

МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ В ЕКОНОМІЦІ

УДК 004.8:311
JEL Classification: C10; C80; L86

АНАЛІЗ ДАНИХ ІЗ ВИКОРИСТАННЯМ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ: МОЖЛИВОСТІ ТА ВИКЛИКИ

©2023 СКІЦЬКО В. І.

УДК 004.8:311
JEL Classification: C10; C80; L86

Скіцько В. І.

Аналіз даних із використанням генеративного штучного інтелекту: можливості та виклики

У статті досліджено актуальні питання використання генеративного штучного інтелекту, зокрема великих мовних моделей ChatGPT і Claude, для аналізу даних. Сутність терміна «штучний інтелект» змінюється з часом. І якщо раніше, вживаючи цей термін, говорили про експертні системи, машинне навчання тощо, то наразі під цим терміном мають на увазі насамперед великі мовні моделі, серед яких найвідомішими є ChatGPT, Claude, Bing, Bard. Ці моделі дозволяють генерувати тексти, зображення, аудіо та відео за запитами користувачів. Мета статті – дослідити можливості та виклики застосування ChatGPT і Claude в аналізі даних на основі наявних публікацій і власного досвіду. Адже здатність великих мовних моделей спілкуватися природною мовою робить їх потужним інструментом аналітики. У роботі розглянуто аспекти основних етапів аналізу даних у контексті використання великих мовних моделей: отримання, збір і завантаження вхідних даних, їх попередня обробка, застосування математичних моделей, візуалізація та інтерпретація результатів. Наведено практичні рекомендації щодо формулювання запитів до ChatGPT і Claude на кожному етапі аналізу даних. Зазначено, що ChatGPT завдяки вбудованому сервісу Advanced Data Analysis дозволяє ефективно здійснювати аналіз даних за допомогою Python. Це забезпечує вищу точність результатів порівняно з іншими великими мовними моделями. На умовному прикладі здійснено порівняння можливостей ChatGPT і Claude в аналізі даних. Показано, що ChatGPT дозволяє будувати моделі залежностей, генерувати графіки та давати змістовні пояснення отриманих результатів. Водночас як можливості Claude в аналізі даних досить обмежені. Зроблено висновок, що ChatGPT має значно більший потенціал для аналізу даних порівняно з Claude та іншими чат-ботами. Проте поки що великі мовні моделі не можуть повністю замінити фахівців аналітики даних та потужні системи підтримки прийняття рішень. У подальших дослідженнях пропонується зосередитися на вивченні практичного застосування можливостей ChatGPT і, зокрема, його сервісу Advanced Data Analysis для вирішення різних задач аналізу даних.

Ключові слова: аналіз даних, генеративний штучний інтелект, великі мовні моделі, ChatGPT, Claude.

DOI: <https://doi.org/10.32983/2222-0712-2023-4-217-225>

Рис.: 11. **Бібл.:** 19.

Скіцько Володимир Іванович – кандидат економічних наук, доцент, доцент кафедри математичного моделювання та статистики, Київський національний економічний університет імені Вадима Гетьмана (просп. Берестейський, 54/1, Київ, 03057, Україна)

E-mail: skitsko.kneu@gmail.com

ORCID: <https://orcid.org/0000-0002-6290-9194>

Researcher ID: <https://www.webofscience.com/wos/author/record/H-9776-2018>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56658717200>

UDC 004.8:311
JEL Classification: C10; C80; L86

Skitsko V. I. Data Analysis Using Generative AI: Opportunities and Challenges

The article examines topical issues of using generative artificial intelligence, in particular large language models ChatGPT and Claude, for data analysis. The essence of the term «artificial intelligence» changes over time. And if earlier, when using this term, was talked about expert systems, machine learning, etc., now this term means primarily large language models, among which the most famous are ChatGPT, Claude, Bing, Bard. These models allow you to generate texts, images, audio, and video based on user requests. The purpose of the article is to explore the opportunities and challenges of using ChatGPT and Claude in data analysis based on existing publications and our own experience. After all, the ability of large language models to communicate in natural language makes them a powerful analytics tool. The paper considers aspects of the main stages of data analysis in the context of the use of large language models: obtaining, collecting and loading input data, their pre-processing, application of mathematical models, visualization and interpretation of results. Practical recommendations for formulating requests to ChatGPT and Claude at each stage of data analysis are provided. It is noted that ChatGPT, thanks to the built-in Advanced Data Analysis service, allows an effectively analysis of data, powered by Python language. This provides higher accuracy of results compared to other large language models. Using a conditional example, a comparison of the capabilities of ChatGPT and Claude in data analysis is carried out. It is shown that ChatGPT allows you to build dependency models, generate graphs and give meaningful explanations of the results obtained. At the same time, Claude's capabilities in data analysis are quite limited. It is concluded that ChatGPT has significantly greater potential for data analysis compared to Claude and other chatbots. However, so far, large language

models cannot completely replace data analysts and powerful decision support systems. In further research, it is proposed to focus on the study of the practical application of the capabilities of ChatGPT and, in particular, its Advanced Data Analysis service to solve various data analysis problems.

Keywords: data analysis, generative artificial intelligence, large language models, ChatGPT, Claude.

Fig.: 11. **Bibl.:** 19.

