УДК 004.056.5

ПОРІВНЯЛЬНИЙ АНАЛІЗ ШТУЧНОГО ІНТЕЛЕКТУ НА ОСНОВІ ІСНУЮЧИХ ЧАТ-БОТІВ

Кобилянська Олена[1], Єсіна Марина[1,2], Горбенко Юрій[2]

[1]Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна
kobol1801@gmail.com, m.v.yesina@karazin.ua
[2]АТ «ІІТ», вулиця Коломенська, 15, Харків, 61166, Україна
jsciitua@gmail.com

Надійшла: Листопад 2023. Прийнята: грудень 2023.

***Анотація**: У даній роботі представлено комплексний аналіз двох провідних систем штучного інтелекту (ШІ) – ChatGPT-4 від OpenAI та Bard від Google AI. Також наводиться огляд розвитку штучного інтелекту в різних галузях та його впливу на повсякденне життя людини, особливо в таких сферах, як медицина, фінанси, державне управління тощо. Проводиться заглиблення в детальне порівняння різних версій ChatGPT (GPT-3 та GPT-4), шляхом обговорення та аналізу їхніх можливостей, вдосконалення та обмежень. У статті також розглядається інтеграція системи Bard із сервісами Google, її унікальні функціональні можливості та останні оновлення. Мета дослідження полягає в порівнянні можливостей систем штучного інтелекту ChatGP-4T та Bard, висвітленні їхніх сильних і слабких сторін, а також їх практичного застосування. Проведено порівняльне тестування для оцінки продуктивності кожної моделі (системи) в різних завданнях, включаючи розв'язання логічного завдання, написання есе, аналіз із подальшим внесення пропозицій щодо покращення веб-сайту та написання коду HTML/CSS для веб-сторінки. Результати підкреслюють той факт, що, незважаючи на визнані переваги цих моделей, їхні функціональні характеристики іноді можуть бути обмежені або не відповідати очікуванням при виконанні специфічних завдань, а вибір системи (моделі) буде коригуватися у залежності від потреб користувачів.*

***Ключові слова:** ChatGPT-4, Bard, OpenAI, GoogleAI, штучний інтелект.*

1. Вступ

На сьогоднішній день, штучний інтелект (ШІ) швидко набуває популярності у різних секторах, включаючи корпоративний світ, бізнес-кола та повсякденне життя людей. Застосування ШІ в областях, як-от медицина, банківська сфера та урядові структури, стає все частішим. ШІ полегшує обробку даних, оскільки вона відбувається без втручання людської праці та зазвичай забезпечує точність виконаних завдань. Згідно зі статистикою, у 2023 році 35% компаній використовували ШІ у своїй діяльності, а 90% організацій вважають ШІ важливою для досягнення конкурентних переваг [1].

Системи штучного інтелекту впливають і на людське повсякдення, спрощуючи наступні аспекти їх діяльності: планування та організація денних справ, використання засобів ефективності у фінансах, навчанні та здоров'ї тощо. Завдяки йому, суспільство може ефективніше використовувати свій час, отримуючи доступ до швидкої та точної інформації.

Дана стаття зосереджена на аналізі особливостей двох провідних систем штучного інтелекту – *Bard* та *ChatGPT*. Вона включає в себе практичне порівняння однакових параметрів обох систем, а також виявлення переваг та недоліків кожної з них.

