'язковий
індикатор зміни
    }
  ],
  "dataTransform": "return { totalSales: data.reduce((sum, item) =>
sum + item.sales, 0) }" // Функція трансформації, яка отримує повний набір
даних та повертає дані для обчислення метрик
}

```

5. Віджет зведення:

```

{
  "type": "summary",
  "title": "Зведення даних",
  "content": "Текст, що підсумовує весь набір даних та ключові
висновки"
}

```

ВАЖЛИВІ РЕКОМЕНДАЦІЇ:

1. Для трансформацій даних надайте дійсний текст функції JavaScript, який приймає вихідні дані як вхідні та повертає перетворені дані.
2. Переконайтеся, що всі дані правильно форматовані для типу віджета.
3. Для діаграм переконайтеся, що структура даних відповідає тому, що очікує тип діаграми.
4. Надайте змістовні інсайти для кожної візуалізації.
5. Поверніть ТІЛЬКИ дійсний масив JSON-конфігурацій віджетів.
6. КОЖЕН віджет повинен включати функцію dataTransform, навіть якщо вона просто повертає оригінальні дані.

ФОРМАТ ВІДПОВІДІ:

```

[
  {віджет1},
  {віджет2},
  ...
]

```

ДОДАТОК Б

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Дата звіту 6/5/2025

Дата редагування ---



Звіт не був оцінений

Звіт подібності

метадані

Назва організації

Kharkiv National University of Radio Electronics

Заголовок

2025_M_ПІ_ІПЗ-23-3_Будник_M_O_скорочений

Автор

Науковий керівник / Експерт

Будник Максим Олексійович Голян В.В./Нечволод В.Ю.

підрозділ

каф. ПІ

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



КП 1



КЦ

25

Довжина фрази для коефіцієнта подібності 2

8287

Кількість слів

65499

Кількість символів

Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		0
Інтервали		0
Мікропробіли		0
Білі знаки		0
Парафрази (SmartMarks)		2

Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз

Колір тексту

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	Барипо 12/13/2024 National University "Lviv Politechnika" (NULP2)	10 0.12 %
2	https://www.akademprostir.com/innovatsijni-pidkhodi-do-vizualizatsiji-danikh-u-naukovikh-stattakh-vid-infografiki-do-vr/	9 0.11 %
3	https://www.akademprostir.com/innovatsijni-pidkhodi-do-vizualizatsiji-danikh-u-naukovikh-stattakh-vid-infografiki-do-vr/	9 0.11 %

ДОДАТОК В
Слайди презентації

1

ДОСЛІДЖЕННЯ МЕТОДІВ ВІЗУАЛІЗАЦІЇ ТА АНАЛІЗУ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ LLM

Будник М.О. ІПЗм-23-3
Керівник: Доц. Голян В.В.

ХНУРЕ | 2025

2

АКТУАЛЬНІСТЬ ТЕМИ ДОСЛІДЖЕННЯ

- Експоненційне зростання обсягів даних вимагає ефективних інструментів для їх аналізу та візуалізації.
- Існуючі рішення (Tableau, Power BI) часто орієнтовані на технічних користувачів, мають складний інтерфейс або обмежену підтримку автоматизованого аналізу.
- Це обмежує доступність аналітики для широкого кола користувачів, особливо без навичок програмування.
- Великі мовні моделі (LLM) пропонують потенціал для автоматизації цих процесів, але їх інтеграція в аналітичні системи залишається недостатньо дослідженою.
- Автоматизація аналізу за допомогою LLM може спростити отримання інсайтів та зробити аналітику доступною для некваліфікованих користувачів.

3

МЕТА ТА ЗАДАЧІ ДОСЛІДЖЕННЯ

- **Мета роботи:** Дослідження принципів застосування LLM для аналізу та візуалізації даних, розробка концептуальної архітектури та програмного прототипу такої системи, дослідження методів генерації інсайтів та візуалізацій, порівняння ефективності різних моделей LLM (зокрема GPT-4o, Claude 3, Gemini Pro) та дослідження оптимальних технік побудови промптів для даної задачі.
- **Задачі дослідження:**
 - Проаналізувати існуючі підходи до візуалізації та аналізу даних.
 - Розробити архітектуру прототипу системи для аналізу даних з різних джерел (CSV, JSON) та автоматизованого створення візуалізацій за допомогою LLM.
 - Дослідити методи аналізу та візуалізації інформації за допомогою LLM.
 - Описати алгоритми аналізу та візуалізації даних з використанням LLM.
 - Дослідити проблеми використання LLM для аналізу даних (конфіденційність, достовірність, обмеження).
