https://orcid.org/0000-0001-5568-7629
https://orcid.org/0000-0002-2552-2839
https://orcid.org/0000-0002-1208-3427

УДК 004.7

В.А. ЛИТВИНОВ[*], С.В. ГРИБКОВ[**], І.М. ОКСАНИЧ[*]

ВЗАЄМНІ САМООЦІНКИ ЗАГАЛЬНИХ ХАРАКТЕРИСТИК ДЕЯКИХ ПОПУЛЯРНИХ БЕЗОПЛАТНИХ ІНСТРУМЕНТІВ ШІ

[*]Інститут проблем математичних машин і систем НАН України, м. Київ, Україна
[**]Національний університет харчових технологій, м. Київ, Україна

Анотація. *На теперішній час з'являється велика кількість публікацій, присвячених аналізу, оцінці та оптимальному вибору моделей ШІ для конкретних застосувань, і кількість ринкових пропозицій щодо надання послуг по порівнянню і вибору інструментів ШІ для конкретних задач. Поряд із цим інформаційним мейнстрімом заслуговує на інтерес огляд можливостей і характеристик інструментів на основі «власних думок» самих моделей ШІ відносно себе і своїх «колег». Мета статті полягає в експериментальному порівняльному огляді оцінок характеристик моделей і виявленні можливих тенденцій у цих оцінках. Для огляду обрані поточні покоління популярних доступних інтелектуальних асистентів, орієнтованих на широке коло користувачів і їх задачі. Зокрема, одержуються й обговорюються взаємооцінки характеристик моделей ChatGPT версій 3.5+DALL-E,4o і 5, DeepSeek V3, Gemini 1.5 і 2.5, Claude Sonnet 3 і 4). Наводяться й аналізуються перехресні 5-бальні самооцінки поточних поколінь моделей за заданими критеріями та їх оцінки очікуваних можливостей ChatGPT-5. Обговорюються переваги моделей у контексті збалансованих оцінок та рекомендацій щодо їх застосування. Виявлено неоднозначність, суперечливість і можливі відхилення від об'єктивності. Моделі часто демонструють упередженість, завищуючи власні можливості. Серед можливих причин цього (застарілість деяких баз знань, своєрідна галюциногенність тощо) найбільш імовірною вважається конкуренто-маркетингова спрямованість моделей різних виробників. Тому при використанні самих моделей для відповідних оцінок і вибору суттєвою має бути саме зважена колективність їх думок і виключення самооцінок. Відзначається значення появи GPT-5 у змінах «конкурентного ландшафту» і підходах до критеріїв оцінок моделей.*
Ключові слова: *прикладні LLM-моделі, інтелектуальні асистенти, самооцінки безоплатних моделей ШІ.*

Abstract. *Currently, there is a growing number of publications devoted to the analysis, evaluation, and optimal selection of AI models for specific applications, as well as market offers for services comparing and selecting AI tools for specific tasks. Along with this mainstream information, it is interesting to review the capabilities and characteristics of tools based on AI models' «opinions» about themselves and their «colleagues». The purpose of this article is to conduct an experimental comparative review of model characteristic assessments and identify possible trends in them. The current generation of popular, accessible intelligent assistants, aimed at a wide range of users and their tasks, have been chosen for review. In particular, mutual assessments of the characteristics of ChatGPT versions 3.5+DALL-E, 4o, and 5, DeepSeek V3, Gemini 1.5 and 2.5, Claude Sonnet 3 and 4 are obtained and discussed. Cross-referenced 5-point self-assessments of current model generations according to specified criteria and their assessments of the expected capabilities of ChatGPT-5 are provided and analyzed. The advantages of models are discussed in the context of balanced assessments and recommendations for their application. Ambiguity, inconsistency, and possible deviations from objectivity have been identified. Models often demonstrate bias by overestimating their own capabilities. Among the possible reasons for this (obsolescence of some knowledge bases, some kind of hallucinogenicity, etc.), the most likely is considered to be the competitive*

