The background features a complex network of grey and red lines and nodes, overlaid with mathematical symbols and diagrams. Visible elements include a limit formula  $f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$ , a triangle with dashed lines, and a function graph labeled  $f(x)$ .

# Розробка чат-боту для розв'язання математичних задач на основі L1emna

Виконала: студентка 4-го курсу навчання,  
Освітньої програми «Прикладна математика», 113  
Сапко Марія Сергіївна  
Науковий керівник: Швай Н.О.

# Мета та основні завдання роботи

- Дослідження ефективності моделі `lemma` та її компетентність у вирішенні математичних задач різної складності.
- Виявлення унікальних особливостей притаманні моделі, щоб визначити перспективи її використання та можливості застосування, зокрема в освітньому процесі.
- Проведення порівняльного аналізу з іншими відомими чат-ботами: ChatGPT, Claude, Gemini та Perplexity AI.
- Створення навчального помічника на основі вищезгаданої моделі.

# Модель Llemma

- EleutherAI
- Code Llama + Proof-Pile-2 = Llemma
- Llemma\_7b, llemma\_34b
- Proof-Pile-2 (55 мільярдів токенів):
  - AlgebraicStack (code): 11
  - OpenWebMath (web): 15
  - ArXiv (papers): 29

		GSM8k	OCW	MMLU-STEM	SAT	MATH
Llama 2	7B	11.8%	3.7%	29.9%	25.0%	3.2%
Code Llama	7B	10.5%	4.4%	25.1%	9.4%	4.5%
Minerva	8B	16.2%	7.7%	35.6%	-	14.1%
LLEMMA	7B	<b>36.4%</b>	<b>7.7%</b>	<b>37.7%</b>	<b>53.1%</b>	<b>18.0%</b>
Code Llama	34B	29.6%	7.0%	40.5%	40.6%	12.2%
LLEMMA	34B	<b>51.5%</b>	<b>11.8%</b>	<b>49.0%</b>	<b>71.9%</b>	<b>25.0%</b>
Minerva	62B	52.4%	12.0%	53.9%	-	27.6%
Minerva	540B	58.8%	17.6%	63.9%	-	33.6%

# Спосіб проведення дослідження

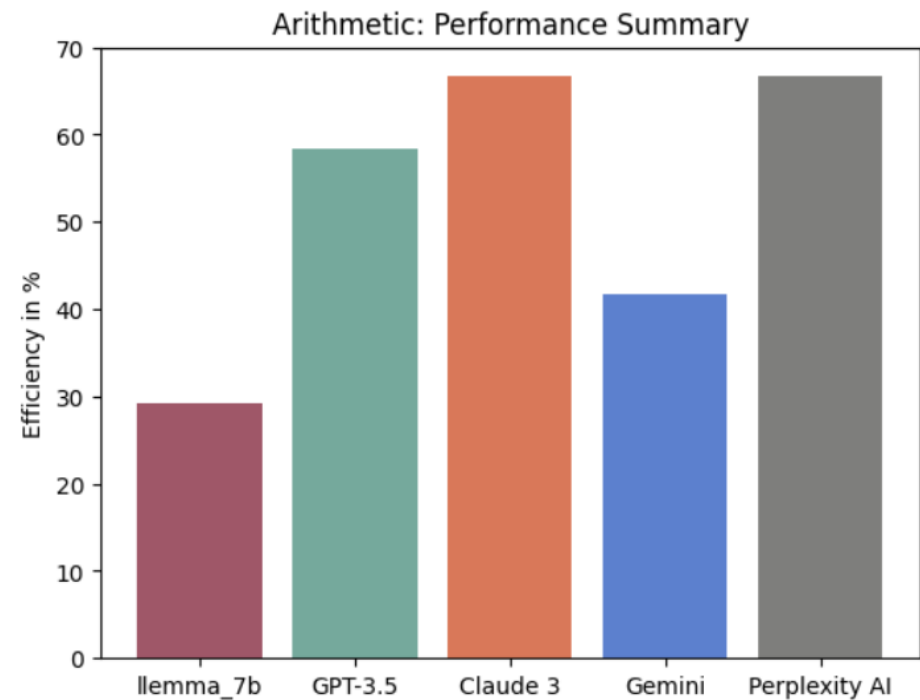
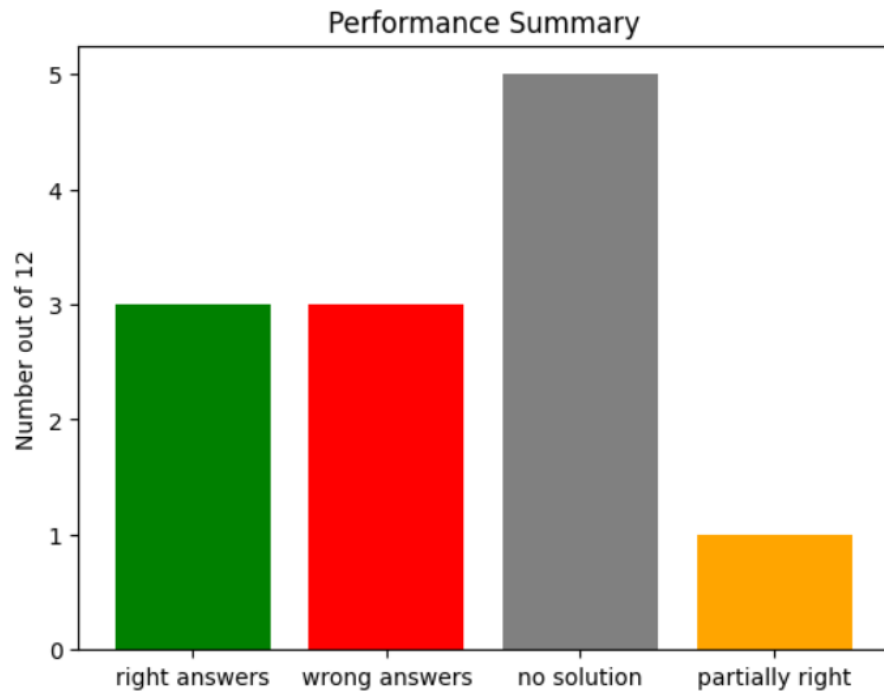
- Набір даних для тестування:
  - 12 розділів, що містять по 12 завдань: 11 розділів математики і завдання підвищеної складності та на доведення
- 95% запитів сформовані англійською мовою, інші 5% написані мовами Python та Isabelle
- Максимальна кількість символів вихідного результату моделі: 200 – 700 символів. (200 – 4 хвилин, 700 – 13 хвилин.

$$\bullet SR = \frac{r + \frac{1}{2}p}{a} * 100\%,$$

Де  $r$  – кількість правильних відповідей,  $p$  – кількість частково правильних відповідей,  $a$  – кількість усіх завдань.