 - Провести порівняльний аналіз використання різних LLM.

4

ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

- **Об'єкт дослідження:** Процес дослідження та розробки системи для візуалізації та аналізу даних з використанням великих мовних моделей (LLM). (Альтернативно з ВСТУПУ: Процес обробки, аналізу та візуалізації даних з використанням великих мовних моделей)
- **Предмет дослідження:** Методи візуалізації та аналізу інформації, що використовують LLM для автоматизації процесу пошуку інсайтів та візуалізації, а також теоретичні основи її функціонування.

НАУКОВА НОВИЗНА

- Розроблено методику генерації інтерактивних візуалізацій на основі динамічних промптів, адаптованих під специфіку даних.
- Встановлено залежність між якістю промптів та точністю результатів аналізу.

ОГЛЯД ІСНУЮЧИХ РІШЕНЬ ТА ЛІТЕРАТУРИ

- **Існуючі інструменти:**
 - Datadog, Grafana, Google Data Studio, Microsoft Power BI, Tableau.
 - Переваги: Потужні можливості візуалізації, інтеграції.
 - Недоліки: Орієнтація на технічних користувачів, складність налаштування, обмежений автоматизований аналіз, вартість.
- **Літературний огляд:**
 - Принципи роботи LLM (Transformer, attention mechanisms), їх застосування в аналізі даних.
 - Порівняння LLM: Gemini (мультимодальність), GPT (обробка тексту), Llama (відкритість, ефективність).
 - Візуалізація даних на JavaScript (D3.js, Chart.js, Plotly.js).
 - Методи аналізу даних з LLM (NLP, ML).
 - Актуальність підтверджена аналізом джерел, виявлено потребу в дослідженні інтеграції LLM та prompt engineering для аналітичних систем.

ПОСТАНОВКА НАУКОВО-ТЕХНІЧНОЇ ЗАДАЧІ

Розробити інструмент автоматизованого аналізу даних із застосуванням LLM, що забезпечує:

- Підтримку структурованих (CSV, JSON) і текстових форматів даних.
- Генерацію інсайтів і візуалізацій за допомогою природної мови.
- Вбудовану перевірку достовірності результатів (зважаючи на "галюцинації" LLM).
- Порівняльний аналіз декількох LLM (GPT-4, Claude 3, Gemini Pro) з використанням об'єктивних метрик точності, вартості та часу відгуку.

Обмеження дослідження:

- Аналіз 3-х LLM (GPT-4o, Claude 3 Opus, Gemini 1.5 Pro).
- Тестування на публічних датасетах.
- Фокус на текстових та структурованих даних.

ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ: LLM В АНАЛІЗІ ДАНИХ

Архітектурні особливості LLM (GPT-4, Claude 3, Llama 3):

- Базуються на архітектурі трансформерів з механізмами уваги.
- Дозволяють аналізувати зв'язки в структурованих даних, генерувати семантичні уявлення для неструктурованих текстів, адаптуватися до різних мовних шаблонів.

Обробка структурованих даних (CSV, JSON):

- Дані серіалізуються в текстовий формат.
- Важливість формату вводу та "structured prompting" для точності.
- Використання режимів виводу JSON або "function calling" для гарантування коректності структури.

Аналіз текстових даних:

- Класифікація тексту, аналіз настрою (few-shot, zero-shot).
- Розпізнавання іменованих сутностей (NER) – складніше для загальних LLM порівняно зі спеціалізованими моделями.

ПОРІВНЯННЯ LLM ТА РОЛЬ ПРОМПТ-ІНЖИНІРИНГУ

Порівняльні характеристики LLM:

- GPT-4o (OpenAI): Економічність, високі оцінки на бенчмарках MMLU, MGSM, HumanEval.
- Claude 3 (Anthropic): Надійність, мультидисциплінарність, обробка візуальної інформації (Opus, Sonnet, Haiku).
- Gemini (Google): Масштабованість, підтримка довгого контексту (до 1M tokenів у Gemini 1.5 Pro).
- Вибір залежить від балансу "швидкість-якість-ціна" та специфіки завдання.

Роль prompt engineering:

- Ключовий для підвищення продуктивності LLM.
- Техніки: Zero-shot, Few-shot, Chain-of-Thought (CoT), system prompts.
- CoT значно покращує результати у складних задачах.
- Правильна побудова промпту критична для точності та уникнення "галюцинацій".

Обмеження LLM та проблема "галюцинацій":

- Обмежений розмір контекстного вікна.
- Негарантована достовірність відповідей ("галюцинації").
- Вартість та ресурсоемність.
- Стратегії боротьби: Retrieval-Augmented Generation (RAG), самодіагностика, заохочення чесності.