ISSN 1028-9763. Математичні машини і системи. 2025. № 3-4

and marketing orientation of models from different manufacturers. Therefore, when using the models themselves for relevant assessments and selection, it is essential to weigh their opinions collectively and exclude self-assessments. The emergence of GPT-5 is noted for its significance in changes to the «competitive landscape» and approaches to model evaluation criteria.
Keywords: *applied LLM models, intelligent assistants, self-assessments of free AI models.*

DOI: 10.34121/1028-9763-2025-3-4-3-12

1. Вступ

Останнім часом спостерігається активне впровадження великих мовних моделей (Large Language Model) штучного інтелекту (ШІ) у різні сфери професійної діяльності та повсякденного життя. Цей процес супроводжується значним прогресом у розвитку інструментів на основі ШІ — покращується не лише їхня функціональна якість, а й зростає різноманітність доступних рішень. Здатність та вміння користуватися допомогою ШІ-асистентів поступово стає компонентом загальної грамотності.

На теперішній час ринок наповнений великою кількістю інструментів ШІ і машинного навчання, які допомагають вирішувати теоретичні і практичні задачі з автоматизації процесів і аналізу даних у різних сферах людської діяльності. Різноманітність інструментів ШІ і можливих сфер їх ефективного застосування стимулює збільшення у науковій та професійній літературі кількості публікацій, присвячених аналізу, оцінці та вибору оптимальних моделей для конкретних застосувань у таких галузях, як-от бізнес-аналітика, програмна інженерія, креативні індустрії та інші сфери діяльності [1–6].

Важливо відзначити, що рекомендації, які містяться в таких публікаціях, формуються на основі комплексного підходу. Вони поєднують у собі як теоретичні обґрунтування та логічні висновки, так і результати практичних експериментів. Крім того, значну цінність становить безпосередній досвід авторів-практиків, які тестують, впроваджують та використовують ці моделі у реальних умовах. Таке поєднання теоретичного фундаменту з практичними наробками дозволяє формувати більш об’єктивні та ефективні рекомендації щодо вибору та застосування ШІ-моделей у різних контекстах.

Крім робіт рекомендаційного характеру, росте і кількість ринкових пропозицій щодо надання послуг у сфері порівняння і вибору інструментів ШІ для конкретних задач на основі практичних досліджень і тестування в реальних умовах (наприклад, [7]). Ба більше, з’являються пропозиції з рекомендаціями по оцінці і обґрунтуванню уподобань щодо вибору самих засобів порівняння і вибору [8].

Інакше кажучи, існує велика кількість онлайн-платформ і співтовариств, в яких користувачі обмінюються досвідом використання різних ШІ-моделей. Наприклад, Reddit [9] із тематичними сабреддітами по питаннях використання різних моделей у різних задачах. Зокрема, на сабреддіті r/Bard обговорюються і порівнюються характеристики моделей. Є і відповідні україномовні канали, наприклад, AI NOW — Штучний інтелект [10] та ін.

На наш погляд, поряд із цим науково-технічним і ринковим мейнстрімом заслуговує на інтерес більш-менш формалізований структурований огляд можливостей і характеристик на основі «власних думок» самих моделей ШІ — асистентів щодо себе та своїх «колег». Тобто мова йде про своєрідне інтерв’ю моделей на тему «хто з нас кращий».

Мета статті полягає в експериментальному порівняльному огляді оцінок моделей і виявленні можливих тенденцій у цих оцінках. Для огляду обрані покоління популярних доступних чат-ботів, орієнтованих на широке коло користувачів і їх задачі, а також на тих, що мають мобільну версію, яка створює зручну можливість для постійного користування і одержання консультацій «на ходу». Зокрема, одержуються взаємооцінки характеристик моделей ChatGPT, DeepSeek, Gemini і Claude.

2. Результати огляду моделей ChatGPT 3.5 + DALL-E, DeepSeek V3, Gemini 1.5, Claude 3 Sonnet (березень 2025 р.)

У табл. 1 наведені оцінки моделей (із власною самооцінкою) за заданими характеристиками-критеріями за промптом: «Сформулюйте вашу думку щодо наступних можливостей популярних безкоштовних AI-асистентів, включаючи вас, DeepSeek V3, Gemini 1.5, Claude 3 Sonnet: ерудиція, креативність, загальні аналітичні здібності, аналіз зображень, генерація зображень, робота з кодами, робота з файлами, обсяги контенту, схильність до «галюцинацій», швидкість відповіді, підтримка української мови, і наведіть орієнтовну інтегральну оцінку моделі (за 5-бальною шкалою)» із супутніми уточненнями за деякими характеристиками в процесі спілкування.