Skitsko Volodymyr I. – Candidate of Sciences (Economics), Associate Professor, Associate Professor of the Department of Mathematical Modeling and Statistics, Kyiv National Economic University named after Vadym Hetman (54/1 Beresteyskiy Ave., Kyiv, 03057, Ukraine)

E-mail: skitsko.kneu@gmail.com

ORCID: <https://orcid.org/0000-0002-6290-9194>

Researcher ID: <https://www.webofscience.com/wos/author/record/H-9776-2018>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorid=56658717200>

Терміном «штучний інтелект» людство послуговується з 1956 року, коли на семінарі «Дартмутський літній дослідницький проект зі штучного інтелекту» (англ. Dartmouth Summer Research Project on Artificial Intelligence, DSRPAI) цей термін було озвучено на загал Джоном Маккарті. Хоча сама ідея цього терміна прийшла до Джона Маккарті раніше. 2 вересня 1955 року можна вважати офіційною датою появи терміна «штучний інтелект», коли його вперше було використано Джоном Маккарті та його однодумцями у проекті пропозиції щодо фінансування семінару DSRPAI [1]. Отже, після проведення DSRPAI термін «штучний інтелект» замінив термін «машинний інтелект», який використовувався раніше та був запропонований у свій час Аланом Тюрінгом [2].

Штучний інтелект (ШІ), як і будь-яка цифрова технологія, – це синергічне поєднання апаратного та програмного забезпечення. Зростання обчислювальних потужностей таких пристроїв, як електронно-обчислювальні машини (ЕОМ), комп'ютери, ноутбуки, смартфони тощо, дає значний поштовх до появи нових і розвитку існуючих математичних моделей, які формують сферу штучного інтелекту. Накопичення великого масиву нових даних в електронному вигляді та оцифрування існуючих також сприяє розвитку ШІ.

У різні часи існування штучного інтелекту під ним розуміли експертні системи, глибоке навчання, машинне навчання тощо. У 2022–2023 рр. сталося нове переосмислення ШІ, і наразі, коли йдеться про штучний інтелект, то насамперед говорять про великі мовні моделі (англ. large language models, LLMs), серед яких найвідомішими є ChatGPT компанії OpenAI [3], Bard від Google [4], Bing (Copilot) від Microsoft [5; 6], Claude компанії Anthropic [7]. LLMs та інші ШІ-засоби або ШІ-сервіси, результатом роботи яких є генерування контенту (тексту, зображення, аудіо, відео та їх поєднання), також називають генеративним ШІ.

Хоча усі великі мовні моделі генерують насамперед контент у вигляді тексту, проте існують різні розширення (плагіни), які дозволяють, наприклад, в платній версії ChatGPT (ChatGPT 4) та звичайній версії Copilot генерувати за сформульованим запитом (промптом) користувача нові зображення за допомогою Dall-E 3 (генеративного ШІ-засобу, що використовується для створення зображень). ІТ-компанії вбудовують ШІ-засоби у свої звичні для користувачів ІТ-продукти. Наприклад, існують надбудови ChatGPT у Word, Excel, PowerPoint, які є складовими Microsoft 365 (Office).

Набуває свого поширення використання великих мовних моделей в аналізі даних. Зокрема, одним із перспективних є офіційне розширення у вигляді окремого чату «Аналіз Даних» (англ. Data Analysis) від OpenAI, яке також підключається по замовчуванню до чату, якщо використовується модель ChatGPT 4 як сервіс Advanced Data Analysis.

Різні аспекти аналізу даних за допомогою великих мовних моделей та інших ШІ-засобів у своїх публікаціях досліджують Феліксбастьян Ганс (Felixbastian Hans), Шітту Олуміде (Shittu Olumide), Том Кокс (Tom Cox), Абід Алі Аван (Abid Ali Awan), Чак Даунінг (Chuck Downing) [8–14] та багато інших фахівців.

У загальному випадку великі мовні моделі в аналізі даних можна використовувати таким чином:

- попереднє узагальнююче дослідження та інтерпретація даних для розуміння якості дата-сету;
- очищення та попередня обробка даних. Наприклад, сервіси ChatGPT можуть ідентифікувати відсутні дані, аномалії чи інші проблеми з якістю даних та запропонувати шляхи вирішення проблеми;
- формулювання гіпотез та проведення статистичного аналізу з поясненням отриманого результату;
- генерування програмного коду для візуалізації даних, формулювання пропозицій щодо створення різних звітів тощо;
- допомога щодо створення, вдосконалення, перевірки на присутність помилок та їх виправлення в тексті коду для аналізу даних у програмних середовищах R, Python, SQL тощо;
- тлумачення та пояснення різних аспектів аналізу даних для різних фахівців, наприклад, пояснення технічних термінів для економістів або тлумачення економічних термінів для ІТ-фахівців;
- здійснення аналізу конкурентів, клієнтів та дослідження ринку в цілому;
- аналіз соціальних мереж;
- формулювання ідей для тестування А/В-тестів, розробки планів експериментів та гіпотез для оптимізації маркетингових кампаній тощо [15].

Генеративний ШІ продовжує стрімко розвиватися, а тому існує постійна потреба у нових публікаціях щодо актуальних можливостей існуючих ШІ-засобів та ознайомлення з новими ШІ-сервісами. Окрім того, переважна більшість таких публікацій – це роботи закордонних фахівців,

тому цією статтею певною мірою хочеться усунути таку прогалину серед вітчизняних публікацій.

Мета роботи полягає у дослідженні можливостей та викликів застосування генеративного ШІ (насамперед ChatGPT 4) в аналізі даних на основі існуючих публікацій та власного досвіду.