РОЗРОБКА ПРОТОТИПУ: АРХІТЕКТУРА ТА ТЕХНОЛОГІЇ

Розроблено веб-додаток для аналізу даних з використанням LLM.

Архітектура: Модульна, забезпечує гнучкість інтеграції різних LLM.

Технологічний стек:

- Next.js: Фреймворк для фронтенду та бекенду (серверний рендеринг, API-маршрути).
- AI SDK (Vercel AI SDK): Уніфікована взаємодія з LLM (OpenAI, Anthropic, Google).
- ShadcnUI: Створення користувацького інтерфейсу.
- Recharts: Побудова інтерактивних графіків.
- Zod: Валідація схем даних та типобезпека.

ФУНКЦІОНАЛЬНІ МОЖЛИВОСТІ

11

Домашня сторінка: Відображення карток раніше створених візуалізацій.

Створення нової візуалізації:

- Завантаження файлу даних (JSON, CSV, до 10 МБ).
- Вибір LLM (GPT-4o, Claude 3 Opus, Gemini 1.5 Pro).
- Вибір мови результатів (українська, англійська).
- Автоматичне скорочення даних при перевищенні лімітів токенів моделі.

Сторінка візуалізації та аналізу даних:

- Динамічно згенеровані віджети: графіки (різні типи), таблиці, текстові інсайти, метрики, загальний висновок.
- Кожен віджет супроводжується заголовком, описом та текстовими інсайтами.
- Можливість перегляду та завантаження вихідних даних.

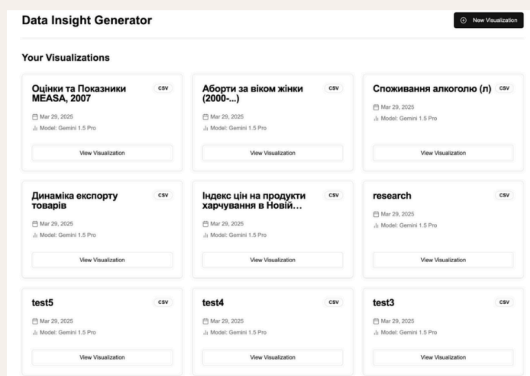
Взаємодія з LLM API:

- Обробка та підготовка даних (csv-parse, truncateData).
- Формування промптів (заголовковий, для аналізу даних).
- Використання Zod для валідації відповідей LLM.
- AI SDK для уніфікованої взаємодії з OpenAI, Anthropic, Google.
- Метод generateObject для отримання структурованих відповідей.
- Динамічна трансформація даних для віджетів (JavaScript-функції dataTransform на клієнті).

ДЕМОНСТРАЦІЯ ПРОТОТИПУ

12

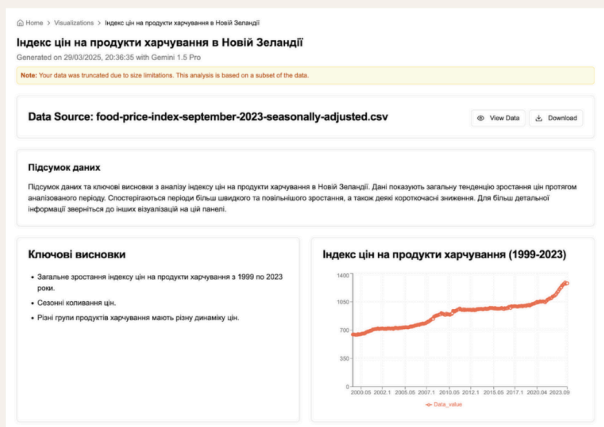
- **Домашня сторінка:** Огляд створених візуалізацій, швидкий доступ до результатів.
- **Діалогове вікно створення візуалізації:** Завантаження даних, вибір моделі LLM та мови.



СТОРІНКА ВІЗУАЛІЗАЦІЇ

13

- Відображення результатів аналізу у вигляді інтерактивних віджетів: графіки, таблиці, інсайти.
- Інформаційна панель: назва, дата, модель, індикатор скорочення даних.
- Приклад віджетів: підсумок даних, ключові висновки, графік (напр., індекс цін).



14

ПОБУДОВА ПРОМПТУ ДЛЯ LLM

Розроблено ефективний промпт для керування LLM у процесі аналізу та візуалізації даних. Ключові компоненти промпту:

1. Встановлення ролі та контексту: "Ви експерт з візуалізації даних".
2. Надання вхідних даних: Дані у форматі JSON/CSV, повідомлення про скорочення (якщо є).
3. Чіткі інструкції: Проаналізувати дані, створити дашборд, повернути JSON-конфігурації віджетів, забезпечити функцію dataTransform, використовувати вказану мову.