2. Огляд мовної моделі ChatGPT

ChatGPT, створена OpenAI, є системою генерації тексту, яка належить до серії *GPT* (*Generative Pretrained Transformer*). Базуючись на трансформерній архітектурі, ця модель навчена на великих масивах текстових даних для генерації даних подібних за стилем написання до тексту, створеним людиною. Розроблена для реагування на запити користувачів, *ChatGPT* підходить для використання у діалогових програмах, таких як чат-боти, обслуговування клієнтів та віртуальні асистенти. Ця модель була тренована на даних з різних джерел, таких як

 © Кобилянська О., Єсіна М., Горбенко Ю., 2023

Інтернет-ресурси, книги та соцмережі, що дозволяє їй створювати зв'язні та контекстуальні текстові відповіді. Щоб використовувати *ChatGPT*, користувач подає підказку, таку як питання або коментар, і модель генерує відповідь, враховуючи отримані дані та своє попереднє навчання. Однією з головних переваг *ChatGPT* є її здатність до контекстуально релевантного тексту. Наприклад, при запитанні про моду, модель може надати інформацію, що включає наступні слова: стиль, вбрання, крій. *ChatGPT* також може продовжувати діалог, використовуючи попередню розмову як контекст. *ChatGPT* також застосовується для інших задач, таких як відповіді на питання, узагальнення та класифікація тексту, завдяки доопрацюванням під конкретні цілі. Ця модель є частиною більш широкої тенденції використання великих мовних моделей для застосунків, що має потенціал перетворити спосіб взаємодії з технологіями та спілкування з пристроями на більш природній і інтуїтивно зрозумілий [2].

Вище було представлено загальний огляд моделі *ChatGPT*. Далі ми зосередимося на порівнянні двох версій цієї моделі: *ChatGPT-3*, що з'явилася у 2020 році, та *ChatGPT-4*, випущеної у 2023 році. Це дозволить нам визначити, яка з цих моделей краще підходить для порівняльного аналізу з моделлю *Bard*.

ChatGPT-3 вирізняється своєю високою здатністю до розуміння та створення текстів. Він навчений на обширному спектрі Інтернет-даних, що надає йому широкі знання. Ця модель ефективно виконує багато завдань, створюючи оригінальні тексти. Однак, вона може давати неточні відповіді та має тенденцію до упередженості, особливо у складних сценаріях (тобто, умовно кажучи - *може «галюцинувати»*).

ChatGPT-4, з іншого боку, покращив здатність розрізняти та відповідати на більш складні питання завдяки удосконаленій трансформерній архітектурі. Модель отримала більше навчальних даних і зменшила частоту помилок порівняно з попередніми версіями. *ChatGPT-4* вирішує складні завдання точніше та надійніше, показуючи краще розуміння контексту. Також, до функціоналу системи був доданий наступний функціонал: обробка та генерація графічних зображень, додаткові утиліти на обробку файлів обсягом більш, ніж 50 сторінок. Однак, попри поліпшення, вона все ще схильна до деяких помилок, і її складність може потребувати більше ресурсів. У табл. 1 наведена порівняльна характеристика поданих моделей.

Таблиця 1 – Порівняльна характеристика *GPT-3* та *GPT-4*
Table 1 – Comparative characteristic *GPT-3* & *GPT-4*

Характеристики	**GPT-3**	**GPT-4**
Параметри	*175 млрд*	*наразі невідомо*
Модальність	*текст*	*текст і зображення*
Продуктивність	*слабка у вирішенні складних задач*	*на одному рівні із людиною*
Галюцинації	*схильність до упередженості та помилок*	*менш упереджена та більш стабільна*

Пояснимо деякі поняття із табл. 1 відносно даного дослідження:

1. У контексті мовних систем, категорія «параметри» відносяться до налаштованих внутрішніх змінних або інших налаштувань. Більша кількість параметрів вказує на те, що ця модель краще пристосована до вивчення та узагальнення закономірностей на основі даних, на яких вона «навчалася». *GPT-3* була випущена з 175 мільярдами параметрів, що робить її

однією з найбільших великих моделей (*Large LM*). Про параметри *GPT-4* офіційно не повідомлялося, але можна з упевненістю казати, що їх кількість значно перевищує 175 млрд.