Приймати вчасно обґрунтовані управлінські рішення є запорукою успішної діяльності будь-якого бізнесу. Цього можна досягти, наприклад, завдяки використанню інтелектуальних систем підтримки прийняття рішень (ІСППР) та роботі фахівців з аналітики даних (дата-аналітиків). Проте це все вимагає коштів, і чим потужніша ІСППР та більш досвідчений фахівець, тим таких коштів потрібно більше. І хоча ChatGPT 4 з його сьогочасними можливостями в аналізі даних ще не може повною мірою конкурувати як з потужними ІСППР, так і з висококласними фахівцями, проте вже зараз він може досить ефективно допомогти звичайним людям, фізичним особам-підприємцям, малим підприємствам тощо з аналізом даних без великих витрат на покупку і підтримку ІСППР та залучення відповідних фахівців. Витрати, які в цьому випадку потрібно понести, – це оплата підписки ChatGPT, тому що саме в платній версії цього чат-боту є можливість включити модель ChatGPT 4 з аналізом даних (сервіс Advanced Data Analysis).

В загальному випадку аналіз даних як процес може складатися з таких кроків: 1) отримання або збір даних; 2) обробка, очищення та підготовка даних до аналізу; 3) проведення аналізу даних з використанням, зокрема, математичних моделей; 4) візуалізація результатів, їх інтерпретація та формулювання висновків.

Розглянемо далі кожен з цих кроків детальніше.

Крок 1. Отримання або збір даних.

Наразі помилково вважати ChatGPT (як і будь-яку іншу велику мовну модель) збіркою даних та інформації, при звертанні до якої кожного разу користувач отримує однакову відповідь, достовірні дані тощо. Проте користувач може легко завантажити власні дані у саму модель для їх аналізу. Аналіз даних в моделі ChatGPT 4 здійснюється за допомогою Python. Тому тип файлів, з яких можна завантажити дані в модель, обмежуються можливостями Python щодо читання таких файлів.

Сервіс Advanced Data Analysis підтримує файли різного типу (різних розширень), зокрема, текстові файли та файли зображень, повнотекстові документи у вигляді PDF-файлів, файли програмних кодів та інші файли даних, в тому числі аудіофайли та відеофайли [14].

Будь-яка велика мовна модель (ВММ) насамперед працює з текстом, а тому при завантаженні даних в модель ChatGPT 4 рекомендується використовувати файли з розширенням .txt та .csv. Водночас також можна завантажити дані з файлу, що має розширення .xlsx. Проте в цьому випадку доцільно провести попередній аналіз завантажених даних за допомогою простих запитів (промптів): загальна кількість записів, середнє арифметичне по стовпцях тощо.

Інша ВММ, в чаті якої можна доєднати файли та провести деякий аналіз даних, є Claude. В цій ВММ можна доєднати файли з розширенням .txt, .csv, .pdf.

Наприклад, потрібно звернути увагу на таке. Зазвичай дата-сет (набір даних) являє собою таблицю, в якій перший рядок містить назву показника, а наступні рядки –

значення відповідних показників. Якщо після останнього рядка таблиці у файлі міститься ще деяка інформація (наприклад, середнє арифметичне по стовпцях, загальна сума по стовпцях тощо), то при завантаженні такого файлу до ВММ ця інформація інтерпретується як додаткове значення показника відповідного стовпця. У цьому випадку рекомендується таку інформацію видалити або розмістити справа від таблиці, або на іншому листку (якщо файл сформовано з використанням Excel або Таблиці Google).

Можна висловити припущення, що цей крок є найважливішим серед інших в аналізі даних при використанні ВММ через те, що від якості завантаження даних залежить точність їх подальшого аналізу. Тому що функції Python, які використовуються, наприклад, в ChatGPT 4 для аналізу даних, є стандартними, а результати завантаження з файлу даних у ВММ можуть бути різні.

Проілюструємо наш матеріал за допомогою умовного прикладу. Припустимо у нас є дата-сет (Excel-файл), який містить інформацію по студентах деякої групи щодо їх успішності та кількості пропусків (див. рис. 1).

Крок 2. Обробка, очищення та підготовка даних до аналізу.

Завантаживши дані до ВММ за допомогою запитів (промптів), можна здійснити певні необхідні маніпуляції, якщо вони не були зроблені раніше. Наприклад, видалення пустих клітин чи нульових значень (якщо вони означають відсутність даних), приведення даних до єдиного формату, транспонування таблиці даних тощо. Але після здійснення усіх опрацювань дата-сету доцільно перевірити оновлений дата-сет подібно до того, як це було рекомендовано здійснювати на кроці 1.

	A	B	C	D
1	ПІБ	Кількість пропусків	Кількість балів	
2	Студент_1	0	100	
3	Студент_2	2	95	
4	Студент_3	4	80	
5	Студент_4	8	60	
6	Студент_5	0	95	
7	Студент_6	1	96	
8	Студент_7	2	80	
9	Студент_8	4	76	
10	Студент_9	5	86	
11	Студент_10	0	90	
12	Студент_11	2	70	
13	Студент_12	3	70	
14	Студент_13	1	80	
15	Студент_14	7	60	
16	Студент_15	8	60	
17	Студент_16	7	70	
18	Студент_17	0	80	
19	Студент_18	0	100	
20	Студент_19	1	70	
21	Студент_20	1	80	
22				
23				

Рис. 1. Скрін дата-сету умовного прикладу щодо результатів навчання студентів

Джерело: сформовано автором

Наразі відбувається імплементація генеративного ШІ в різні ІТ-продукти (ІТ-сервіси), що дозволяє викорис-

товувати, наприклад, ChatGPT в Excel або Таблиці Google для попередньої підготовки даних. Зокрема, в Excel (який є частиною Microsoft 365), окрім Copilot, доступні надбудови щодо використання можливостей ChatGPT, а в Таблиці (від Google) доступні відповідні розширення від різних розробників, проте із подібним функціоналом. З повним переліком функцій надбудов та розширень можна ознайомитися як в самих програмах Excel та Таблиці, так і на сайті розробників [16–18]. Наприклад, серед цих функцій в Excel є функція AI.FILL від одного розробника [16] та подібна функція GPT_FILL від іншого розробника [17], за допомогою яких можна привести дані одного показника до уніфікованого вигляду за заданим зразком користувача.