# Арифметика

$$\frac{f(x+h)-f(x)}{h}$$



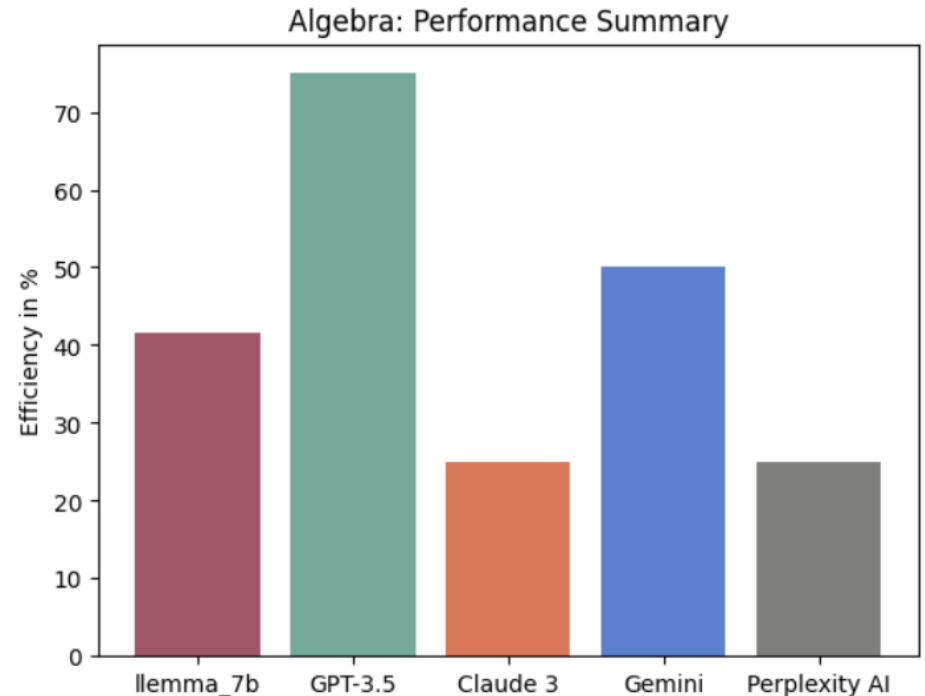
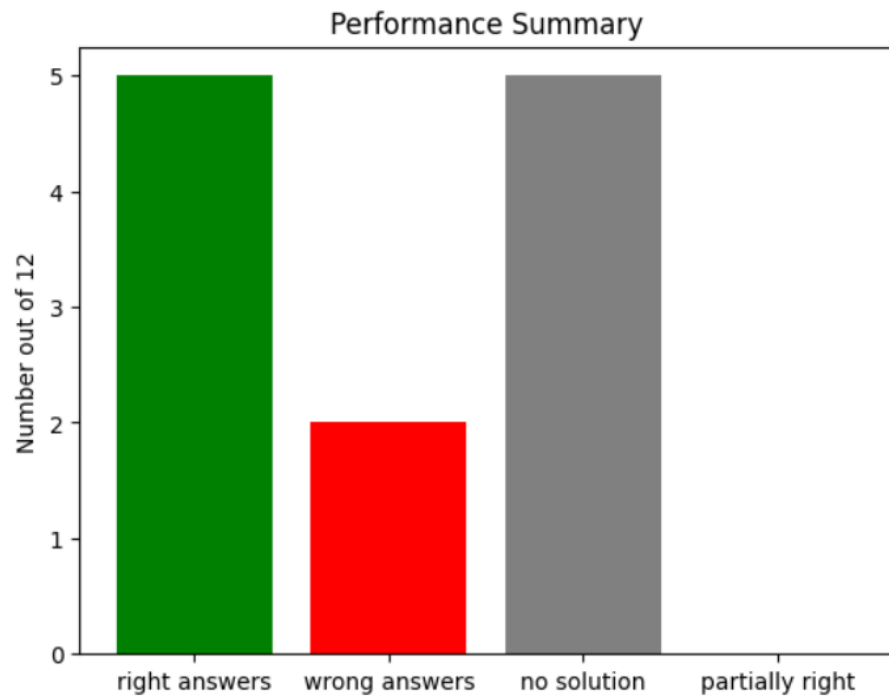
**Особливості llerna\_7b:** найбільш оптимальний формат подання задачі (mod y)

- це  $(x ** y) \% z$

**Загальні тенденції:** неточності в обчисленнях, де фігурують двоцифрові числа і більше

# Алгебра

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

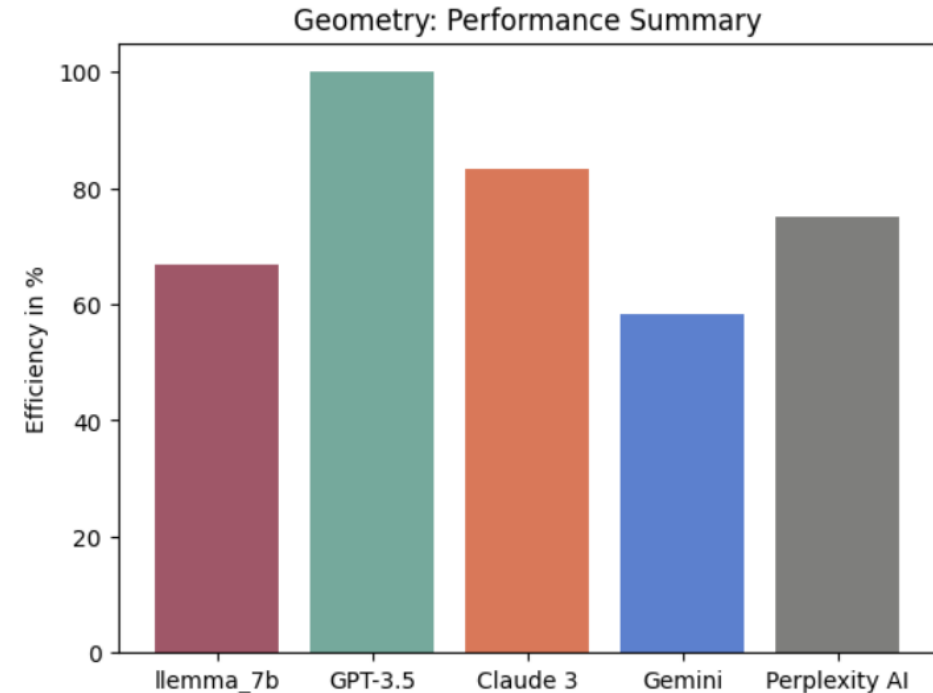
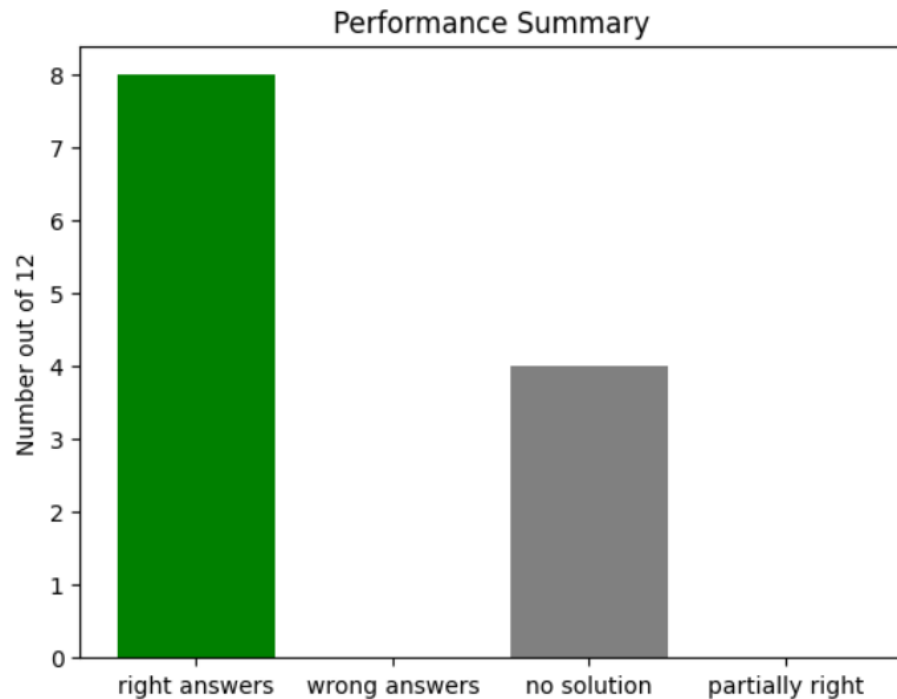