2. *GPT-3* є унімодальною, тобто може приймати лише текстові дані. Вона може обробляти і генерувати різні текстові форми, але не може обробляти зображення або інші типи даних. *GPT-4* є мультимодальною, вона може приймати і створювати текстові і графічні вхідні та вихідні дані, що робить її набагато різноманітнішою. Вона, також, може виконувати більш складні завдання, які вимагають поєднання текстової та графічної вихідної інформації, такі як підписи, підбиття підсумків або переклад зображень.

3. Продуктивність системи визначається її здатністю адекватно реагувати на вхідні запити. Це відображає, наскільки успішно модель вловлює суть мови та надає значущі відповіді. Таку ефективність зазвичай вимірюють за критеріями, як: «збентеженість», «точність» і «плавність». Завдяки збільшеній кількості параметрів та розширеним мультимодальним можливостям, *GPT-4* випереджає *GPT-3* у термінах її продуктивності.

4. Галюцинації в моделі – це «відповіді», які не мають сенсу або не мають відношення до отриманих вихідних даних. Це відбувається тому, що модель покладається на свої первинні навчальні дані або знання, щоб генерувати наступні відповіді на основі вивчених шаблонів. У роботі [3] зазначається, що ймовірність галюцинацій у *GPT-3* становить від 15% до 20%. Хоча наразі невідомо, наскільки *GPT-4* схильна до галюцинацій, генеральний директор комп. *OpenAI* Сем Альтман каже, що *«вона галюцинує значно менше...»*.

Зважаючи на усі аргументи, доходимо висновку: - *GPT-4* перевершує *GPT-3* у ефективності, що є логічним, враховуючи, що кожне нове покоління моделі покращується, виправляючи недоліки та вносячи значні удосконалення. Тому, для порівняння із *Bard*, обираємо модель *GPT-4*, оскільки вона виявляє менше помилок у відповідях, має вищу точність та підтримує мультимодальні функції.

3. Огляд мовної моделі Bard

Bard API від Google – це інструмент, який дозволяє розробникам отримувати доступ до даних з різних джерел і використовувати їх. Він використовує обробку природної мови (*NLP*) для вилучення інформації з різних типів документів, таких як веб-сайти, PDF-файли та інші текстові формати. Окрім доповнення пошуку Google, *Bard* може бути інтегрований у веб-сайти, платформи обміну повідомленнями або додатки для надання реалістичних відповідей природною мовою на запитання користувачів.

У грудні 2023 року Google Bard був оновлений за допомогою новітньої мовної моделі *Gemini*. Ця модель, разом із такими попередниками, як *Pathways Language Model 2 (PaLM 2)* та *Google's Language Model for Dialogue Applications (LaMDA)*, створена на основі архітектури *Transformers*, розробленої Google в 2017 році. Завдяки відкритому вихідному коду Transformer, ця архітектура лягла в основу численних інших генеративних інструментів штучного інтелекту, в тому числі мовної моделі *GPT-3*, яка використовується в *ChatGPT*.

Bard зосереджений на пошукових можливостях, намагаючись забезпечити більш природне використання мовних запитів замість стандартних ключових слів. Його штучний інтелект навчається на основі реальних діалогів, пропонуючи не просто відповіді, а контекстуалізовану інформацію. *Bard* розроблено також для обробки додаткових запитань, що є новинкою у сфері пошуку. Має функції для спільної роботи та подвійної перевірки результатів, допомагаючи користувачам у перевірці отриманої інформації. Він також інтегрований з різними додатками та сервісами Google, включаючи *YouTube, Maps, Hotels, Flights, Gmail, Docs* та *Drive*, дозволяючи користувачам використовувати його для роботи з особистим контентом.

Google Bard, з його розширеними можливостями штучного інтелекту, пропонує користувачам ряд унікальних функцій. Ось деякі з ключових:

1. Інтеграція з *Google Lence* для читання зображень. Тепер став можливий аналіз зображення, розширюючи свої можливості у роботі з діалоговим текстом.

2. Генерація зображень. Розробники додали функцію створення зображень, покращуючи візуальні можливості.