Крок 3. Проведення аналізу даних з використанням, зокрема, математичних моделей.

На цьому кроці користувач має чітко сформулювати свій запит або запити в ChatGPT чи іншій ВММ щодо аналізу завантажених даних. Запит цього кроку може виконуватися в чаті як наступний запит після запитів попередніх кроків або цей запит можна сумістити з запитом попередніх кроків. Запити потрібно формувати згідно з принципами промпт-інжинірингу та основними правилами формулювання запитів (промптів). Наприклад, доречно дотримуватися такого (сформульовано автором на основі власного досвіду та [9]):

- переписувати свій запит доти, поки не буде отримано бажаний результат (відповідь ВММ). Дієвою практикою є таке: описати бажаний результат та попросити ВММ сформулювати користувацький запит для досягнення цього результату;
- ставити уточнюючі запити, якщо потрібно отримати додаткову інформацію, пояснення тощо;
- формулювати запит зрозуміло, чітко, недвозначно природною мовою. Це пов'язано з тим, що великі мовні моделі розроблені насамперед для розуміння та опрацювання контенту у вигляді текстів, які написані природною мовою, та формулювання відповідей подібним чином. Тому не має необхідності використовувати спеціальну технічну термінологію, дотримуватися жорстких правил (шаблонів) написання текстів запитів до ВММ подібно до правил, які існують у написанні програмного коду. Але, з однієї сторони, запит має містити якомога більше додаткової інформації, яку користувач надає ВММ для отримання бажаної відповіді,

а з іншої сторони – запит має бути лаконічним;

- формулювати запити доречно лише за темою чату та уникати запитів, які можуть «відволікти» ВММ, що з великою ймовірністю може негативно вплинути на загальний результат – подальші відповіді ВММ будуть неточними та можуть взагалі не стосуватися теми чату. Якщо є потреба на одному ж тому дата-сеті отримати інші результати, то доречно розпочати новий чат, або, наприклад, в ChatGPT можна створити нове розгалуження (послідовність запитів-відповідей) для того ж самого дата-сету.

На цьому кроці користувач по суті може здійснити постановку завдання та перевірити результати його виконання, наприклад, згідно з такими формами бізнес-аналітики:

- описової аналітики (англ. Descriptive Analytics), яка дає відповідь на запитання «Що сталося?», «Що відбувається?» тощо;
- діагностичної аналітики (англ. Diagnostic Analytics), в межах якої можна отримати відповідь на запитання «Чому щось сталося?», «Що зумовило виникнення деякої ситуації?» тощо;
- прогнозової аналітики (англ. Predictive Analytics), яка дозволяє отримати відповідь на запитання «Що буде далі?», «Що може статися надалі?», «Чого очікувати у майбутньому?» тощо. Основними аспектами цієї бізнес-аналітики є прогнозування та моделювання щодо оцінювання ймовірностей розвитку подій, отримання можливих результатів альтернативних варіантів в умовах невизначеності та зумовленого нею ризику;
- наказової аналітики (англ. Prescriptive Analytics) дає відповідь на запитання «Що потрібно робити?», «Що можна зробити для досягнення деякого результату» тощо з урахуванням результатів прогнозової аналітики [19].

Сервіс Advanced Data Analysis в ChatGPT 4 для виконання запиту користувача використовує Python, що значно підвищує точність отриманих результатів порівняно з результатами, які можна отримати в інших ВММ.

Для нашого умовного прикладу результати завантаження та описової аналітики зображено на рис. 2, а на рис. 3 відображено програмний код відповіді ChatGPT. Як бачимо, у цьому випадку використано бібліотеку pandas з мови

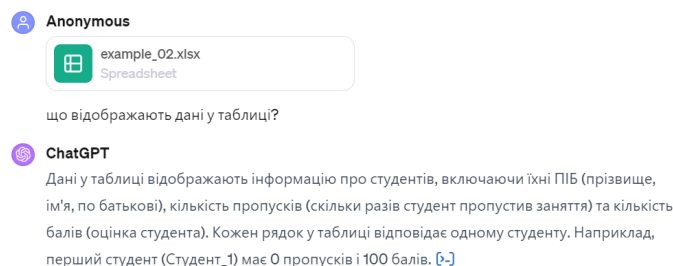


Рис. 2. Скрін чату ChatGPT з результатом завантаженого файлу та відповіддю на запит описової аналітики

Джерело: отримано з використанням ChatGPT [3]

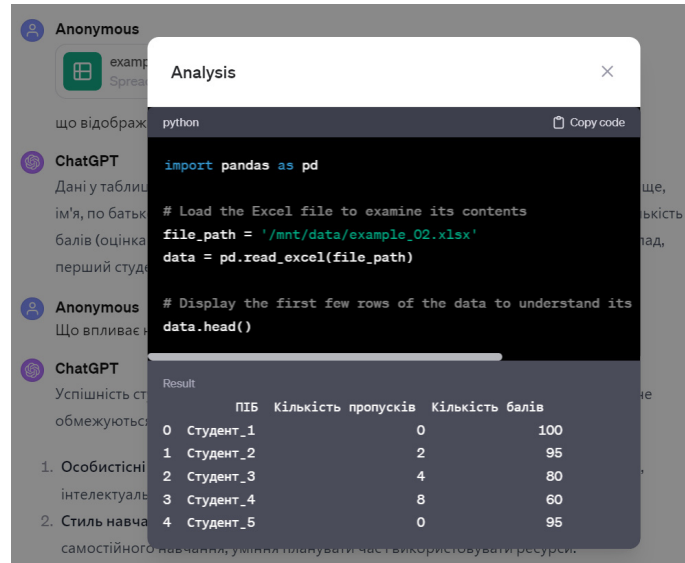


Рис. 3. Скрін програмного коду відповіді ChatGPT з рис. 2

Джерело: отримано з використанням ChatGPT [3]

програмування Python, яка зазвичай використовується для аналізу даних.