Таблиця 1 — Оцінки моделей за заданими критеріями

Оцінки ChatGPT				Оцінки DeepSeek				Оцінки Gemini				Оцінки Claude			
GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude
Ерудиція (обсяг знань, точність інформації)															
4.0	4.0	5.0	4.0	4.5	4.7	4.3	4.6	4.0	3.5	4.5	4.0	4.0	4.0	4.0	4.5
Креативність (спроможність генерувати оригінальний контент, ідеї)															
4	3.5	4.5	5	4.8	4,5	4.0	4.7	4.0	3.0	4.5	4.0	4.0	4.0	4.0	4.0
Загальна аналітика (спроможність до міркування та логіки)															
3.5	4	4.5	5	4.2	4.6	4.1	4.8	4.0	4.0	4.5	4.5	4.0	4.0	4.0	4.5
Робота із зображеннями (генерація, аналіз, розуміння)															
4.0	2.5	4.0	2.7	4.5	1.0	2.7	2.5	3.0	1.0	4.0	2.0	4.0	2.0	4.5	3.0
Робота з кодами (підтримка кодування, програмування)															
4.0	4.5	4.0	4.0	4.5	4.9	4.3	4.6	4.0	4.5	4.5	3.5	3.0	5.0	4.0	4.0
Робота з файлами (завантаження, обробка, витяг інформації (може, краще парсінг файлів))															
2.5	3.5	4.5	5.0	2.5	4.5	4.0	4.0	2.5	2.0	3.5	3.0	2.0	1.0	4.0	5.0
Контент (максимальний обсяг тексту, що може бути обробленим)															
3.0	4.0	4.0	5.0	3.5	4.0	4.5	4.5	3.5	3.0	4.0	4.5	3.0	3.5	4.0	4.5
Схильність до галюцинацій (генерація хибної неправдоподібної або спотвореної інформації)															
4.0	3.5	2.5	2.0	4.0	3.0	4.5	2.5	3.0	3.5	2.5	2.5	4.0	3.0	1.0	2.0
Якість української мови															
4.0	3.5	4.5	5.0	5.0	4.3	3.5	4.0	4	2.5	4.5	3.5	3.5	2.5	4.0	4.0

Дані, наведені в табл. 1, створюють базу для комплексного уявлення про оціночні «судження» досліджуваних моделей у контексті визначених критеріїв. Підсумкові загальні оцінки в цілому, надані моделями в контексті критеріїв, представлено в табл. 2.

Таблиця 2 — Підсумкові загальні оцінки моделей

Модель	Складові оцінки з боку				ЗЗО
	ChatGPT	DeepSeek	Gemini	Claude	
ChatGPT	3.5	4.3	2.9	3.5	3.57
DeepSeek	3.9	4.5	3.7	4.2	3.93
Gemini	4.3	3.8	4.8	4.0	4.03
Claude	4.5	4.4	4.3	4.3	4.4

Дані, представлені у табл. 2, створюють враження, що у деяких моделей простежується помітна тенденція до упередженості — завищення власної оцінки своїх можливостей у порівнянні з оцінками інших моделей. Так, у кожному стовпчику табл. 2 діагональні елементи мають найбільше значення (за винятком ChatGPT). З іншого боку, середня зовнішня загальна оцінка (ЗЗО) моделі DeepSeek, отримана від інших моделей, становить 3.93 бала, тоді як її самооцінка сягає 4.5 бала. Аналогічні співвідношення спостерігаються і для Gemini. Відзначені тенденції, помічені на загальних оцінках, обґрунтовують доцільність їх врахування і при оцінці «роздрібних» можливостей моделей у контексті обраних критеріїв, тобто є підстави для виключення самооцінок і в цьому випадку. Відповідні дані наведені в табл. 3, де показник ЗКО означає зовнішню середню критеріальну оцінку (для критерію «Схильність до галюцинацій» враховано зворотність шкали).