**Особливості llama\_7b:** ефективна в проведенні розрахунків з дробовими числами

**Загальні тенденції:** складнощі у завданнях на розкриття дужок та спрощення виразів

# Геометрія

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

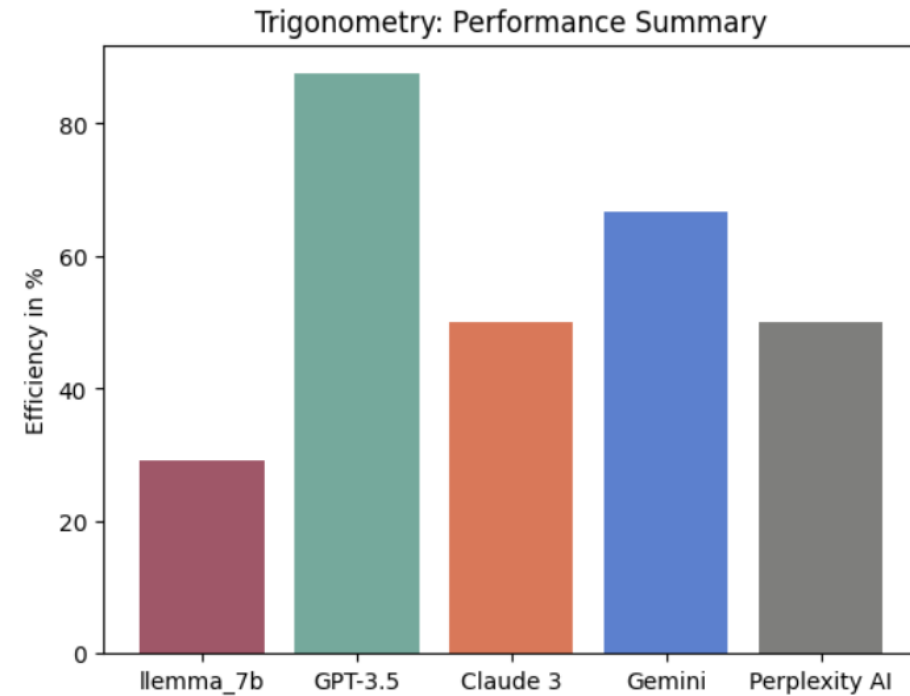
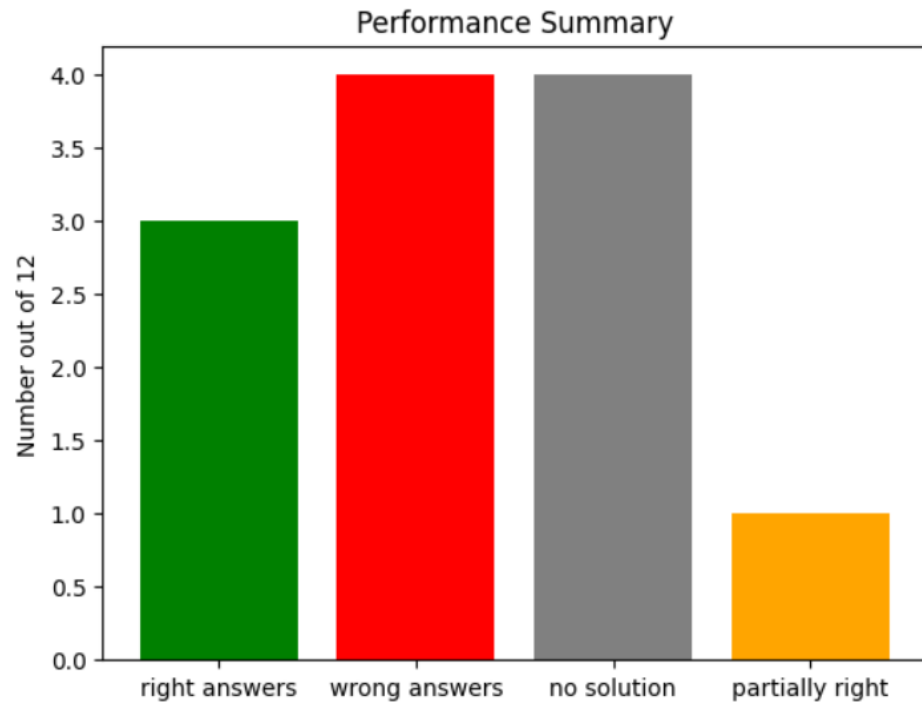


**Особливості llerna\_7b:** правильно розв'язує однокрокові завдання з невеликими числами

**Загальні тенденції:** коректно визначені формули, допущені помилки у розрахунках

# Тригонометрія

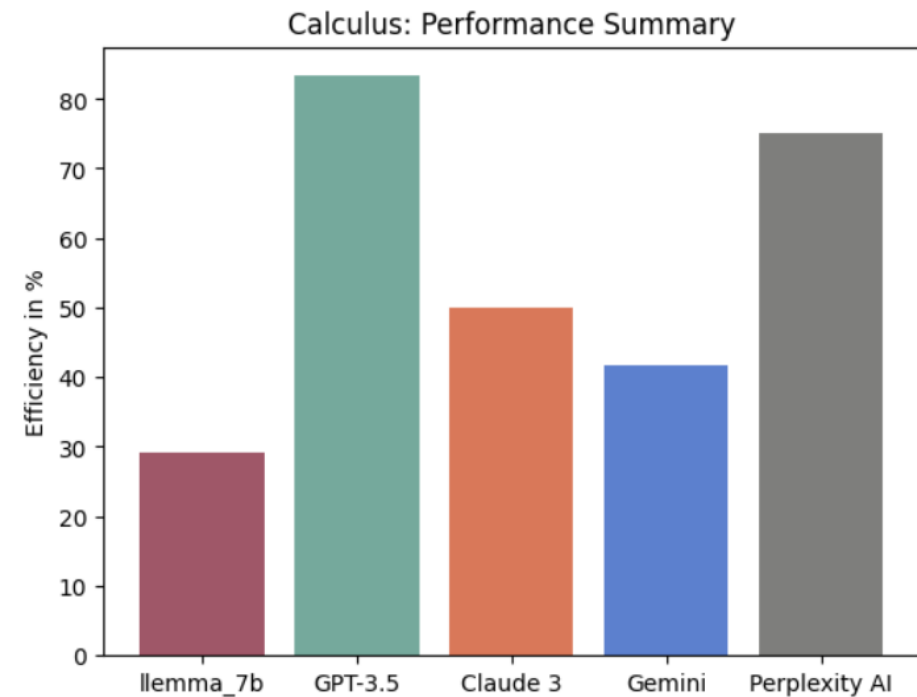
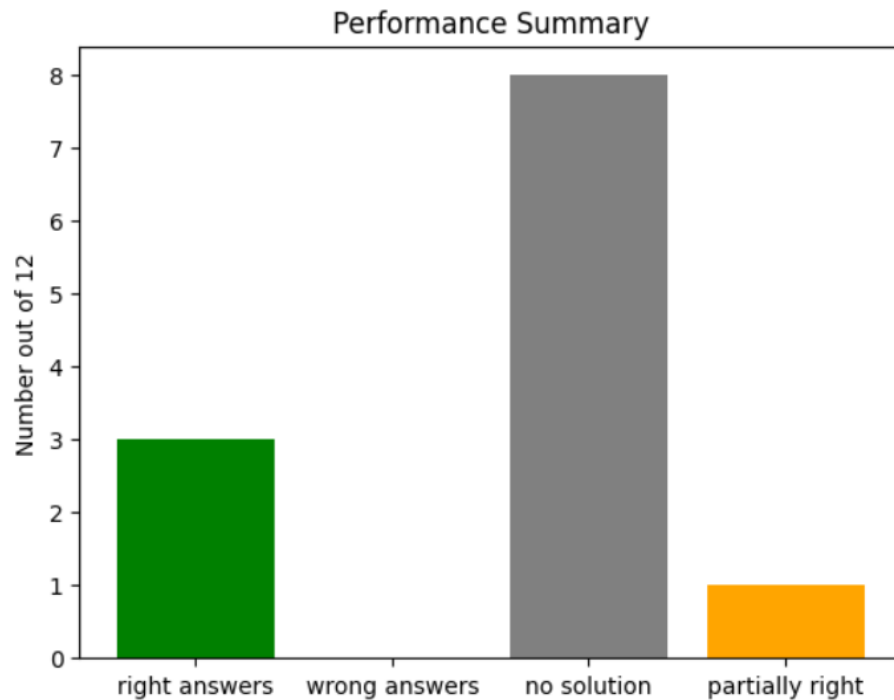
$$\frac{f(x+h)-f(x)}{h}$$



