

## ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ МАТЕМАТИЧЕСКОГО АНАЛИЗА

*Дроздов А.И., студент гр.250502*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Луцакова И. Н. – канд. физ.-мат. наук*

**Аннотация.** В связи с быстрым ростом нейронных сетей в этом году данная работа проводит сравнительный анализ трех самых популярных решений в прикладных задачах математического анализа. Сравниваются ChatGPT (GPT-4), Google Bard и Bing Chat по введённой системе оценки.

**Ключевые слова.** Нейронные сети, математический анализ, chatgpt, google bard, bing chat.

В последнее время наблюдается бум в развитии нейронных сетей (основанных на технологии GPT – Generative Pre-Trained Transformer). В связи с постоянным потоком новостей о невероятных результатах этих нейронных сетей в совершенно разных сферах [5], цель этой работы – проанализировать самые популярные из решений на задачах математического анализа.

Что такое нейронные сети? Нейронные сети – это метод в искусственном интеллекте, который учит компьютеры обрабатывать данные таким же способом, как и человеческий мозг. Используется множество слоев, начиная от входного, проходя через огромное количество промежуточных слоев, и до выходного. Нейронная сеть способна учиться на своих ошибках на предыдущих слоях.

Что такое GPT, который лежит в основе всех нейронных сетей, рассматриваемых в этой работе? GPT – это особый вид трансформера (модель глубокого обучения, которая по-разному взвешивает каждую часть входных данных и анализирует каждую часть данных с полным контекстом из предыдущих данных), которому не нужно обучение с учителем. Это позволяет гораздо быстрее тренировать сети на гораздо больших наборах данных. Обучение состоит из двух этапов: обучение и настройка параметров на большом объеме необработанных данных, а затем более точная “подкрутка” под финальный результат под наблюдением.

Было выбрано 3 решения: ChatGPT (GPT-4) [2], Google Bard [3] и Bing Chat [4]. Использовались примеры из сборника задач [1].

Для сравнения была создана своя шкала оценки: каждый пример оценивается по 3 критериям (от 1 до 10): корректность (правильно ли решен пример), рациональность (можно ли было его решить проще) и полнота объяснения (насколько бы было понятно решение студентам).

Для оценки результатов использовались следующие формулы:

$$S_i = [(correct_i + rational_i + clear_i)/3], \quad (1)$$

где  $S_i$  – оценка нейронной сети по конкретному примеру, а  $correct_i$ ,  $rational_i$ ,  $clear_i$  – оценка примера по корректности, рациональности и полноте объяснения соответственно.

$$S = \left[ \frac{\sum_{i=1}^n S_i}{n} \right], \quad (2)$$

где  $S$  – итоговая оценка нейронной сети, а  $n$  – число примеров.

В работе приведены наиболее характерные скриншоты ответов и их краткая характеристика.

Теория вещественных чисел и множеств

1. Доказать, что  $1^2 + 2^2 + \dots + n^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{6}$

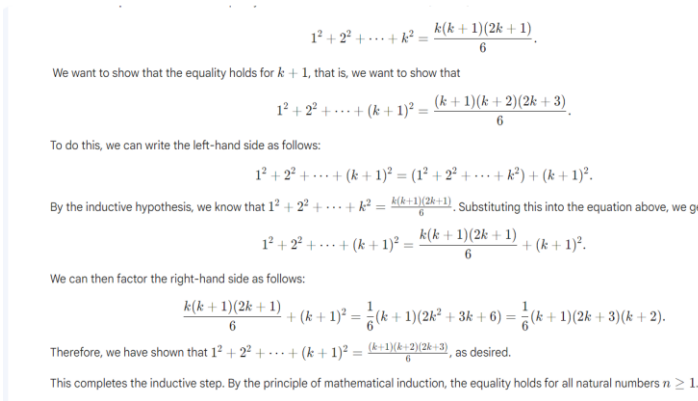
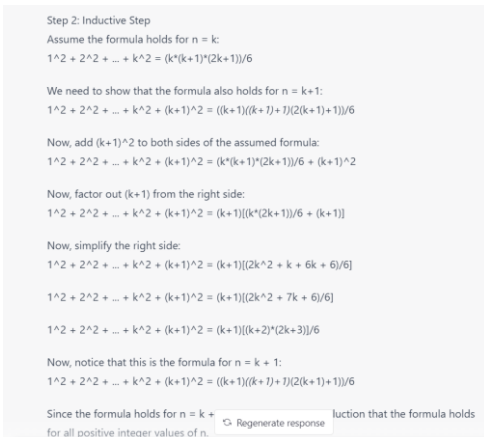


