

ПОРІВНЯЛЬНИЙ АНАЛІЗ АНАЛІТИЧНИХ МЕТОДІВ ТА АЛГОРИТМІВ РІЗНИХ LLM МОДЕЛІВ ШІ ПРИ РОЗВ'ЯЗУВАННІ ВИЗНАЧЕНИХ ІНТЕГРАЛІВ

Вінницький національний технічний університет

Анотація

Роботу присвячено порівняльному дослідженню можливостей сучасних нейромережових архітектур (GPT-4o, DeepSeek, Grok, Claude, Gemini) у розв'язанні задач інтегрального числення. На прикладі визначеного інтеграла, що потребує застосування універсальної тригонометричної підстановки, проаналізовано різні алгоритмічні стратегії: від використання готових аналітичних формул до покрокового академічного виведення результату. Особливу увагу приділено верифікації математичної точності та виявленню дефектів логіки обчислень. У ході дослідження ідентифіковано проблему «контекстуального дрейфу» — схильності ШІ до самовільної зміни вхідних параметрів задачі під впливом розповсюджених шаблонів, що критично впливає на достовірність результатів. Отримані дані дозволяють визначити межі застосування ШІ як допоміжного інструменту в курсі вищої математики.

Ключові слова: вища математика, штучний інтелект, LLM, визначений інтеграл, підстановка Вейерштрасса, символічні обчислення, контекстуальний дрейф, верифікація даних.

Abstract

The work is dedicated to a comparative study of the capabilities of modern neural network architectures (GPT-4o, DeepSeek, Grok, Claude, Gemini) in solving integral calculus problems. Using the example of a definite integral that requires the universal trigonometric substitution, various algorithmic strategies are analyzed: from the use of ready-made analytical formulas to step-by-step academic derivation of the result. Special attention is paid to the verification of mathematical accuracy and the detection of computational logic defects. During the study, the problem of «contextual drift» was identified — the tendency of AI to arbitrarily change the input parameters of a task under the influence of common patterns, which critically affects the reliability of the results. The obtained data allow for determining the boundaries of using AI as an auxiliary tool in the higher mathematics course.

Keywords: higher mathematics, artificial intelligence, LLM, definite integral, Weierstrass substitution, symbolic computation, contextual drift, data verification.

Вступ

Великі мовні моделі (LLM) штучного інтелекту — це сучасний інструмент символічних та чисельних обчислень, що активно інтегрується в освітній процес з вищої математики. На відміну від класичних систем комп'ютерної алгебри, які діють за жорсткими детермінованими алгоритмами, ШІ генерує розв'язки на основі ймовірнісних підходів, що дозволяє імітувати покрокове людське мислення, але водночас створює ризики математичних похибок.

Результати дослідження

Візьмемо для прикладу інтеграл $\int_0^{\frac{\pi}{2}} \frac{1}{2 + \cos x} dx$.

Для перевірки точності, спочатку розв'яжемо вираз за допомогою системи комп'ютерної алгебри, а потім порівняємо з результатами отриманими з використанням сучасних нейромережових архітектур.

Для верифікації було використано систему комп'ютерної алгебри WolframAlpha, яка надала точний аналітичний результат: $I = \frac{\pi}{3\sqrt{3}} \approx 0,6046$. Основним методом розв'язання було визначено

універсальну тригонометричну підстановку $t = tg\left(\frac{x}{2}\right)$, що приводить підінтегральну функцію до

раціонального вигляду $\frac{2}{t^2 + 3}$. Цей результат прийнято за еталон (Ground Truth) для подальшого порівняння нейромережових моделей.

Результати досліджень

GPT-4o — флагманська мультимодальна ШІ-модель від OpenAI. Для розв’язання даного інтеграла використала готову складну функцію $\int \frac{dx}{a + b \cos(x)}$. Розв’язання виглядає дуже акуратно і швидко веде

до відповіді. З мінусів можна зазначити відсутність еталонного рішення, бо маловідома формула може заплутати користувача.

DeepSeek — потужна генеративна модель ШІ та однойменна китайська дослідницька компанія. При розв’язанні вказана модель використала правильну термінологію «підстановка Вейерштрасса». Точність результату в межах $[0, 1]$. Особливістю розв’язку можна зазначити те, що дана модель використала запис через квадрат $\sqrt{3}^2$ у знаменнику — це дуже хороший педагогічний хід, який полегшує перевірку формули арктангенса.

Grok — чат-бот зі ШІ, створений компанією xAI Ілона Маска. Дуже швидко будує складні математичні вирази. Навіть при помилковій відповіді, сама структура розв’язання виглядає професійно та послідовно. Найбільша слабкість — це здатність самовільно змінити умову задачі (наприклад, межу інтегрування) під впливом типових шаблонів, то ж потребує обов’язкової перевірки кожного кроку, оскільки може видати «правильне рішення іншої задачі».