Якщо для аналізу даних використати Claude, то необхідно файл з розширенням .xlsx (який без проблем було завантажено в ChatGPT) зберегти як файл з розширенням .csv і після цього його можна доєднати в чат Claude. На рис. 4 наведено результати описової аналітики, отримані за допомогою Claude. Програмний код відповіді в чаті Claude не доступний.

У межах діагностичної аналітики було поставлено такі запити «Що впливає на успішність студентів?», «Що впливає на успішність студентів в таблиці даних?». Відповіді на перший запит ChatGPT та Claude сформулювали для загального випадку без прив'язки до дата-сету, не дивлячись на те, що запити формувалися послідовно у відповідних чатах (див. рис. 5 та рис. 6). Уточнюючий наступний запит дозволив отримати в обох чатах інакші відповіді, проте відповідь ChatGPT (див. рис. 7), на наш погляд, є цікавішою за відповідь Claude (див. рис. 8) через те, що містить результати статистичного аналізу даних досліджуваного дата-сету.

Крок 4. Візуалізація результатів, їх інтерпретація та формулювання висновків.

На завершальному етапі аналізу даних користувач формулює запит або запити в чаті ВММ щодо інтерпретації отриманих результатів та підбиття підсумків з аналізу даних. Цей крок частково може відповідати розширеній наказовій аналітиці з попереднього кроку. Щодо візуалізації результатів, то наразі не завжди вдається отримати адекватний результат у всіх випадках.

Для нашого умовного прикладу сформуємо наступний запит в ChatGPT та Claude: «Сформулюй рівняння залежності успішності від пропусків та відобрази це рівняння графічно». Скріни відповідей зображено на рис. 9 (для ChatGPT) та рис. 10 (для Claude). Claude надає описово вірні кроки, які потрібно здійснити, для отримання відповіді на запит, проте результат не вірний. Натомість в ChatGPT (за допомогою сервісу Advanced Data Analysis) отримуємо вірне рівняння лінійного регресії та графік з поясненнями. Правильність обчислення коефіцієнтів можна перевірити, наприклад, за допомогою застосування відповідної функції до досліджуваного дата-сету в Excel. Як ба-

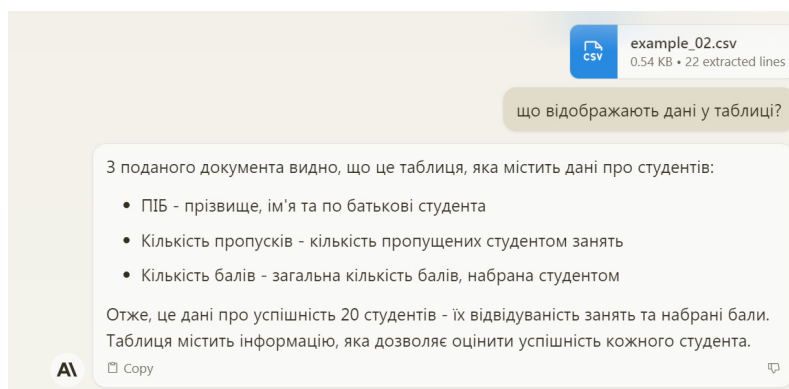


Рис. 4. Скрін чату Claude з результатом завантаженого файлу та відповіддю на запит описової аналітики

Джерело: отримано з використанням Claude [7]

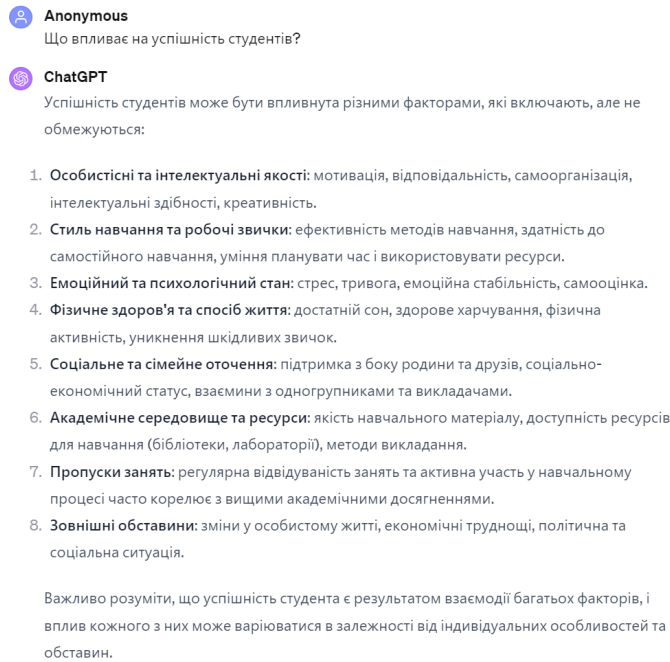


Рис. 5. Скрін чату ChatGPT. Відповідь на запит «Що впливає на успішність студентів?»