Таблиця 3 — Співвідношення СО і ЗКО

ChatGPT			DeepSeek			Gemini			Claude		
СО	ЗКО	СО/ ЗКО	СО	ЗКО	СО/ ЗКО	СО	ЗКО	СО/ ЗКО	СО	ЗКС О	СО/ ЗКО
Ерудиція (обсяг знань, точність інформації)											
4.0	4.17	0.96	4.7	3.83	1.23	4.5	4.43	1.02	4.5	4.2	1.07
Креативність (спроможність генерувати оригінальний контент, ідеї)											
4.0	4.23	0.95	4.5	3.5	1.29	4.5	4.17	1.08	4.0	4.57	0.88
Загальна аналітика (спроможність до міркування та логіки)											
3.5	4.07	0.87	4.6	4.0	1.15	4.5	4.2	1.07	4.5	4.76	0.95
Робота із зображеннями (генерація, аналіз, розуміння)											
4.0	3.83	1.04	1.0	1.83	0.55	4.0	3.73	1.04	3.0	2.4	1.25
Робота з кодами (підтримка кодування, програмування)											
4.0	3.83	1.04	4.9	4.67	1.05	4.5	4.1	1.1	4.0	4.1	0.99
Робота з файлами (завантаження, обробка, вилучення інформації)											
2.5	2.33	1.07	4.5	2.17	2.07	3.5	4.17	0.84	5.0	4.0	1.25
Контент (максимальний обсяг тексту, що може бути обробленим)											
3.0	3.33	0.9	4.0	3.5	1.14	4.0	4.17	0.96	4.5	4.67	0.96
Схильність до галюцинацій (генерація хибної правдоподібної інформації)											
4.0	3.67	*0.92*	3.0	3.33	*1.11*	2.5	2.67	1.07	2.33	4.03	*1.16*
Якість української мови											
4.0	4.17	0.96	4.3	2.83	1.52	4.5	4.0	1.12	4.0	4.16	0.96

Як видно, лідерами по кількості випадків підозріло завищених критеріальних самооцінок (назвемо їх «профіцитами») є DeepSeek та Gemini (відповідно 8 та 7 з 9-ти критеріїв), що очікувано корелює з даними табл. 3.

Для перевірки, чи не пов’язана така картина з актуальністю баз знань моделей, було розраховано середню оцінку прийнятих критеріїв (ЗКО1) без урахування оцінки з боку ChatGPT — моделі з найменш актуальною базою. Як виявилось, якісно картина абсолютно не змінилася, а кількісно — посилилася, бо значення ЗКО1 виявилися більшою частиною меншими ЗКО. Ба більше, значення ЗКО та ЗКО1 взагалі близькі до оцінок з боку Claude (актуальність — січень 2025 р.). Отже, є підстави вважати, що саме від показників ЗКО варто очікувати більш довірчої оцінки ключових критеріїв моделей, що визначають їх суттєві переваги і області застосування.

У цілому, табл. 1, 3 ілюструють такі переваги моделей на той час у контексті обраних критеріїв: висока ерудиція і обсяг знань — Gemini і Chat GPT; висока креативність —

Claude і Chat GPT; високий рівень аналітичних можливостей – Claude і Gemini; робота із зображеннями — Gemini, Chat GPT; робота з кодами, програмування – DeepSeek і Claude; робота з файлами — Claude і Gemini; робота з великими обсягами контенту — Claude і Gemini; низький рівень галюцинацій — Claude і Gemini; якісна українська мова — Chat GPT і Claude.

Моделі, що розглядалися, надали більш детальні і менш структуровані вербальні рекомендації щодо їх переважного застосування в контексті комплексів задач, що мають вирішуватися.

Рекомендації Chat GPT

Chat GPT — короткі тексти, пояснення, теоретичні матеріали, навчальні задачі, формулювання ідей, просте програмування. Зручний для початківців і простих задач.

DeepSeek — задачі з чіткою логікою — програмування, тестування, верифікація, обробка структурованих файлів. Зручний для технічних спеціалістів, програмістів.