**Особливості llerna\_7b:** нездатність відтворювати значення тригонометричних функцій для стандартних табличних кутів

**Загальні тенденції:** складнощі у розв'язанні тригонометричних рівнянь

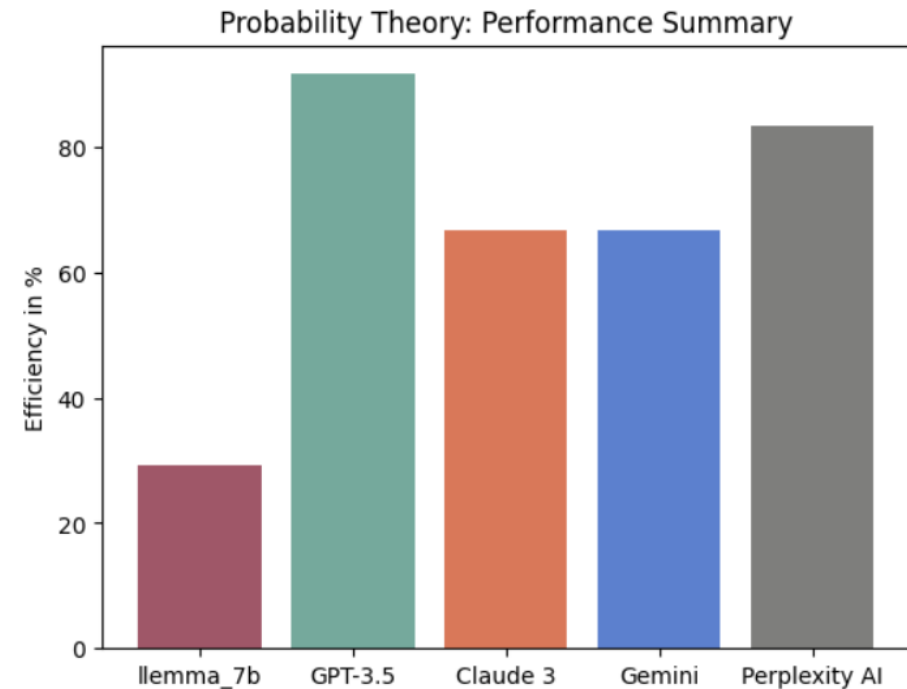
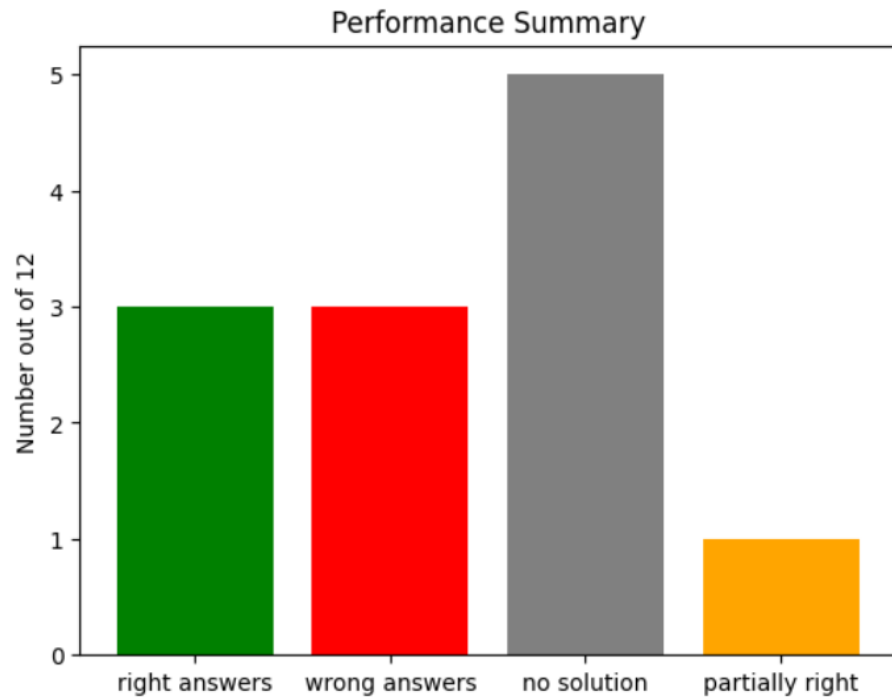
# Математичний аналіз



**Особливості llama\_7b:** успішна у розв'язанні завдань на похідну, не ефективна для обчислень інтегралів

**Загальні тенденції:** неточності в обчисленні обмежених інтегралів та максимального значення функції

# Теорія ймовірності

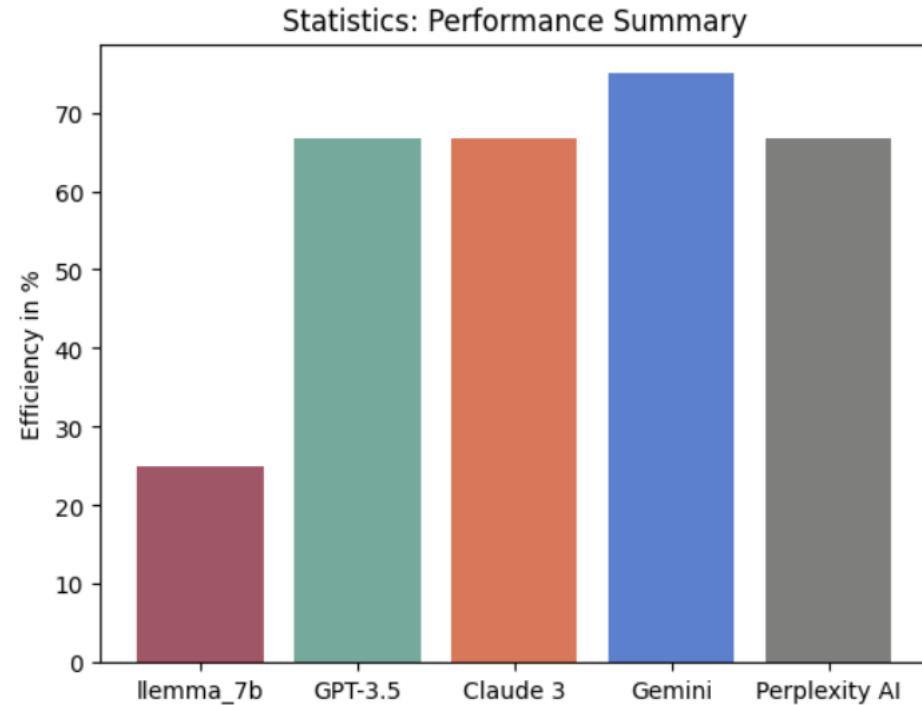
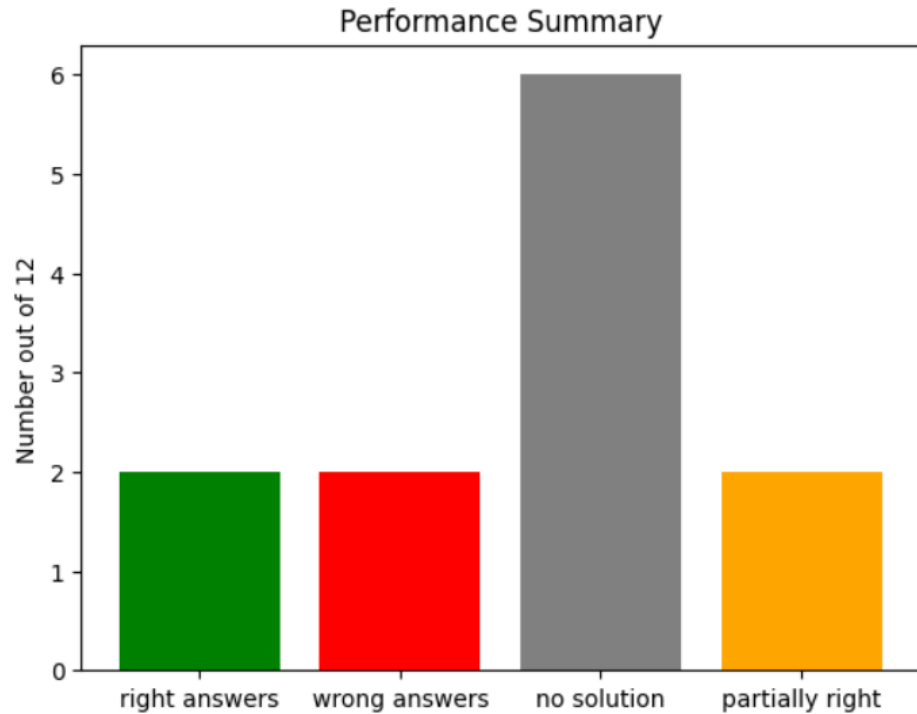