Рисунок 1 – Ответ ChatGPT

Рисунок 2 - Ответ Google Bard

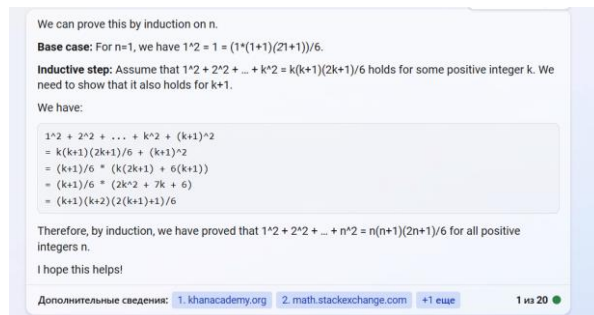


Рисунок 3 – Ответ Bing Chat

Как видно из скриншотов, все сети смогли решить данный пример. Однако, в форматировании ответа и объяснениях были некоторые проблемы.

Например, в проверке базы индукции ChatGPT вывело 21 вместо  $2 \cdot 1$  (наверное, подставляя 1 вместо  $n$ , не заметив, что получилось другое число). Также использовались квадратные скобки вместо круглых (такая проблема есть в старых версиях MathCAD), не использовался отформатированный вывод. Объяснение даже слишком подробное (обычно нет нужды описывать словами каждый шаг, только если это не что-то неочевидное).

Google Bard имеет хорошее и рациональное решение. По объему объяснений не слишком много и не слишком мало – золотая середина. Однако он допустил ошибку в конце доказательства:

## 59-я научная конференция аспирантов, магистрантов и студентов БГУИР

он сказал, что  $\frac{1}{6}(k+1)(2k^2+3k+6) = \frac{1}{6}(k+1)(2k+3)(k+2)$ , однако правильно  $\frac{1}{6}(k+1)(2k^2+7k+6) = \frac{1}{6}(k+1)(2k+3)(k+2)$ , что очень похоже на “подгонку под ответ”.

Bing Chat же также имел проблемы с форматированием  $2*1$ , но решил правильно. Объяснение довольно краткое и может запутать изначально.

Оценка (correct/rational/clear):

ChatGPT: 10/9/6

Google Bard: 7/10/10

Bing Chat: 10/10/8

2. Доказать, что  $n! < \left(\frac{n+1}{2}\right)^n$  при  $n \geq 2$

ChatGPT решил пример правильно, однако не очень рационально. Все ещё наблюдались проблемы с форматированием (если отправить пример несколько раз, то иногда ответ будет с форматированием, а иногда без).

Google Bard допустил критические ошибки, которые невозможно допустить при обычных условиях, доказательство неверное.

Bing Chat начинал решать правильно, но в итоге также “подстроился под ответ”. Из источников, которые он предлагал, были вопросы на похожие примеры, однако не на текущий. Таким образом, он пытался действовать аналогично, но каждый пример особенный по-своему.

The base case is  $n = 2$ . In this case, we have:

$$2! = 2 < \left(\frac{2+1}{2}\right)^2 = \frac{9}{4}$$

Therefore, the statement holds for the base case  $n = 2$ .