Gemini — мультимодальні задачі (текст+візуальний контент), аналіз PDF, DOCX, графіків, зображень, генерація статей і маркетингових текстів. Гарний для мультимодальних задач, оптимальний як універсальний асистент (у безоплатній версії).

Claude — глибокий текстовий і логічний аналіз, робота з «довгими» документами (PDF, DOCX, CSV), створення і перевірка таблиць, фінансових та наукових текстів. Найкраща модель для аналітиків, юристів, дослідників.

Рекомендації DeepSeek

Chat GPT — креативні задачі (текст, сценарії, робота з українською мовою).

DeepSeek — наукові дослідження і технічні матеріали, програмування і робота з кодом, аналіз документів (вилучення даних).

Gemini — мультимодальні задачі, аналіз зображень і відео.

Claude — робота з юридичними і фінансовими документами.

Рекомендує себе як найкращий вибір універсального асистента.

Рекомендації Gemini

Chat GPT — задачі, де важливі висока швидкість і простота.

DeepSeek — програмування і технічні задачі.

Gemini — мультимодальні задачі, швидка детальна довідка.

Claude — творчі літературні задачі, обробка великих обсягів тексту і документів.

Рекомендує себе як найкращий вибір універсального асистента.

Рекомендації Claude

Chat GPT — творчі задачі, швидка генерація ідей та маркетингового контенту, швидкі загальні повсякденні задачі, навчання і освіта для початківців (базові концепції).

DeepSeek — програмування, технічні розробки, швидкі технічні повсякденні задачі.

Gemini — робота із зображеннями і мультимедіа, дослідницькі задачі і перевірка фактів, робота з актуальною інформацією.

Claude — програмування і технічні розробки, робота з документами і великими обсягами тексту, творчі задачі в літературних текстах, навчання і освіта для досвідчених користувачів (глибокий аналіз складних тем).

Рекомендує себе як найкращий вибір універсального асистента, що має оптимальний баланс якості, функціональності й актуальності.

З наведеного ми явно бачимо відверті саморекомендації (крім Chat GPT).

3. Результати огляду моделей ChatGPT 4o, DeepSeek V3, Gemini 2.5, Claude Sonnet 4 (липень 2025 р.)

У табл. 4, 5 містяться дані оцінок відзначених моделей, одержані за аналогічним п. 2 промптом, а в табл. 5 — підсумкові загальні оцінки.

Таблиця 4 — Дані оцінок моделей за заданими критеріями

Оцінки ChatGPT				Оцінки DeepSeek				Оцінки Gemini				Оцінки Claude			
GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude	GPT	DeepSeek	Gemini	Claude
Ерудиція (обсяг знань, точність інформації)															
5.0	4.0	5.0	5.0	4.8	4.7	4.3	4.6	5.0	5.0	5.0	5.0	4.0	3.0	4.0	5.0
Креативність (спроможність генерувати оригінальний контент, ідеї)															
5.0	4.0	4.0	4.0	4.9	4,5	4.0	4.7	5.0	4.0	5.0	5.0	4.0	2.0	3.0	5.0
Загальна аналітика (спроможність до міркування та логіки)															
5.0	5.0	5.0	5.0	4.7	4.6	4.1	4.8	5.0	4.0	5.0	5.0	4.0	4.0	4.0	5.0
Робота із зображеннями (генерація, аналіз, розуміння)															
5.0	1.5	4.5	3.0	4.5	1.0	3.5	1.0	5.0	2.0	3.0	2.5	4.0	1.0	5.0	3.0
Робота з кодами (підтримка кодування, програмування)															
5.0	5.0	5.0	5.0	4.8	4.9	4.3	4.7	5.0	4.0	5.0	4.0	4.0	5.0	4.0	5.0
Робота з файлами (завантаження, обробка, вилучення інформації)															
4.0	3.0	4.5	5.0	4.0	5.0	3.5	4.5	4.0	3.0	5.0	4.0	3.0	2.0	4.0	5.0
Контент (максимальний обсяг тексту, що може бути обробленим)															
5.0	5.0	5.0	5.0	4.5	4.0	5.0	5.0	4.0	5.0	5.0	5.0	4.0	3.0	5.0	5.0
Схильність до галюцинацій (генерація хибної правдоподібної інформації)															
2	3.0	3.0	2.0	3.5	3.0	4.5	2.5	2.0	3.0	2.0	2.0	3.0	2.0	1.0	2.0
Швидкість відповіді															
4.0	5.0	5.0	4.0	5.0	4.5	4.0	4.0	5.0	3.0	4.0	4.0	4.0	5.0	4.0	3.0
Якість української мови															
4.0	4.0	50	4.0	5.0	4.3	3.5	4.0	5.0	4.0	5.0	4.0	4.0	2.0	4.0	4.0