**Особливості llerna\_7b:** потребує формулювання умови в контексті тематики книг, кульок або грального кубика

**Загальні тенденції:** складнощі у виявленні та розв'язанні завдань на умовну ймовірність 9

# Статистика

$$\frac{f(x+h)-f(x)}{h}$$

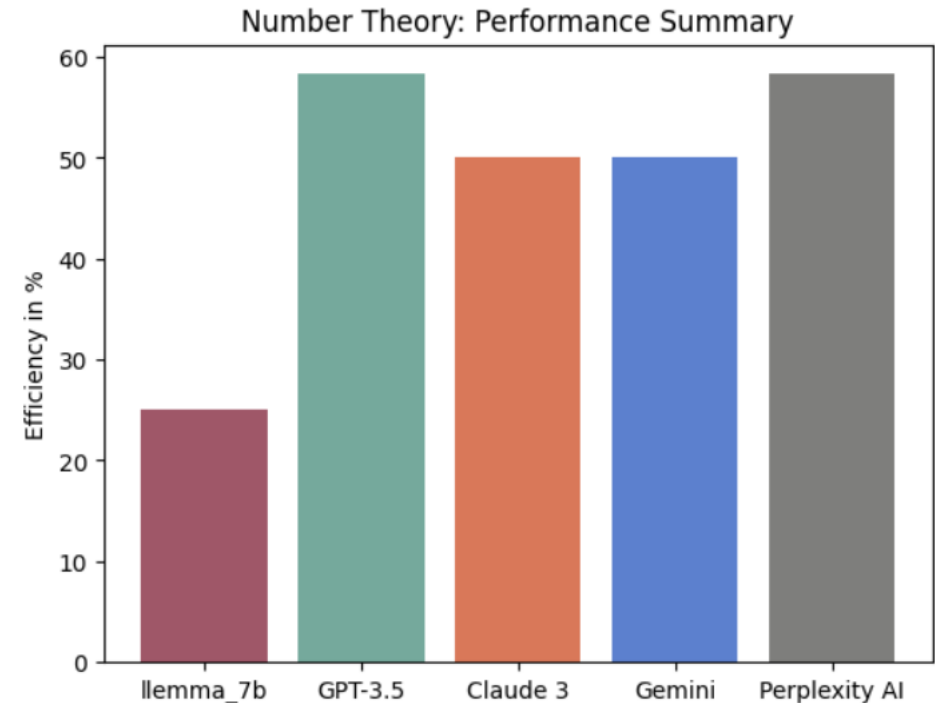
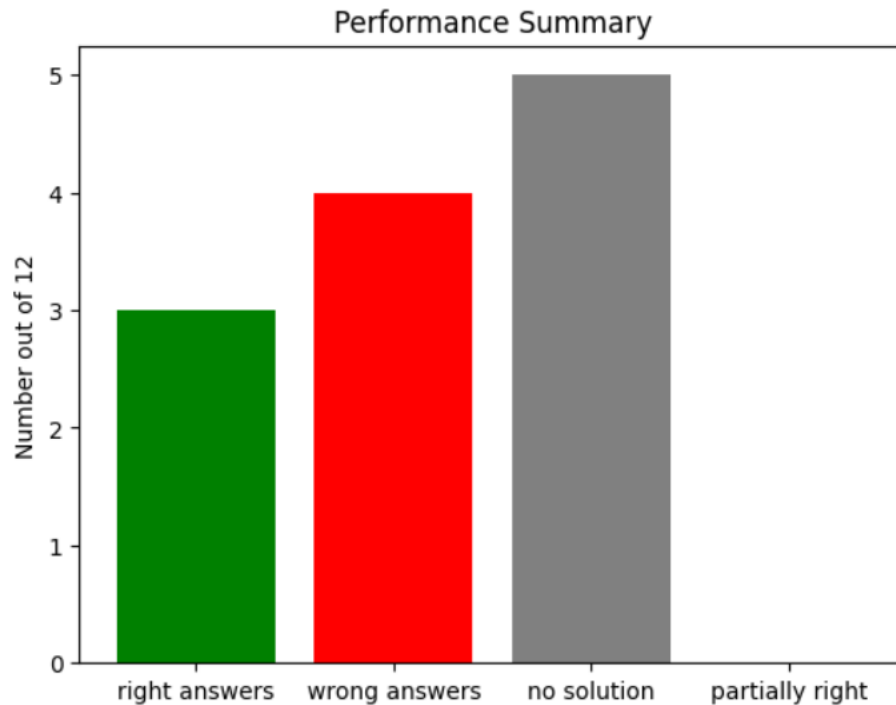


**Особливості llemma\_7b:** неефективна для даного розділу. Може визначити лише моду та розмах вибірки на основі числового ряду.

**Загальні тенденції:** неточності у розумінні поняття кватилів та міжквартильного розмаху

# Теорія чисел

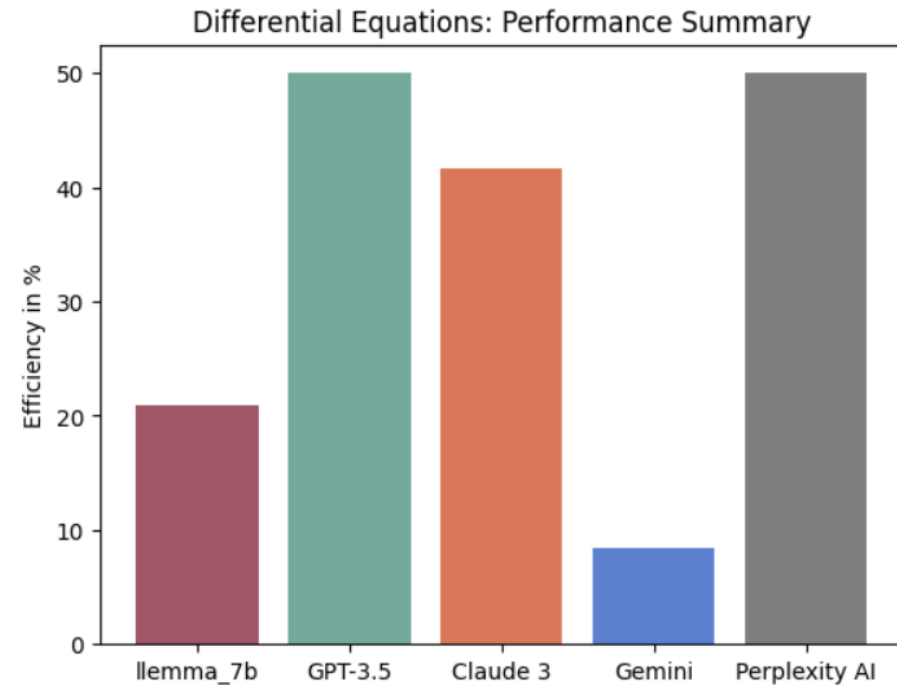
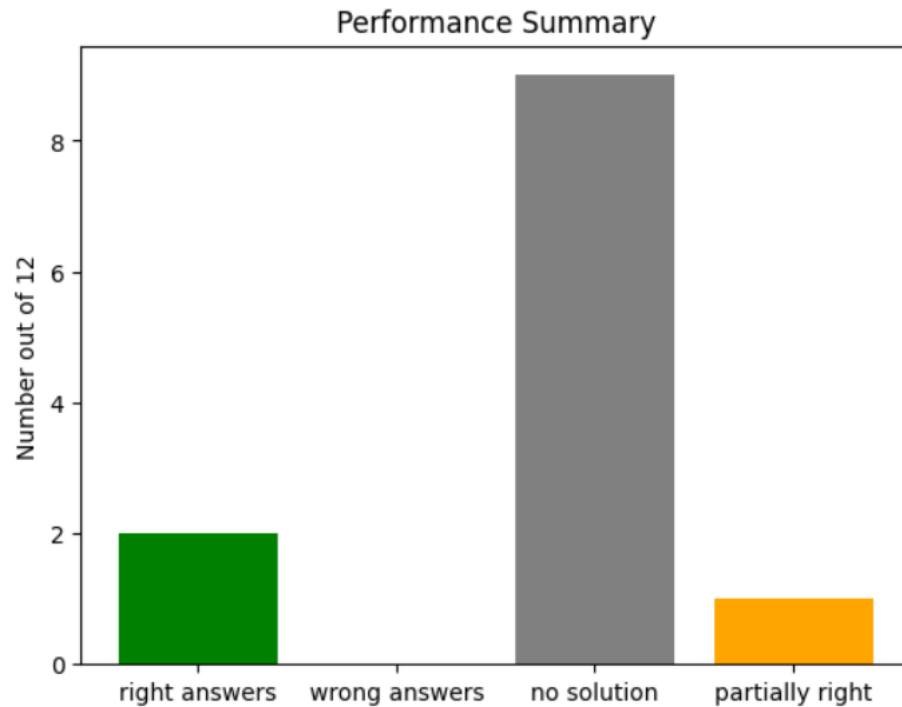
$$\frac{f(x+h)-f(x)}{h}$$