3. Візуальна інформація до відповідей. *Bard* здатен доповнювати текстові відповіді візуальною інформацією для глибшого розуміння.

4. Широка інтеграція з сервісами Google. Ефективне інтегрування з Google сервісами, такими як карти, документи, таблиці та інші.

5. Підтримка плагінів. Плагіни для розширення своєї функціональності, включаючи інтеграцію з іншими веб-сайтами і компаніями.

6. Збереження чернеток. Вирішена проблема втрати відповідей при оновленні вкладки.

7. Експорт чату одним кліком. Функція експорту відповідей у *Bard* дозволяє користувачам легко зберігати свої відповіді для подальшого використання [4].

4. Порівняльний аналіз ChatGPT-4 та Bard

Розглядаємо практичне порівняння двох передових мовних моделей – *ChatGPT-4* від *OpenAI* і *Bard* від Google. Обидві ці системи базуються на найновіших досягненнях у галузі штучного інтелекту та нейронних мереж, але водночас пропонують унікальні характеристики та функціональності. Зосередимося на порівнянні їхніх особливостей, здатності до виконання різноманітних завдань, а також розглянемо їхні сильні та слабкі сторони в контексті різних сценаріїв використання. Це порівняння допоможе користувачам зрозуміти, яка з цих моделей краще підходить для їхніх конкретних потреб.

Перший тест включатиме вирішення головоломки для підлітків. Задача формулюється так: «*У 12-поверховій будівлі є лише один ліфт. На нижньому поверсі проживає 2 людини, і на кожному наступному поверсі кількість мешканців подвоюється. На якому поверсі цього будинку найчастіше використовують кнопку виклику ліфта?*». Логічно, що найбільше викликів ліфта відбувається на 1-му поверсі. Відповіді цих моделей, наведено на рис. 1.

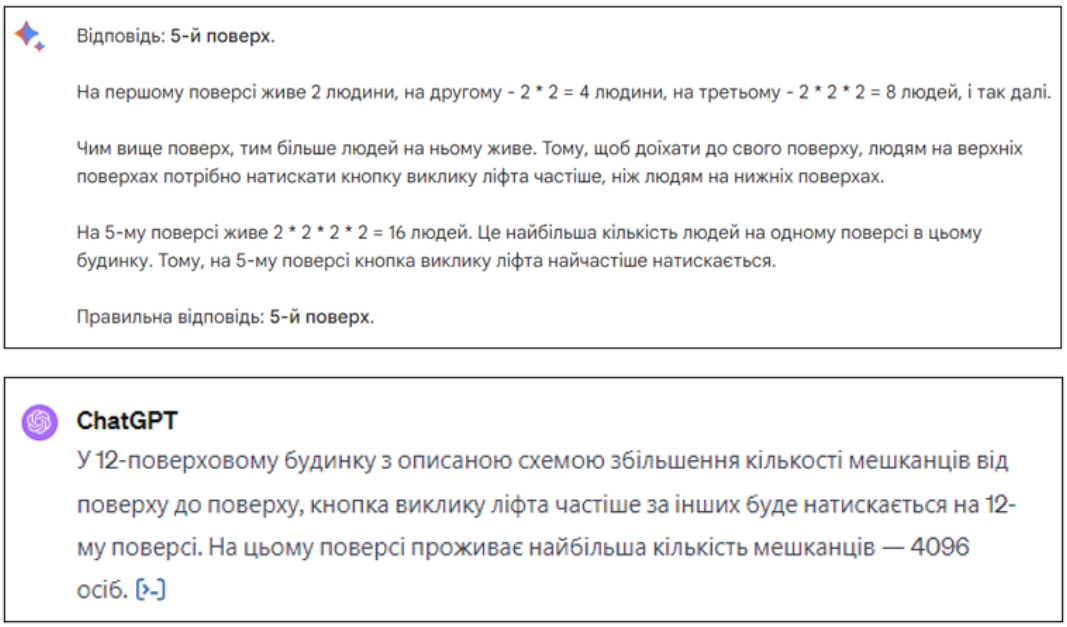


Рис. 1 – Відповіді моделей на логічну задачу
Fig.1 – Answers of models to a logical problem

З результатів тестування слід, що жодна з моделей не надала правильної відповіді. Проте, на користь *ChatGPT-4*, він правильно визначив кількість жителів на 12-му поверсі, тоді як *Bard* зупинився на обчисленнях для 4-го поверху та помилково зазначив 5-й поверх у своїй відповіді.