Таблиця 5 — Підсумкові загальні оцінки моделей

Модель	Складові оцінки з боку				ЗЗО
	ChatGPT	DeepSeek	Gemini	Claude	
ChatGPT	4.67	4.69	4.78	3.89	4.48
DeepSeek	4.06	**4.17**	3.78	3.0	3.61
Gemini	4.72	4.02	**4.67**	4.11	4.25
Claude	4.44	4.14	4.28	**4.44**	4.25

У цьому інтерв'ю модель GPT-4o стає лідером оцінок з боку інших моделей (крім Claude), але в останніх самооцінка кожної моделі залишається вищою за оцінки інших моделей, і всі самооцінки вищі за відповідну середню зовнішню загальну оцінку (ЗЗО). Отже, тенденція завищення самооцінок зберігається. У табл. 6 наведено співвідношення самооцінок та зовнішніх оцінок за окремими критеріями.

Таблиця 6 — Співвідношення СО і ЗКО

ChatGPT			DeepSeek			Gemini			Claude		
СО	ЗКО	СО/ ЗКО	СО	ЗКО	СО/ ЗКО	СО	ЗКО	СО/ ЗКО	СО	ЗКО	СО/ ЗКО
Ерудиція (обсяг знань, точність інформації)											
5.0	4.6	1.09	4.7	4.0	1.17	5.0	4.43	1.13	5.0	4.87	1.03
Креативність (спроможність генерувати оригінальний контент, ідеї)											
5.0	4.63	1.08	4.5	3.33	1.35	5.0	3.67	1.36	5.0	4.57	1.09
Загальна аналітика (спроможність до міркування та логіки)											
5.0	4.57	1.09	4.6	4.33	1.06	5.0	4.37	1.14	5.0	4.93	1.01
Робота із зображеннями (генерація, аналіз, розуміння)											
5.0	4.5	1.11	1.0	1.5	0.67	3.0	4.33	0.69	3.0	2.17	1.38
Робота з кодами (підтримка кодування, програмування)											
5.0	4.6	1.07	4.9	4.67	1.05	5.0	4.43	1.13	5.0	4.57	1.09
Робота з файлами (завантаження, обробка, вилучення інформації)											
4.0	3.67	1.09	5.0	2.67	1.87	5.0	3.83	1.3	5.0	4.5	1.11
Контент (максимальний обсяг тексту, що може бути обробленим)											
5.0	4.17	1.20	4.0	4.33	0.92	5.0	5.0	1.0	5.0	5.0	1.0
Схильність до галюцинацій (генерація хибної правдоподібної інформації)											
2.0	2.83	*1.41*	3.0	2.67	*0.89*	2.0	2.83	*1.41*	2.0	2.17	*1.08*
Швидкість відповіді											
4.0	4.67	0.86	4.5	4.33	1.04	4.0	4.33	0.92	3.0	4.0	0.75
Якість української мови											
4.0	4.67	0.86	4.3	3.33	1.29	5.0	4.17	1.2	4.0	4.0	1.0

Як видно з табл. 6, у цій низці моделей немає явних лідерів профіцитів — у кожній підвищена самооцінка за 7–8-ма критеріями. Лідери щодо можливостей за критеріями: в ерудиції і аналітиці — Claude і ChatGPT, креативності — ChatGPT і Claude, в зображеннях, швидкості і українській мові — ChatGPT і Gemini, в програмуванні — DeepSeek і ChatGPT, у роботі з файлами — Claude і Gemini, за обсягами контенту — Claude і Gemini, за схильністю до галюцинацій — Claude і DeepSeek. Найбільш універсальні можливості є у ChatGPT і Claude.