**Особливості llama\_7b:** ефективна лише для найпростіших завдань даного розділу (пошук НСД, НСК невеликих чисел)

**Загальні тенденції:** неспроможність визначити чи достатньо даних в умові для розв'язання завдання та вирішити завдання середнього та високого рівнів складності

# Диференціальні рівняння

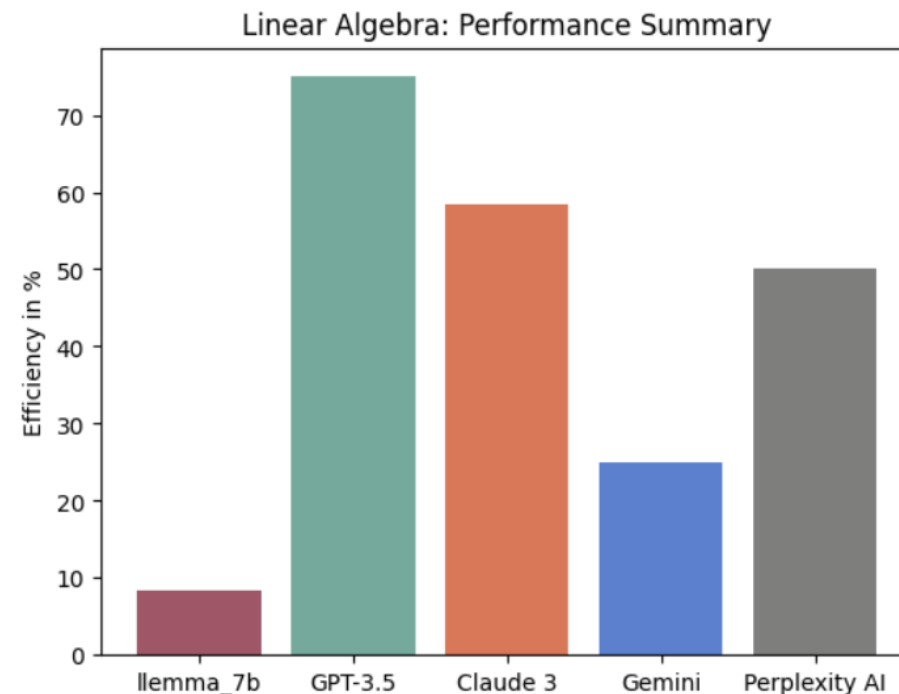
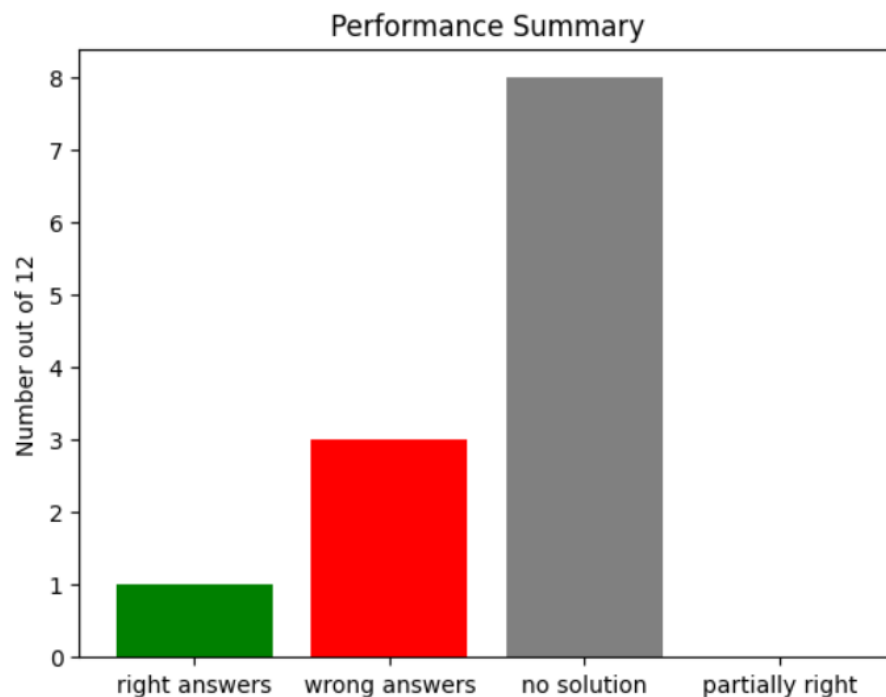