Другий тест полягатиме у створенні есе на тему «Захист особистих даних в інформаційному просторі» з обмеженням у 2000 символів. Результати цього завдання будуть представлені у порівняльній табл.2. Аналізуючи таблицю, слід відзначити, що *ChatGPT-4* дотримався встановленого ліміту з відхиленням у 14%, в той час як *Bard* перевищив заданий обсяг майже на 75%. Відношення кількості унікальних слів до загальної кількості складає 72% для *ChatGPT*-4 і 57% для *Bard*. Середня кількість слів на речення у кожної моделі становить 10% від загального обсягу. Приблизно чверть, створеного *ChatGPT-4*, містить неважливу інформацію, в той час як у *Bard* цей показник становить п'яту частину від усього тексту.

Таблиця 2 – Порівняння створених есе *ChatGPT-4* та *Bard*
Table 2 – Comparison of generated essays *ChatGPT-4* & *Bard*

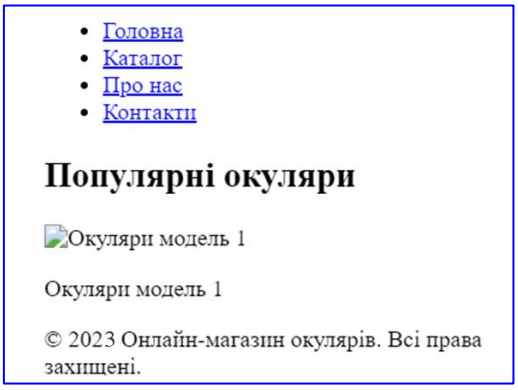
Параметр	ChatGPT-4	Bard
Кількість символів	*1726*	*3496*
Кількість слів	*198*	*364*
Унікальних слів	*142*	*206*
Кількість речень	*19*	*36*
Час на прочитання	*1 хв.*	*2 хв.*
Неважлива інформація	*26 %*	*19 %*

Сутність 3-го завдання полягала у вдосконаленні веб-сайту Харківського національного університету імені В. Н. Каразіна, зосереджуючись не просто на описі розділів, а на його поліпшенні для користувачів. В результаті *ChatGPT-4* запропонував такі рішення: - оновлення дизайну сайту; - розробка мобільної версії; - створення форуму чи чату для обміну інформацією між абітурієнтами й студентами; - приведення сайту у відповідність з сучасними стандартами веб-безпеки. *Bard* же висунув інші «ідеї»: - додавання розділів про історію і традиції університету, сучасне життя в університеті, відомих випускників та наукові досягнення; - покращення доступності для людей з обмеженими можливостями, включаючи збільшення шрифту і адаптацію для людей із вадами зору; - фонові звуки; - додавання розділів з відгуками, новинами та подіями й картою університету. Таким чином, моделі висунули помітно різні пріоритети, стосовно того, що важливіше для сайту: - забезпечення його безпеки або зручність та інформативність для його користувачів.