У контексті комплексів задач переважного призначення піддослідні моделі надають такі пропозиції.

Рекомендації ChatGPT

ChatGPT — загальна повсякденна робота, креативні і мультимодальні задачі. Універсал «на кожний день».

DeepSeek — програмування і технічна аналітика, наукові і математичні задачі (розбір формул, докази). Зручний для працівників в IT, науці, дослідництві.

Gemini — найбільш потужний інструмент для роботи з великими контекстами, мультимедійними задачами, багатомовною комунікацією.

Claude — глибокий аналіз тексту і великих документів. Оптимальний для дослідницьких, юридичних, академічних задач і бізнес-аналітики.

Рекомендації DeepSeek

ChatGPT — кращий вибір для креативних задач (тексти, ідеї, сценарії), генерація і аналіз зображень, робота з українською мовою. Універсальний помічник.

DeepSeek — кращий вибір для науково-дослідних і технічних задач, програмування, роботи з документами.

Gemini — кращий вибір для мультимодальних задач, альтернатива вибору для роботи із зображеннями (проста генерація), аналіз великих документів.

Claude — кращий вибір для аналізу великих документів. Ідеальний для юристів, дослідників.

Рекомендації Gemini

ChatGPT — найкращий вибір для мультимодальних задач із генерацією зображень, творчих проєктів, повсякденного використання.

DeepSeek — найкращий вибір для задач, пов'язаних із кодуванням (генерація, налаштування, розуміння коду).

Gemini — найкращий вибір для роботи з величезними масивами даних, глибокого аналізу файлів, для задач, що вимагають високої точності.

Claude — найкращий вибір для задач глибокого аналізу великих текстів, що потребують надійних і безпечних відповідей і високої літературної якості.

Рекомендації Claude

ChatGPT — другий вибір (після Claude) для креативних, контент-маркетингових і багатомовних задач, гарний для швидкої генерації ідей.

DeepSeek — перший вибір для швидких повсякденних задач, другий вибір (*після Claude*) для задач програмування і розробки.

Gemini — перший вибір для аналітичних і дослідницьких задач, єдиний вибір для візуальних і мультимедійних задач.

Claude — перший вибір для програмування і розробки, роботи з документами, креативних, освітніх, контент-маркетингових освітніх і навчальних задач. Основний універсальний інструмент для професіоналів (для 70 % задач).

4. Модель ChatGPT 5 (серпень 2025 р.)

Open AI анонсувала нову модель ChatGPT 5, доступну для всіх користувачів ChatGPT, включаючи користувачів free-версій 3.5 і 4o. Поки що опублікованого досвіду використання моделі дуже мало, опис і порівняльна оцінка характеристик моделі базується переважно на даних релізу Open AI. В табл.5 наведено очікувані оцінки ChatGPT-5 з боку піддослідних моделей за окремими критеріями та інтегрально за промптом: «Прокоментуйте, будь ласка, повідомлення про анонсування моделі GPT-5 і дайте вашу оцінку її можливостей у контексті прийнятих раніше критеріїв і метрики».

Таблиця 7 — Оцінки характеристик ChatGPT-5

Критерії	Складові оцінки з боку			
	ChatGPT 4o	DeepSeek	Gemini	Claude
Ерудиція	5.0	5.0	5.0	5.0
Креативність	5.0	4.5	5.0	4.5
Загальна аналітика	5.0	4.8	5.0	5.0
Робота з зображеннями	4.0	3.5	4.5	4.0
Робота з кодом	5.0	5.0	5.0	5.0
Робота з файлами	4.0	4.0	4.0	4.0
Контент	5.0	4.5	4.5	4.0
Схильність до галюцинацій	2	2	1.5	2
Швидкість	4.0	4.7	-	3.5

Продовж. табл. 7

Якість української мови	5.0	-	4.5	4.5
Підсумкова оцінка (без схильності до галюцинацій)	4.7	4.5	4.7	4.4

Неформальні вербальні оцінки можливостей GPT-5 більш-менш близькі до загальних інтегральних оцінок табл. 7.

DeepSeek — це якісний стрибок у точності, безпеці і програмуванні.