**Особливості llerna\_7b:** здатна розв'язати лише рівняння невеликої складності

**Загальні тенденції:** ефективні для розв'язання простих екземплярів кожного типу рівнянь

# Лінійна алгебра

$$\frac{f(x+h)-f(x)}{h}$$

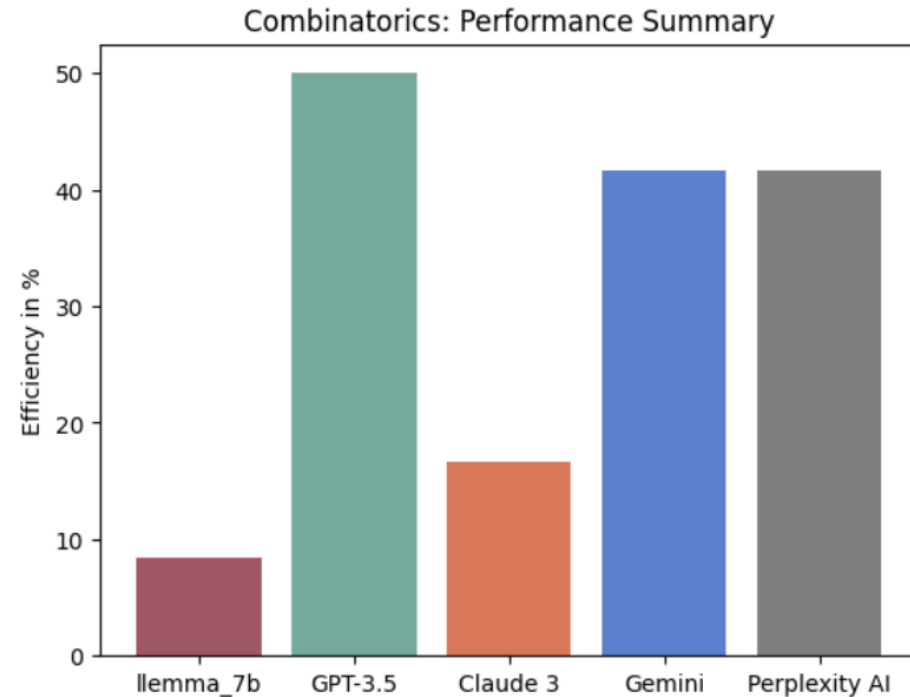
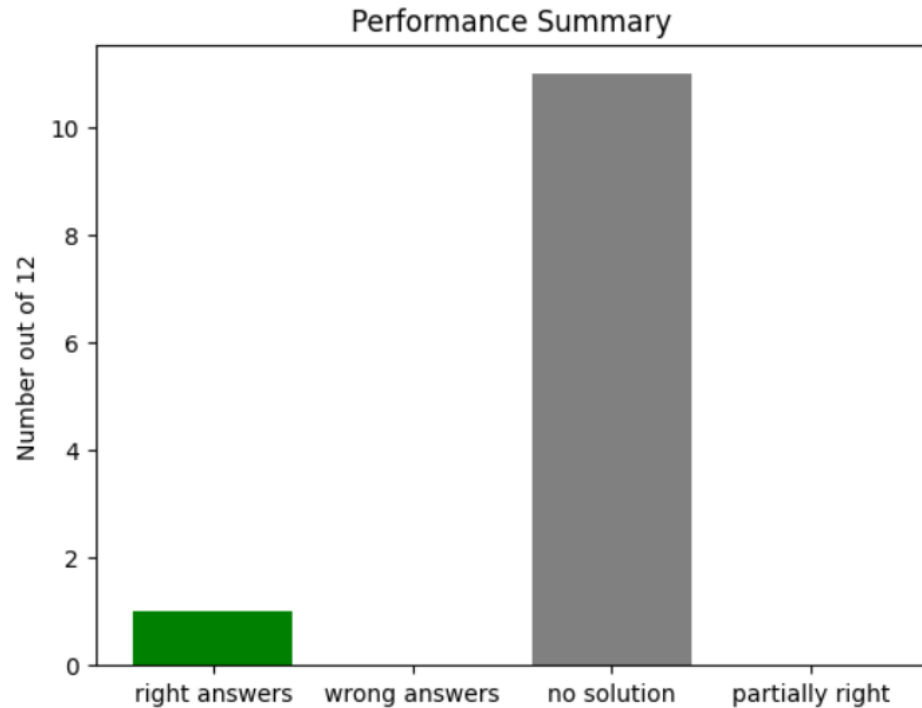


**Особливості lllemma\_7b:** неефективна для даного розділу. Може виконувати арифметичні дії над матрицями  $2 \times 2$ .

**Загальні тенденції:** складнощі у роботі з матрицями  $3 \times 3$  і більше, нездатність вирішити систему рівнянь та визначити власні числа і власні вектори

# Комбінаторика

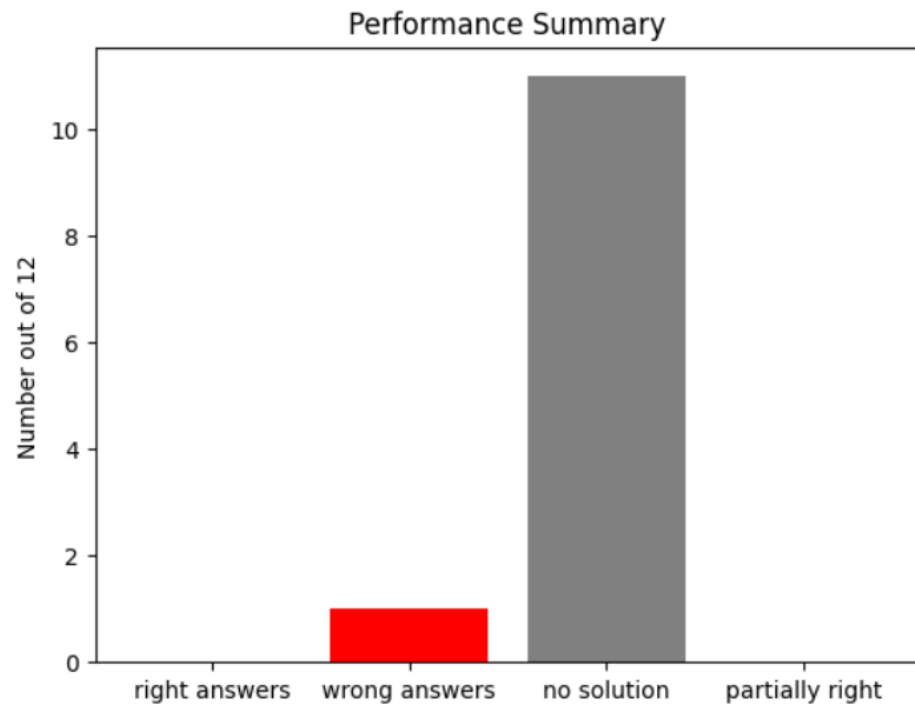
$$\frac{f(x+h)-f(x)}{h}$$



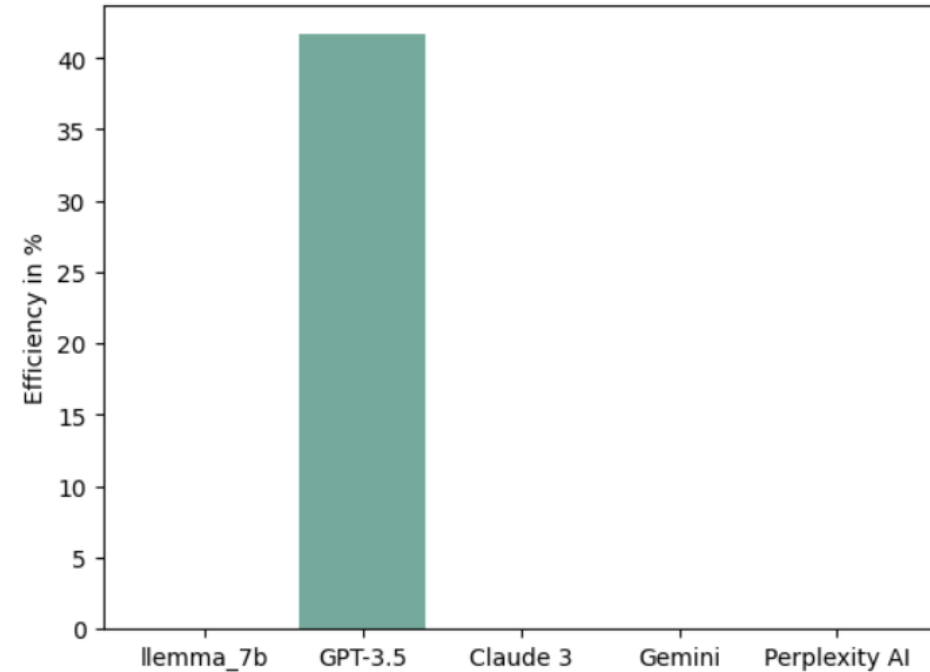
**Особливості llemma\_7b:** не здатна виявити необхідні патерни у неструктурованій формі подання умови