4. Детальний опис шаблонів віджетів: Для діаграм, таблиць, інсайтів, метрик, зведення, з описом поля dataTransform.
5. Важливі рекомендації: Щодо валідності JavaScript, форматування даних, змістовних інсайтів, повернення лише валідного JSON.
6. Чіткий формат відповіді: Приклад очікуваного масиву JSON-об'єктів.

Функціональні можливості промпту: Комплексний аналіз, підтримка різних типів візуалізацій, програмний компонент для трансформації даних.

РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ

- Прототип тестовано на різних наборах даних (продажі, метрики активності, часові ряди, категоріальні дані).

Порівняння LLM:

- Claude 3 Opus: Найкращі результати в генерації змістовних текстових інсайтів. Найкращий вибір доречних типів візуалізацій.
- GPT-4o: Висока точність у математичних обчисленнях. Генерація коректних функцій трансформації даних (dataTransform).
- Gemini 1.5 Pro: Переваги у швидкості обробки великих наборів даних. Іноді генерує менш структуровані текстові описи.

Ключове спостереження:

- Точність і корисність аналізу значно залежать від якості промптів та структури вихідних даних. Прототип успішно генерує візуалізації та інсайти.

ВИСНОВКИ

- Проведено аналіз предметної галузі, виявлено актуальність автоматизації аналізу даних за допомогою LLM.
- Систематизовано теоретичні основи використання LLM (Gemini, GPT, Llama), методів візуалізації (D3.js, Recharts) та промпт-інженерії.
- Розроблено концептуальну архітектуру та програмний прототип системи на базі Next.js, AI SDK, Zod, Recharts. Прототип реалізує завантаження даних (CSV, JSON), автоматичну генерацію візуалізацій та інсайтів.
- Експериментально досліджено ефективність LLM: Claude 3 Opus (інсайти, типи візуалізацій), GPT-4o (математика, трансформації), Gemini 1.5 Pro (швидкість).
- Обґрунтовано ефективні методи взаємодії з LLM, розроблено деталізований промпт.
- Підтверджено гіпотезу про ефективність використання LLM для спрощення та автоматизації аналізу даних, особливо для користувачів без технічних навичок.

ДЯКУЮ ЗА УВАГУ!

Питання?

ДОДАТОК Г
Апробація результатів роботи



Громадська організація «Молодіжна наукова ліга».
Номер запису в Реєстрі громадських об'єднань: 1506433.
Адреса: вул. Зодчих, буд. 40, офіс 103; м. Вінниця, Вінницька обл., 21037
Організація функціонує як відокремлений підрозділ ТОВ «UKRLOGOS Group».
ЄДРПОУ: 44574526
ІВАН: UA433052990000026002046104529
Банк ВФ АТ КБ «ПриватБанк»; МФО 44574526
Свідоцтво суб'єкта видавничої справи: ДК № 7860 від 22.06.2023.

Д О В І Д К А
ПРО ПРИЙНЯТТЯ СТАТТІ ДО ПУБЛІКАЦІЇ

22.05.2025

Шановний(і) автор(и):

Будник Максим Олексійович,

Редакційний комітет з радістю повідомляє, що стаття «Дослідження методів візуалізації та аналізу інформації з використанням LLM» прийнята до публікації в № 21 студентського наукового журналу «UNIVERSUM», випуск якого заплановано на 20 червня 2025 року.

Опублікована стаття буде доступна з 20.06.2025 за посиланням:

<https://archive.liga.science/index.php/universum/issue/view/june2025>

.....

Електронні сертифікати про публікацію та подяки науковим керівникам також будуть доступні з 20 червня. Розсилка замовлених друкованих примірників, сертифікатів та подяк відбудеться з 3 по 10 липня.

З повагою,

Директор Молодіжної наукової ліги
Голова редакційного комітету
ІГОР КОРЕНЮК



ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008: 2015

1

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)програмної інженерії
(кафедра)ПІЗМ-23-3 □
(група)

Максим БУДНИК

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
7.1.5	Звіт друкують шрифтом Times New Roman чорного кольору прямого накреслення через півтора міжрядкові інтервали кеглем 14. Розмір шрифту для написання заголовків у рядках і колонках таблиць і пояснювальних даних на рисунках і в таблицях встановлює виконавець звіту.	35
7.1.25	Не дозволено розміщувати назву розділу, підрозділу, а також пункту й підпункту на останньому рядку сторінки.	35, далі за текстом
	7.3 Нумерація сторінок звіту	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

Експерт

(підпис)

Вадим НЕЧВОЛОД

(прізвище, ініціали)

04.06.2025