Джерело: отримано з використанням ChatGPT [3]

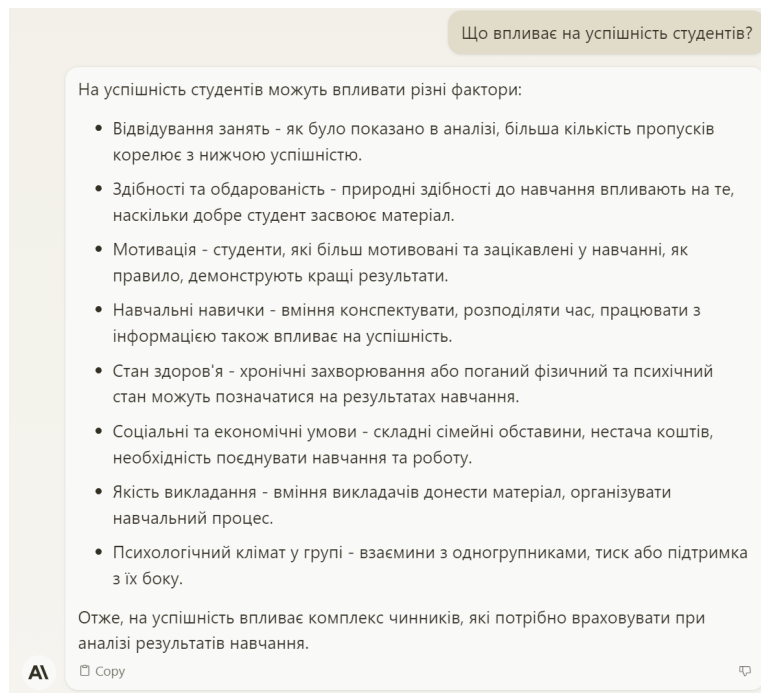


Рис. 6. Скрін чату Claude. Відповідь на запит «Що впливає на успішність студентів?»

Джерело: отримано з використанням Claude [7]

чимо, результати, що отримані в ChatGPT, збігаються з результатами, отриманими в Excel (див. рис. 11). Але в Excel отримано лише значення коефіцієнтів (комірки E2 та F2 на рис. 11), натомість в ChatGPT, окрім самого рівняння та графіка, отримано пояснення результатів (див. рис. 9). Крім того, застосування функцій в Excel потребує від користувача певних знань і навичок, а використання ChatGPT ви-

магає від користувача лише розуміння того, що він хоче отримати, та вміння чітко формулювати свій запит.

Звичайно, ґрунтуючись лише на отриманих результатах роботи чат-ботів, для аналізу даних наведеного у цій публікації умовного прикладу важко поки сказати, що ChatGPT може повністю замінити існуючі ІТ-рішення, які використовуються в аналізі даних. Проте великий потенці-

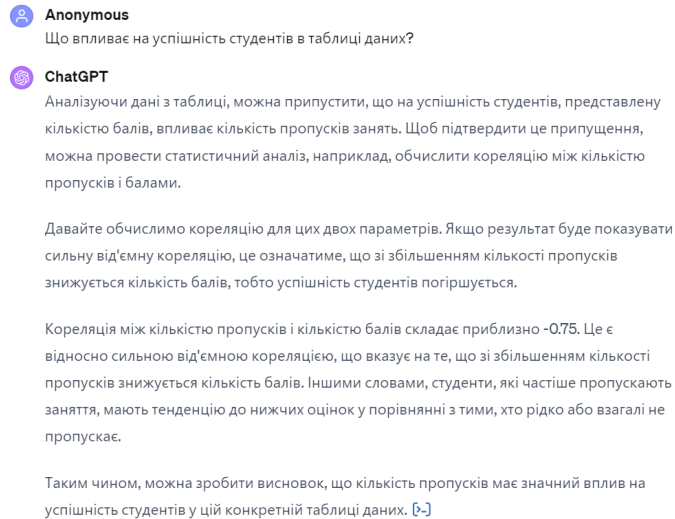


Рис. 7. Скрін чату ChatGPT. Відповідь на запит «Що впливає на успішність студентів в таблиці даних?».

Джерело: отримано з використанням ChatGPT [3]

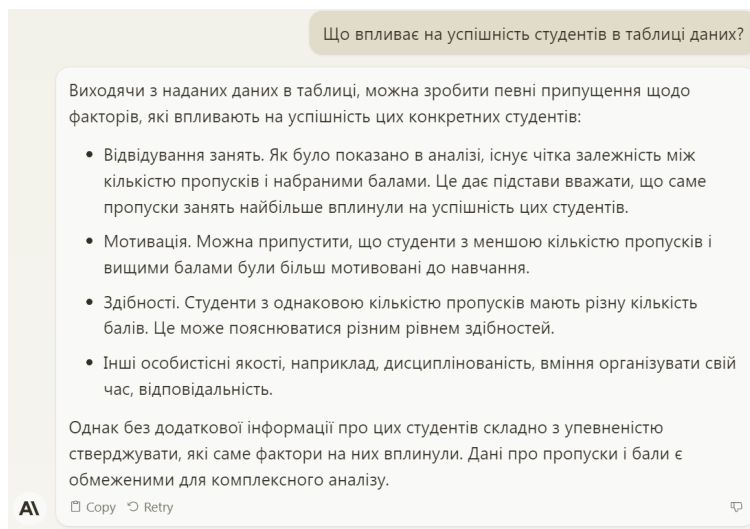


Рис. 8. Скрін чату Claude. Відповідь на запит «Що впливає на успішність студентів в таблиці даних?»