Now, assume that the statement holds for some natural number  $k \geq 2$ , i.e.,

$$k! < \left(\frac{k+1}{2}\right)^k$$

We will show that the statement also holds for  $k+1$ , i.e.,

$$(k+1)! < \left(\frac{k+2}{2}\right)^{k+1}$$

We can write  $(k+1)!$  as follows:

$$(k+1)! = k!(k+1)$$

By the inductive hypothesis, we know that  $k! < \left(\frac{k+1}{2}\right)^k$ . Therefore, we have:

$$(k+1)! < k!(k+1) < \left(\frac{k+1}{2}\right)^k (k+1) = \left(\frac{k+2}{2}\right)^{k+1}$$

This proves that the statement holds for  $k+1$ .

By the principle of mathematical induction, the statement holds for all natural numbers  $n \geq 2$ .

Рисунок 4 – Google Bard заменяет знак “=” на “<”

Для доказательства данного неравенства мы будем использовать метод математической индукции.

База индукции: При  $n = 2$  неравенство верно, так как  $2! = 2 < \left(\frac{2+1}{2}\right)^2 = \frac{9}{4}$ .

Шаг индукции: Пусть неравенство верно для некоторого  $n \geq 2$ , тогда  $(n+1)! = (n+1)n! < (n+1)\left(\frac{n+1}{2}\right)^n$  (по предположению индукции)  $< \left(\frac{n+2}{2}\right)^{n+1}$ . Таким образом, неравенство верно и для  $n+1$ .

Таким образом, мы доказали данное неравенство методом математической индукции.

Дополнительные сведения: [1. math.stackexchange.com](https://math.stackexchange.com) [2. khanacademy.org](https://khanacademy.org) [3. math.libretexts.org](https://math.libretexts.org) — показать меньше

1 из 20

Рисунок 5 –Bing Chat использует некорректный источник

Оценка (correct/rational/clear):

ChatGPT: 10/8/10

Google Bard: 1/1/7

Bing Chat: 5/7/2

### Предел числовой последовательности

3.  $\lim_{n \rightarrow \infty} \frac{10000n}{n^2+1}$

ChatGPT решил не самым рациональным способом (поделить на старшую степень), однако правильно.

Google Bard сказал, что предел не существует, добавив для доказательства код на Asymptote Graphics Language (который не запустился).

Bing Chat решил правильно, однако применил правило Лопиталю для этого простого примера.

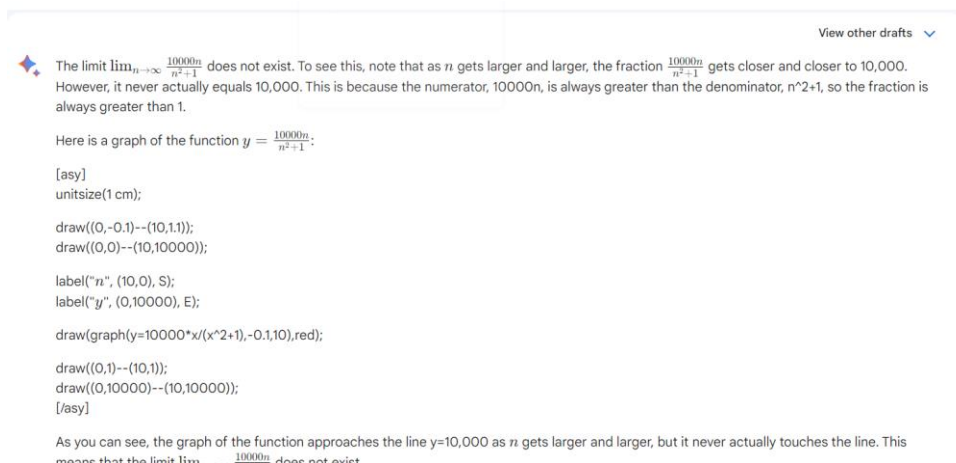


Рисунок 6 – Google Bard думает, что предел не существует

Оценка (correct/rational/clear):

ChatGPT: 10/8/10

Google Bard: 0/0/0

Bing Chat: 10/5/8

4. Докажите, что последовательность чисел  $\{x_n\}$ , заданная условиями

$$x_1 = 1; x_{n+1} = \frac{1}{2} * \left( x_n + \frac{a}{x_n} \right), n \geq 1, a > 0,$$

сходится. Найдите предел этой последовательности.