Claude 3.5 — це один із націпрогресивніших чат-ботів та ШІ на основі великих мовних моделей (LLM), його розробила компанія Anthropic. Проявляє себе як «педагог-аналітик», чиєю беззаперечною перевагою є найвищий рівень математичної культури, використання точної термінології та ідеальне візуальне оформлення символічних записів із наочним виділенням констант, хоча надмірна обережність у твердженнях та значна кількість супровідного тексту іноді можуть розмивати фокус безпосередньо з самих обчислювальних операцій.

Gemini — це сучасний мультимодальний ШІ від компанії Google. Функціонує як «системний контролер», що зосереджується на логічній цілісності процесу та акцентує на критичних точках верифікації, як-от коректність заміни змінної та перевірка області визначення функції, вдало балансує між швидкістю та деталізацією, проте іноді демонструє надмірну багатослівність у поясненнях очевидних математичних переходів, що може дещо ускладнювати сприйняття сухого технічного звіту.

Для узагальнення результатів проведеного аналізу та наочного порівняння особливостей роботи різних LLM-моделей отримані спостереження було систематизовано у вигляді таблиці. У ній відображено основні підходи моделей до розв’язання математичних задач на прикладі визначеного інтеграла, рівень точності отриманих результатів, а також характерні особливості формування пояснень і представлення розв’язку.

Табл. 1. Порівняння якості розв’язання математичних задач різними LLM-моделями

Модель	Методологія	Точність	Пояснення
GPT-4o	Формульний підхід	100%	Стисле, орієнтоване на довідник.
DeepSeek	Академічне виведення	100%	Детальне, з покроковою заміною.
Grok	Шаблонна підстановка	0%	Помилкове (зміна умов).
Claude 3.5	Педагогічний аналіз	100%	Висока наочність перетворень.
Gemini	Системна верифікація	100%	Акцент на перерахунку меж.

Висновки

Проведене дослідження дозволяє стверджувати, що вибір конкретної моделі ШІ для розв'язання задач вищої математики має базуватися на пріоритетах користувача, оскільки кожна архітектура демонструє унікальні функціональні переваги та обмеження.

GPT-4o виявляється найбільш ефективним у ситуаціях, що потребують миттєвої верифікації фінального результату складних обчислень. Завдяки доступу до широкої бази аналітичних шаблонів, модель працює як потужний «інженерний калькулятор». Її доцільно використовувати для перевірки відповідей у прикладних задачах, де покрокове виведення не є критично важливим, а головним критерієм є швидкість.

DeepSeek є оптимальним інструментом для освітнього процесу та підготовки академічних звітів. Здатність моделі до детального відтворення класичних алгоритмів дозволяє використовувати її як інтерактивний посібник. Вона є незамінною у випадках, коли необхідно продемонструвати повний логічний ланцюжок перетворень — від вибору заміни змінної до фінального спрощення виразу.

Claude 3.5 демонструє найкращі результати при підготовці наукових публікацій та тез конференцій. Високий рівень математичної культури та ідеальне візуальне оформлення символічних записів роблять цю модель найкращим вибором для ситуацій, де важлива не лише правильність обчислень, а й професійна презентація результатів із використанням точної термінології.

Gemini найкраще проявляє себе в задачах, що потребують підвищеного контролю за логічною цілісністю та верифікацією «тонких місць» алгоритму. Цю модель варто обирати для складних інтегралів, де існує високий ризик помилитися в області визначення чи при перерахунку меж інтегрування, оскільки вона акцентує увагу саме на системній перевірці цих вузлів.

Отже, найкращим варіантом для навчання та студентської роботи є DeepSeek. Він продемонстрував найбільш «чесний» математичний підхід. Замість того, щоб просто видати цифру, він крок за кроком відтворює логіку підстановки Вейерштрасса, перераховує межі та спрощує вирази. Якщо метою є зрозуміти хід розв'язання, то DeepSeek наразі є найбільш надійним асистентом.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Фіхтенгольц Г. М. Курс диференціального та інтегрального числення .Т. 3. - Київ : Наукова думка, 2003. 800 с.
2. Мейзлер Д. Г. Обчислення інтегралів: навчальний посібник для студентів технічних спеціальностей. - Вінниця: ВНТУ, 2018.
3. Imani S., Du L., Shrivastava H. MathPrompter: Mathematical Reasoning using Large Language Models. *arXiv preprint arXiv:2303.05398*. 2023.

Поліщук Максим Олександрович – студент групи 2 БС-25б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: 04-25-128.stud@vntu.edu.ua

Тютюнник Оксана Іванівна – доцент кафедри вищої математики, Вінницький національний технічний університет, м. Вінниця, e-mail: tyutyunnik@vntu.edu.ua

Maxym Polischuck – student of group 2 SS-25b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: 04-25-128.stud@vntu.edu.ua

Oksana Tiutiunyk – PhD (Pedagogics), Associate Professor of the Department of Higher Mathematics, Vinnytsia National Technical University, e-mail: tyutyunnik@vntu.edu.ua