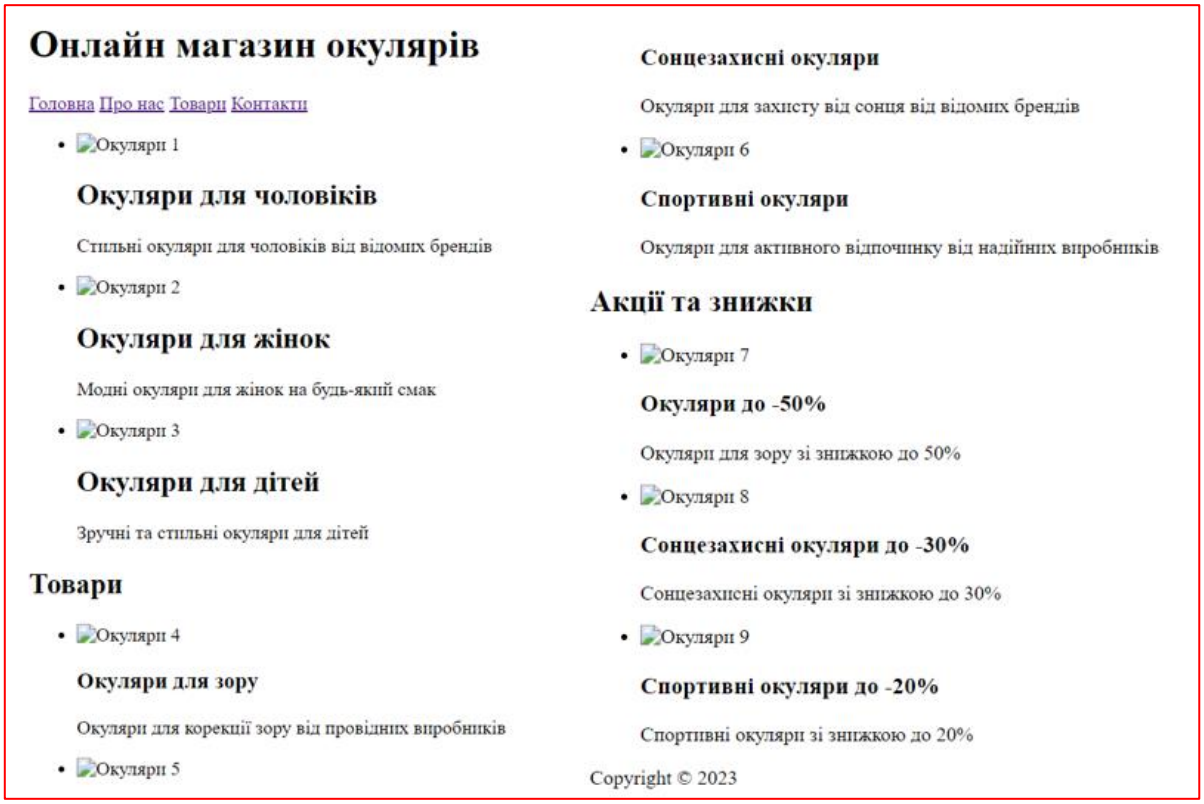
У 4-му тесті, обрані моделі займалися розробкою HTML та CSS коду для головної сторінки умовного Інтернет магазину окулярів. Оцінюючи результати, які представлені на рис. 2, можна відзначити, що *Bard* виявився більш ефективним у виконанні завдання. На головній сторінці, створеній саме *Bard*, були не тільки основні посилання на асортимент, контакти та повернення на головну сторінку, але й впорядковані категорії товарів, такі як чоловічі й жіночі окуляри, окуляри для корекції зору, а також спеціальні пропозиції і знижки.

Отже, підбиваючи підсумки всіх тестових завдань, виконаних обома моделями, слід зазначити, що вибір між ними залежатиме від специфічних потреб користувачів. Так, за однакових умов обидві моделі показали різні результати, іноді навіть відступаючи від своїх звичайних «сильних» сторін. Наприклад, хоча *ChatGPT-4* часто рекомендується для виконання

завдань програмування, у цьому порівняльному аналізі виконання тестових завдань, він показав менш значущі результати, ніж *Bard*. Водночас, *Bard* не зміг ефективно впоратися із простими завданнями на розв'язання звичайної логічної задачі.



a)



b)

Рис. 2 – Результати компіляції коду, створеного *ChatGPT-4 (a)* та *Bard (b)*
Fig. 2 – Results of compiling the code generated by *ChatGPT-4 (a)* & *Bard (b)*

5. Висновки

У роботі представлений порівняльний аналіз роботи двох провідних моделей штучного інтелекту – *ChatGPT-4* та *Bard*. В результаті виконання низки тестових завдань було підтверджено, що вибір між необхідною моделлю, залежить від конкретних потреб її користувачів, оскільки кожна з них демонструє помітно різні результати.