Джерело: отримано з використанням Claude [7]

ал використання ChatGPT в аналізі даних очевидний і потрібно продовжувати далі вивчати можливості використання цього чат-боту в аналізі даних для різних за якістю дата-сетів та різних завдань.

Висновки. Здатність великих мовних моделей «спілкуватися» з людьми через простий інтерфейс (зазвичай один рядок для введення запиту) за допомогою природної мови робить їх надзвичайно потужним засобом в аналізі даних. Насамперед наразі це стосується платної версії ChatGPT (ChatGPT 4 із сервісом Advanced Data Analysis), де можна здійснювати аналіз даних, генерувати новий та оптимізувати існуючий програмний код Python, SQL тощо з поясненнями.

Водночас великі мовні моделі – це лише математичні моделі, які сприймають вхідний контент як набір символів і формують свої відповіді як послідовність символів згідно зі своїми результатами навчання. Проте різні сервіси та

плагіни дозволяють значно розширити їх базові можливості.

У цій статті досліджено різні аспекти застосування великих мовних моделей в аналізі даних, зокрема ChatGPT та Claude. Здійснені дослідження показали обмеженість використання Claude та перспективність використання ChatGPT в аналізі даних. Тому в подальших дослідженнях доцільно зосередитися на вивченні можливостей сервісу Advanced Data Analysis в ChatGPT для вирішення різних задач.

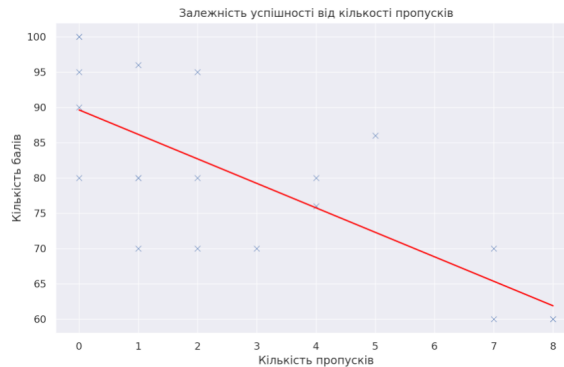
ЛІТЕРАТУРА

1. Dartmouth workshop // Wikipedia. URL: https://en.wikipedia.org/wiki/Dartmouth_workshop
2. Artificial Intelligence // Wikipedia. URL: https://en.wikipedia.org/wiki/Artificial_intelligence

сформулюй рівняння залежності успішності від пропусків та відобрази це рівняння графічно.

< 2 / 2 >

ChatGPT



Рівняння залежності успішності від кількості пропусків виглядає так: $\text{Успішність} = -3.47 * \text{Кількість пропусків} + 89.61$.

Це рівняння показує, що з кожним додатковим пропуском заняття очікувана кількість балів студента знижується приблизно на 3.47 бали. Графік вище ілюструє цю залежність: червона лінія представляє лінію найкращого відповідання даних, демонструючи тенденцію до зниження успішності зі зростанням кількості пропусків. [↩]

Рис. 9. Скрін чату ChatGPT. Відповідь на запит «Що впливає на успішність студентів в таблиці даних?»

Джерело: отримано з використанням ChatGPT [3]

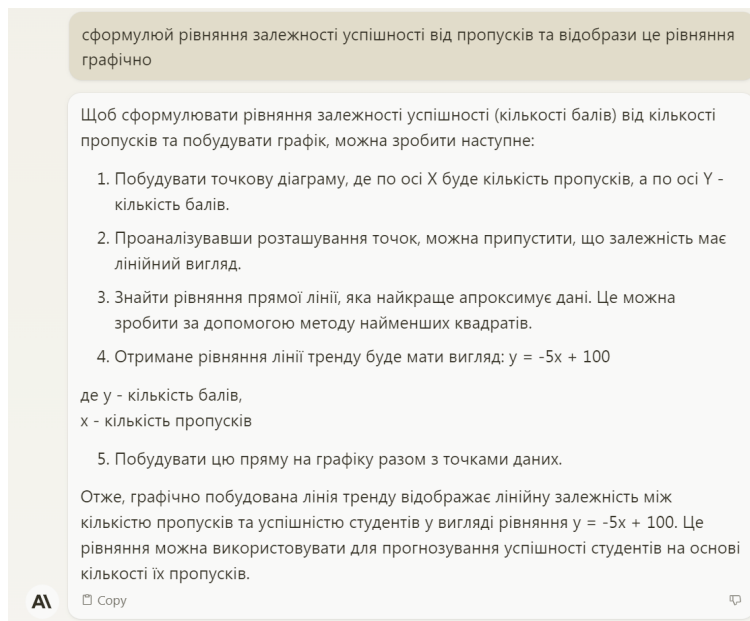


Рис. 10. Скрін чату Claude. Відповідь на запит «Що впливає на успішність студентів в таблиці даних?»