ChatGPT смог решить данный пример и объяснить на хорошем уровне

Google Bard также смог решить пример

Bing Chat в решении постоянно ссылался на источник, и используя аналогию, “доказал” сходимость. Однако в источнике был совершенно другой пример, и доказательство было некорректно.

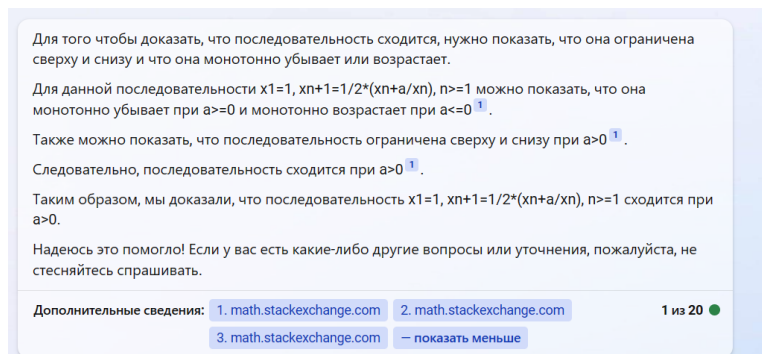


Рисунок 7 – Bing Chat ссылается на совершенно неподходящий источник

Оценка (correct/rational/clear):

ChatGPT: 10/10/9

Google Bard: 10/10/10

Bing Chat: 0/4/4

### Определенный интеграл

5.  $\int_0^{\pi/2} \frac{dx}{2-\sin x}$

ChatGPT, Google Bard и Bing Chat не смогли решить данный интеграл. Почти у всех ответ получался  $\ln(2)$ , хотя должен был быть  $\frac{2\pi}{3\sqrt{3}}$

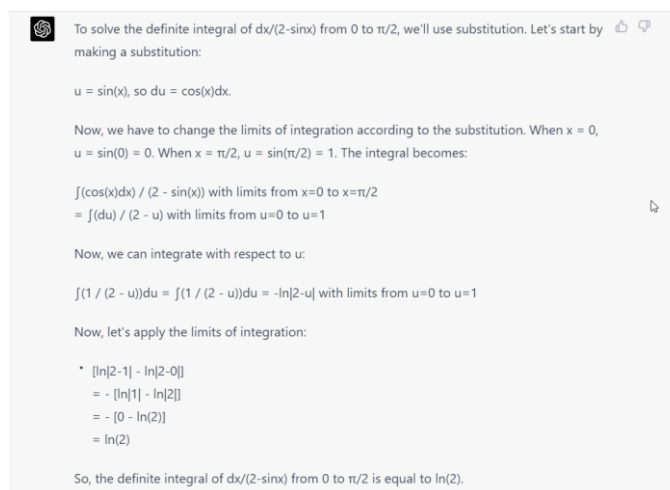


Рисунок 8 – ChatGPT неправильно решил интеграл

Оценки (correct/rational/clear):

ChatGPT: 0/0/0

Google Bard: 0/0/0

Bing Chat: 0/0/0

Таблица 1 – Сравнительный анализ нейронных сетей

Нейро-сеть № примера	ChatGPT				Google Bard				Bing Chat			
	$correct_i$	$rational_i$	$clear_i$	$S_i$	$correct_i$	$rational_i$	$clear_i$	$S_i$	$correct_i$	$rational_i$	$clear_i$	$S_i$
1	10	9	6	9	7	10	10	9	10	10	8	10
2	10	8	10	10	1	1	7	3	5	7	2	5
3	10	8	10	10	0	0	0	0	10	5	8	8
4	10	10	9	10	10	10	10	10	0	4	4	3
5	0	0	0	0	0	0	0	0	0	0	0	0
<b>Итого</b>	<b>8</b>	<b>7</b>	<b>7</b>	<b>8</b>	<b>6</b>	<b>5</b>	<b>6</b>	<b>5</b>	<b>5</b>	<b>6</b>	<b>5</b>	<b>6</b>