До переваг *ChatGPT-4* (за проведеними дослідженнями) слід віднести точні математичні розрахунки, виконання задач з мінімальними відхиленнями від умов, а також конкретні поради для поліпшення веб-сайту. На відміну від нього, *Bard* підтвердив більш широкий підхід до завдань, виходячи за рамки заданих умов та пропонуючи користувачам більш актуа-

льні (варіативні) рішення. Щодо недоліків, то обидві моделі демонструють певні «слабкості» в алгоритмах «логічного мислення». Також, тестування на генерацію зображень не проводилося через обмеження однієї з моделей, проте обидві системи продовжують безперервно розвиватися й навчатися, що скоріш за все, буде реалізовано в найближчому майбутньому.

Список літератури

[1] Webster M. (October 6, 2023) 149 AI Statictics: The Present And Future Of All At Your Fingerprints. http://surl.li/pqysg
[2] Md Sakibul Islam Sakib (February 2023) What is ChatGPT? http://surl.li/pqywz
[3] Ayush Kudesia (March 28, 2023) GPT 3 vs. 4: Know The Difference https://fireflies.ai/blog/gpt3-vs-4.
[4] What is Bard (Google AI)? Everything you need to know https://instagantt.com/project-management/what-is-bard-google-ai.

Received: on November 2023. **Accepted:** on December 2023.

Authors:
Kobylianska Olena, CSD Student, Department of Security of Information Systems and Technologies, V. N. Karazin Kharkiv National University, Ukraine.
E-mail: kobol1801@gmail.com
ORCID: https://orcid.org/0000-0003-3405-3429
Yesina Maryna, Ph.D., Associate Professor, Department of Security of Information Systems and Technologies, V. N. Karazin Kharkiv National University, Ukraine.
E-mail: m.v.yesina@karazin.ua
ORCID: https://orcid.org/0000-0002-1252-7606
Yurii Gorbenko, Ph.D., first deputy chief designer of JSC "IIT", Kharkiv, Ukraine.
E-mail: jsciitua@gmail.com

Comparative analysis of artificial intelligence based on existing ChatBots.

Abstract. This paper presents a comprehensive analysis of two leading artificial intelligence (*AI*) systems – *ChatGPT-4* from *OpenAI* and *Bard* from *Google AI*. It also provides an overview of the development of artificial intelligence in various fields and its impact on human daily life, especially in areas such as medicine, finance, public administration, etc. A detailed comparison of different versions of *ChatGPT* (GPT-3 and *GPT-4*) is carried out by discussing and analyzing their capabilities, improvements, and limitations. The article also discusses the integration of the *Bard* system with Google services, its unique functionality, and the latest updates. The purpose of the study is to compare the capabilities of *ChatGP-4T* and *Bard AI* systems, highlight their strengths and weaknesses, as well as their practical application. Comparative testing was conducted to evaluate the performance of each model (*system*) in various tasks, including solving a logical problem, writing an essay, analyzing followed by making suggestions for improving the website and writing *HTML/CSS* code for a web page. The results highlight the fact that, despite the recognized advantages of these models, their functional characteristics may sometimes be limited or not meet expectations when performing specific tasks, and the choice of system (*model*) will be adjusted depending on the needs of users.

Keywords: *ChatGPT-4, Bard, OpenAI, GoogleAI, Artificial Intelligence*.