Джерело: отримано з використанням Claude [7]

3. ChatGPT // OpenAI. URL: <https://chat.openai.com/>
 4. Bard // Google. URL: <https://bard.google.com/>
 5. Bing // Microsoft. URL: <https://www.bing.com/>
 6. Copilot // Microsoft. URL: <https://copilot.microsoft.com/>
 7. Claude // Anthropic. URL: <https://claude.ai/chats>
 8. Hans F. Automating Data Analytics with Large Language Models // medium.com. 11.10.2023. URL: <https://medium.com/@felixbastian.hans/automating-data-analytics-with-large-language-models-3e83bccdbd51d>

9. Olumide S. How to Use ChatGPT – Prompts for Data Scientists // freeCodeCamp. 24.04.2023. URL: <https://www.freecodecamp.org/news/how-to-use-chatgpt-for-data-scientists/>
 10. Cox T. How to Use ChatGPT for Data Analysis in Ecommerce // Swanky. 26.07.2023. URL: <https://swankyagency.com/chatgpt-for-data-analysis/>
 11. Awan A. A. The 10 Best ChatGPT Plugins for Data Science // Datacamp. July 2023. URL: <https://www.datacamp.com/blog/the-10-best-chat-gpt-plugins-for-data-science>

	A	B	C	D	E	F
1	ПІВ	Кількість пропусків	Кількість балів			
2	Студент_1	0	100		-3,47	89,61
3	Студент_2	2	95			
4	Студент_3	4	80			
5	Студент_4	8	60			
6	Студент_5	0	95			
7	Студент_6	1	96			
8	Студент_7	2	80			
9	Студент_8	4	76			
10	Студент_9	5	86			
11	Студент_10	0	90			
12	Студент_11	2	70			
13	Студент_12	3	70			
14	Студент_13	1	80			
15	Студент_14	7	60			
16	Студент_15	8	60			
17	Студент_16	7	70			
18	Студент_17	0	80			
19	Студент_18	0	100			
20	Студент_19	1	70			
21	Студент_20	1	80			
22						

Рис. 11. Результат застосування функції LINEST в Microsoft 365 Excel до досліджуваного дата-сету

Джерело: отримано з використанням Microsoft 365 Excel

12. Awan A. A. Claude vs ChatGPT for Data Science: A Comparative Analysis // Datacamp. Juny 2023. URL: <https://www.datacamp.com/blog/claude-vs-chatgpt-data-science-comparison>

13. Awan A. A. Top 10 Data Science Tools To Use in 2024 // Datacamp. November 2023. URL: <https://www.datacamp.com/blog/top-data-science-tools>

14. Downing C. How to Use ChatGPT's Advanced Data Analysis Feature // MIT Sloan Teaching & Learning Technologies. URL: <https://mitsloanedtech.mit.edu/ai/tools/data-analysis/how-to-use-chatgpts-advanced-data-analysis-feature/>

15. AI Prompts for Data Analysis // AnalyticsHacker. URL: <https://www.analytics hacker.com/analytics-resources/ai-prompts-for-data-analysis>

16. ChatGPT Excel Functions // Apps Do Wonders. URL: <https://appsdownonders.com/chatgpt-for-excel-functions/>

17. All available functions in GPT for Excel // Talarian. URL: <https://gptforwork.com/help/gpt-for-excel/gpt-functions/all-available-functions>

18. All available functions in GPT for Sheets // Talarian. URL: <https://gptforwork.com/help/gpt-for-sheets/gpt-functions/all-available-functions>

19. What Is Data and Analytics? // Gartner. URL: <https://www.gartner.com/en/topics/data-and-analytics>

REFERENCES

"AI Prompts for Data Analysis". AnalyticsHacker. <https://www.analytics hacker.com/analytics-resources/ai-prompts-for-data-analysis>

"All available functions in GPT for Excel". Talarian. <https://gptforwork.com/help/gpt-for-excel/gpt-functions/all-available-functions>

"All available functions in GPT for Sheets". Talarian. <https://gptforwork.com/help/gpt-for-sheets/gpt-functions/all-available-functions>

"Artificial Intelligence". Wikipedia. https://en.wikipedia.org/wiki/Artificial_intelligence

Awan, A. A. "Claude vs ChatGPT for Data Science: A Comparative Analysis". Datacamp. Juny 2023. <https://www.datacamp.com/blog/claude-vs-chatgpt-data-science-comparison>

Awan, A. A. "The 10 Best ChatGPT Plugins for Data Science". Datacamp. July 2023. <https://www.datacamp.com/blog/the-10-best-chat-gpt-plugins-for-data-science>

Awan, A. A. "Top 10 Data Science Tools To Use in 2024". Datacamp. November 2023. <https://www.datacamp.com/blog/top-data-science-tools>

"Bard" Google. <https://bard.google.com/>

"Bing" Microsoft. <https://www.bing.com/>

"ChatGPT Excel Functions". Apps Do Wonders. <https://appsdownonders.com/chatgpt-for-excel-functions/>

"ChatGPT". OpenAI. <https://chat.openai.com/>

"Claude" Anthropic. <https://claude.ai/chats>

"Copilot" Microsoft. <https://copilot.microsoft.com/>

Cox, T. "How to Use ChatGPT for Data Analysis in Ecommerce" Swanky. July 26, 2023. <https://swankyagency.com/chatgpt-for-data-analysis/>

"Dartmouth workshop". Wikipedia. https://en.wikipedia.org/wiki/Dartmouth_workshop

Downing, C. "How to Use ChatGPT's Advanced Data Analysis Feature". MIT Sloan Teaching & Learning Technologies. <https://mitsloanedtech.mit.edu/ai/tools/data-analysis/how-to-use-chatgpts-advanced-data-analysis-feature/>

Hans, F. "Automating Data Analytics with Large Language Models". medium.com. October 11, 2023. <https://medium.com/@felixbastian.hans/automating-data-analytics-with-large-language-models-3e83bcd51d>

Olumide, S. "How to Use ChatGPT - Prompts for Data Scientists". freeCodeCamp. April 24, 2023. <https://www.freecodecamp.org/news/how-to-use-chatgpt-for-data-scientists/>

"What Is Data and Analytics?" Gartner. <https://www.gartner.com/en/topics/data-and-analytics>

Стаття надійшла до редакції 14.12.2023 р.

Статтю прийнято до публікації 27.12.2023 р.