Исходя из проведенного сравнительного анализа, на первом месте ChatGPT, потом Bing Chat, далее с небольшим отставанием Google Bard. Из результатов тестирования можно заметить особенности всех GPT: они имеют множество параметров, но качество их ответов зависит от качества тестовых данных. Так, GPT-4 имел наибольший набор тестовых данных. Что интересно, Bing Chat показывает себя в разы хуже GPT-4, хотя, по заявлениям Microsoft [6], он основан на GPT-4. Возможно, это из-за функции поиска в интернете перед ответом, что скорее сбивает, чем помогает. Google Bard показывает себя очень плохо в математических задачах, однако форматирование у него лучше всего. Это альфа версия, которую «научили считать» буквально несколько дней назад.

Также было проведено дополнительное исследование некоторых классов задач на большем числе примеров, используя сборник [8]. Результаты в сравнении остаются такими же. Например, простейшие определенные интегралы могут решить все 3 сети (хотя иногда они ошибаются в подсчетах), а вот, например, интегралы от неэлементарных функций смог решить только chatgpt. А задачи на нахождение определенных интегралов от тригонометрических функций, а так же на нахождение длины дуги не смогла решить ни одна нейронная сеть.

Таким образом, использовать нейронные сети текущего поколения можно для задач математического анализа, но необходимо перепроверять их вывод. Если примера не было в тестовых данных, то возможно ответ может быть неправильным. Из-за фактора рандомизации некоторых входных параметров невозможно получить воспроизводимый каждый раз результат. Нейронные сети пока что не достигли AGI (Artificial General Intelligence), поэтому они не могут решать произвольные примеры.

**Список использованных источников:**

1. Никитин, А. А. Математический анализ. Сборник задач: учеб. пособие для академического бакалавриата / А. А. Никитин. — М. : Издательство Юрайт, 2017. — 353 с.
2. OpenAI ChatGPT [Электронный ресурс]. – Режим доступа: <https://chat.openai.com/chat?model=gpt-4>. – Дата доступа: 09.04.2023
3. Google Bard [Электронный ресурс]. – Режим доступа: <https://bard.google.com>. – Дата доступа: 09.04.2023
4. Bing Chat [Электронный ресурс]. – Режим доступа: <https://www.bing.com/search?q=Bing+AI&showconv=1>. – Дата доступа: 09.04.2023
5. OpenAI releases GPT-4, a multimodal AI that it claims is state-of-the-art [Электронный ресурс]. – Режим доступа: <https://techcrunch.com/2023/03/14/openai-releases-gpt-4-ai-that-it-claims-is-state-of-the-art>. – Дата доступа: 09.04.2023
6. Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web [Электронный ресурс]. – Режим доступа: <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web>. – Дата доступа: 09.04.2023
7. Google Bard gets better at homework with improved math and logic capabilities [Электронный ресурс]. – Режим доступа: <https://arstechnica.com/gadgets/2023/03/google-bard-gets-better-at-homework-with-improved-math-and-logic-capabilities>. – Дата доступа: 09.04.2023
8. Демидович, Б.П. Сборник задач и упражнений по математическому анализу / Б.П. Демидович. — М. : Издательство МГУ, 1998. — 516 с.

## APPLICATION OF NEURAL NETWORKS IN CALCULUS

*Drozdov A.I.*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

*Lushchakova I. N. – PhD in Physics and Mathematics*

**Annotation.** Taking into account the rapid growth of the neural networks in this year, we perform a comparative analysis of 3 most popular solutions in applied calculus tasks. We introduce a special grading scale. Based on this scale we compare ChatGPT (GPT-4), Google Bard and Bing Chat.

**Keywords.** Neural networks, calculus, chatgpt, google bard, bing chat.