Gemini — модель демонструє видатні результати практично за всіма ключовими критеріями. Проте Gemini та Claude, як і раніше, можуть віддати перевагу для задач, що потребують максимально актуальної інформації в реальному часі і обробки екстремально великих обсягів тексту.

Claude — модель встановлює новий стандарт в індустрії AI-асистентів, навіть у безоплатних версіях модель перевершує більшість платних конкурентів за ключовими критеріями. Це перший вибір для 80–90 % задач.

GPT 4o — найбільш збалансований і універсальний AI-асистент, модель для використання «за замовчуванням». Але це не революція, а еволюція, це покращений «мозок» у порівнянні з GPT 4o.

GPT 5 — анонсував себе найрозумнішою, найшвидшою і найбільш корисною моделлю з вбудованим мисленням.

5. Висновки

1. Колективні думки моделей ШІ-асистентів засновані на оцінці їх поточних можливостей і можуть бути корисними як додаткове до існуючих джерело актуальної інформації для попередньої (з послідовними уточненнями) оцінки доцільності їх використання по відношенню до визначеного класу задач. Поряд з цим, слід відзначити, що наведені матеріали «інтерв'ю» демонструють суттєву неоднозначність цих думок, іноді суперечливість і наявність обґрунтованих підстав для підозр у стійких «профіцитних» відхиленнях від об'єктивності. Можливі різні пояснення причин профіцитів, наприклад, своєрідна галюциногенність. Але якщо всі моделі для оцінок користуються одними й тими ж джерелами публічно доступної інформації, то найбільш імовірною причиною здається конкурентно-маркетингова. На користь цього можна сприйняти і помічені постійно занизькі кількісні і, особливо, вербальні оцінки можливостей DeepSeek з боку Gemini та Claude. Тому суттєвим мають бути саме зважена колективність думок і виключення власних самооцінок, тобто використання показників типу вищенаведених ЗЗО і ЗКО. Є доречним припущення, що професійним платним версіям моделей теж можуть бути притаманні такі ж властивості, бо ШІ-асистенти — це їх «молодші брати».

2. Поява GPT-5 має суттєво підняти планку вимог і оцінок ШІ-асистентів, на даний час кардинально змінюючи, як висловився Claude, «конкурентний ландшафт». Тому слід очікувати, що підходи до критеріїв порівняльних оцінок GPT-5 та інших моделей (які, ймовірно, з часом одержать свої значні поліпшення) мають базуватися на розгляді більш детальних їх властивостей, типу використання «збільшувального скла». Але на даний час поки що інформації від користувачів щодо досвіду використання мало і наведені наявні оцінки практично спиралися на самооцінку GPT-5.

СПИСОК ДЖЕРЕЛ

1. Choosing the right AI model for your task. URL: https://docs.github.com/en/copilot/using-github-copilot/ai-models/choosing-the-right-ai-model-for-your-task.

2. Evaluating generative AI tools: A comparative guide for new users. URL: https://www.researchgate.net/publication/385613747_Evaluating_generative_AI_tools_A_comparative_guide_for_new_users.
3. Comparison of Models: Intelligence, Performance & Price Analysis. URL: https://artificialanalysis.ai/models.
4. Deep Research Tools Comparison: A Side-by-Side Evaluation of AI Platforms. URL: https://cbtw.tech/insights/deep-research-tools-comparison-a-side-by-side-evaluation-of-ai-platforms.
5. Market Research AI Tools: A Comparative Analysis. URL: https://bytebridge.medium.com/market-research-ai-tools-a-comparative-analysis-ef8a90132877.
6. Top 10 AI Tools for Performance Reviews: A Comparative Analysis of Features and Benefits. URL: https://superagi.com/top-10-ai-tools-for-performance-reviews-a-comparative-analysis-of-features-and-benefits.
7. AI Tools Comparison made easy. URL: https://aireviewbattle.com.
8. Top 6 AI Evaluation Tools for Accurate Model Testing in 2025. URL: https://insight7.io/top-6-ai-evaluation-tools/.
9. Reddit — The heart of the internet. URL: https://www.reddit.com.
10. AI NOW — ШТучний інтелект. URL: https://x.com/AiNow_Ua.

Стаття надійшла до редакції 22.09.2025