**Загальні тенденції:** не завжди враховують повну умову завдання, опускаючи маленькі, але важливі деталі

# Задачі підвищеної складності та на доведення



Tasks of increased complexity and requiring proof: Performance Summary



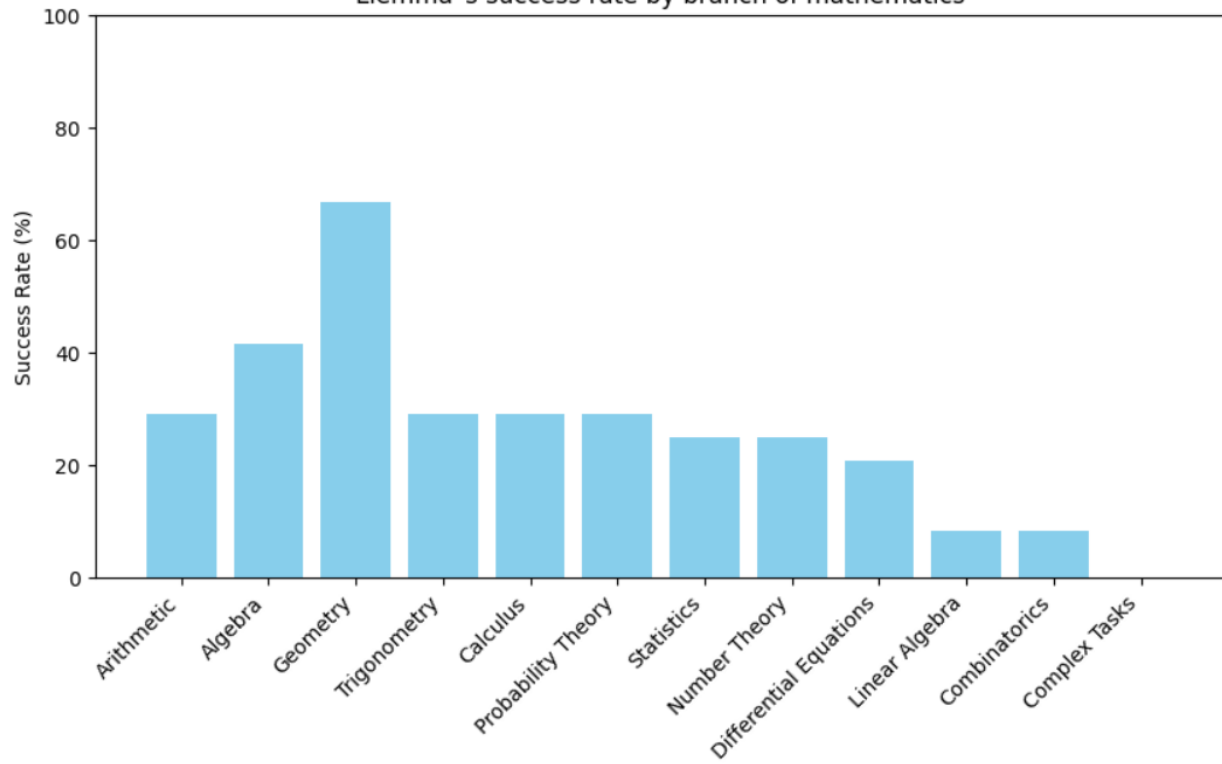
**Особливості llemma\_7b:** зміна мов подання умови не впливає на успішність моделі в інтерпретації завдання

**Загальні тенденції:** не здатні вирішити завдання цього розділу

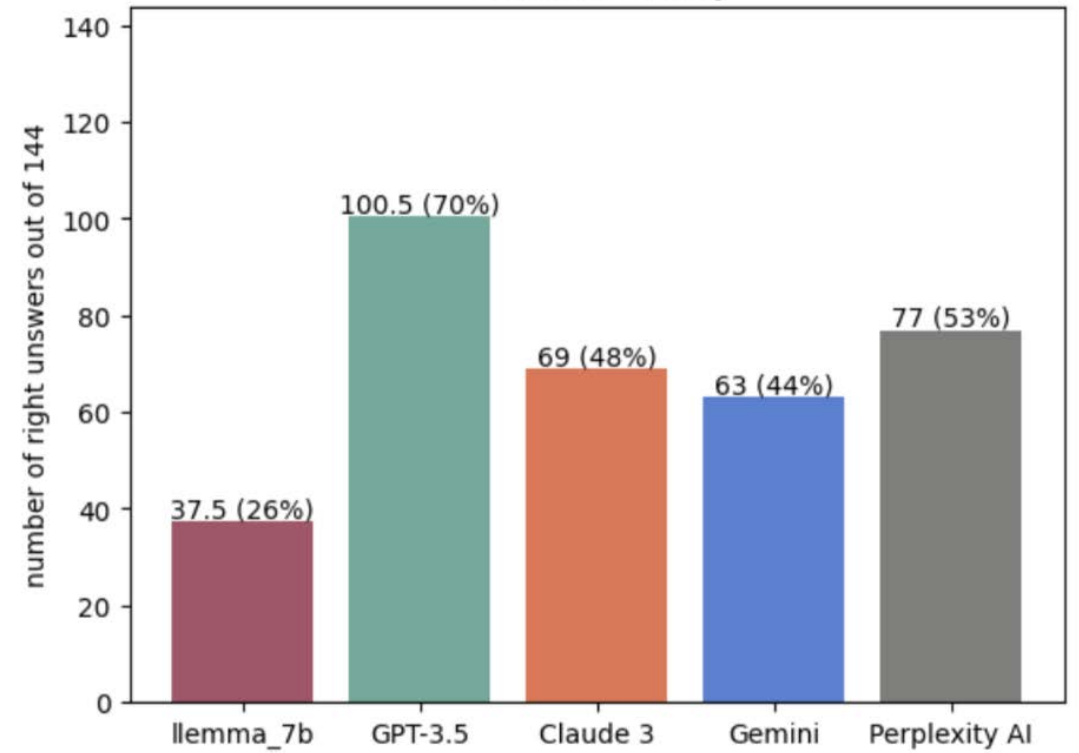
# Підсумки

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Llemma`s success rate by branch of mathematics



Overall summary



# LEMMA\_CHATBOT

**lemma\_7b:** Hello! I'm a chatbot specialized in solving mathematical problems. How can I assist you today?

**user:** Solve the first-order ordinary differential equation  $dy/dx = x^2 + 1$  with initial condition  $y(0) = 1$

**lemma\_7b:** Solve the first-order ordinary differential equation  $dy/dx = x^2 + 1$  with initial condition  $y(0) = 1$ .

Solution:

$$dy/dx = x^2 + 1$$

$$dy = (x^2 + 1) dx$$

Integrating both sides, we get

$$\int dy = \int (x^2 + 1) dx$$

$$\Rightarrow y = x^3/3 + x + C$$

Now,  $y(0) = 1$

$$\Rightarrow 1 = 0/3 + 0 + C$$

$$\Rightarrow C = 1$$

Therefore, the solution of the differential equation is  $y = x^3/3 + x + 1$ .

Type your message...

Send

**user:** Determine the probability of selecting 2 blue marbles from a bag containing 5 blue marbles and 3 red marbles without replacement

**lemma\_7b:** Determine the probability of selecting 2 blue marbles from a bag containing 5 blue marbles and 3 red marbles without replacement.

Solution:

The probability of selecting 2 blue marbles from a bag containing 5 blue marbles and 3 red marbles without replacement is given by

$$P(2 \text{ blue marbles}) = P(1\text{st blue marble}) \times P(2\text{nd blue marble})$$

$$= 5 / 8 \times 4 / 7$$

$$= 5 / 14$$

Therefore, the probability of selecting 2 blue marbles from a bag containing 5 blue marbles and 3 red marbles without replacement is  $5 / 14$ .

**user:** Find the least common multiple (LCM) of 24 and 36

**lemma\_7b:** Find the least common multiple (LCM) of 24 and 36.

Solution:

The LCM of 24 and 36 is 72.

Type your message...

Send

**lemma\_7b:** Hello! I'm a chatbot specialized in solving mathematical problems. How can I assist you today?

**user:** Proof of Fermat's theorem on the sum of two squares: Prove that the equation  $x^2 + y^2 = z^2$  has infinitely many solutions in natural numbers, except for  $x = y = z = 0$


**lemma\_7b:** Sorry, I can't solve this task

Type your message...

Send

# Висновки

- Створено набір даних для комплексного аналізу моделі
- Проведене тестування моделі `lemma_7b`, на основі якого зроблені якісний та кількісний аналізи
- Проведено порівняльний аналіз із найвідомішими моделями
- Визначено сферу застосування чат-боту на основі моделі `Lemma`
- Створено `LEMMA_CHATBOT`


$$\int_V \operatorname{div} F \, dV = \int_{\partial V} F \cdot d\Sigma$$

Дякую за увагу!

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

$F(x)$

$f(t) dt$