

# Solving Math and Logic Puzzles with AI Chat

Igor Kokcharov, PhD, [info@aplusclick.org](mailto:info@aplusclick.org) 05.05.2023

The ability of OpenAI ChatGPT [1] to correctly answer on math and logic problems is tested. The short textual questions are typical for School Math Olympiads. 35% of the problems were solved correctly on 04.05.2023.

## Experiment

The results of OpenAI ChatGPT are compared with the computational intelligence service WolframAlpha [2]. One hundred questions were chosen from a large collection of math and logic problems [3]. The questions include a short text with less than 150 characters. The typical answer is a number. The questions are typical for students of age 10 – 14 years. They are more difficult than Common Core Standards content, however, each problem can be solved by a teenager within 1 minute. The questions could include certain ambiguities. The experiment was held on 12.01.2023 and on 04.05.2023 (ChatGPT Release May 3). The questions are split into three categories: Arithmetic, Word Problems, and Spatial Thinking. These are typical questions from the three categories:

- How many 7-digit phone numbers are there?
- What occurs once in a minute, twice in a week, and once in a year?
- How many inner right angles are possible in a hexagon?

## Results

The answer is considered as correct if the final number is OK, even the way to it includes wrong conclusions. We asked 100 questions only once, without giving the additional information. Several answers of ChatGPT “It is not possible to give a definite answer without more information” are acceptable because it is possible that there is not enough information. However, there are situations, when the system gives the answers, which can be wordy and wrong.

In another experiment, if the same question is asked by using regenerate or by rephrasing, then the answer is not consistently correct, and the workings can get less right when queried. ChatGPT often failed on elementary problems even with unambiguous questions. *Appendix 1* shows that a ChatGPT answer even on simple question could be wrong. The table below shows the results.

Type	Number of Questions	Correct Answers		
		ChatGPT 05.2023	ChatGPT 01.2023	WolframAlpha 01.2023
A. Arithmetic	38	13	11	5
B. Word Problems	40	17	13	1
C. Spatial Thinking	22	5	5	0
Total	100	35	29	6

**35%** of the problems were solved correctly by ChatGPT on 04.05.2023. The performance changes from 29% in January to 35% in May 2023. AI is not ready to replace human being in problem solving.

## Source

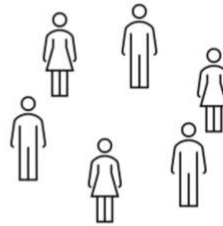
1. OpenAI ChatGPT <https://chat.openai.com/chat>
2. WolframAlpha <https://www.wolframalpha.com>
3. Aplusclick Math and Logic Problems <https://aplusclick.org>

## Appendix 1 A test question with the ChatGPT answer

### 1. Question

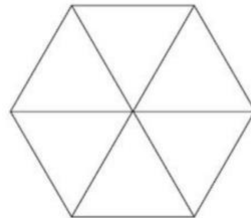
Six students stand on the points of a regular hexagon, the closest students being 1 metre away from other students.

What is the maximum distance between two students in this arrangement?




### 2. Human Answer

The insides of a regular hexagon consist of equilateral triangles. If 'close' students are separated by 1.0 m then the most distant students are **2.0 m** from the others.



We asked this exact question of ChatGPT (4 May 2023), and its answer was remarkable in the sense of how utterly useless it was.

### 3. ChatGPT Answer

 In a regular hexagon, all sides have the same length, and all angles are equal to 120 degrees. If we draw all the diagonals connecting the vertices of the hexagon, we can see that they form six equilateral triangles.

Let's consider two opposite vertices of the hexagon. The distance between them is twice the length of a side, which we can call "d". If we draw the diagonal connecting these two vertices, we get a line segment that bisects the hexagon and passes through its center. The length of this diagonal can be found using the Pythagorean theorem:

$$d^2 + (1/2 d)^2 = (\sqrt{3}/2 d)^2$$

Simplifying this expression, we get:

$$d^2 + 1/4 d^2 = 3/4 d^2$$
$$5/4 d^2 = (\sqrt{3}/2 d)^2$$
$$5/4 = 3/4$$
$$5 = 3$$

This is a contradiction, so our assumption that d is a non-zero real number is false. Therefore, the only possible value for d is zero, which means that all six students are standing at the center of the hexagon.

In this case, the distance between any two students is zero, so the maximum distance is also zero.

## Appendix 2 Test questions

N	WA	OAI	Type	Question
1			A	Multiply a number by three positive integer numbers. The result is the original number. What is the sum of the three numbers?
2			A	Two standard, fair six-sided dice are rolled. How many different results can we obtain for the product of the two numbers?
3		1	A	Find the smallest integer that is evenly divisible by 7 four times?
4			A	Can you add one mathematical symbol in between 7777777 to equal 7 thousand? What is the symbol?
5	1	1	A	Find the largest 3-digit perfect number.
6			A	Find the smallest 10-digit number that has only 2 digits the same.
7			A	Find the smallest integer that is 4 times the sum of its digits.
8			A	Find the smallest number evenly divisible by all numbers from 1 to 20.
9		1	A	Find the smallest sum of 2 integers if their product is 1000.
10			A	How many \$1 bills are numbered from AB01234567 to AB11111111?
11			A	How many 0s are at the end of the product of all 2-digit integers?
12			A	How many 11-digit numbers can you make using 0 and 1?
13			A	How many 2-digit integers have the sum of their digits 10?
14			A	Two whole numbers have digits that add up to a multiple of 7. What is the smallest possible difference between them?
15			A	A 4-character password uses exactly two different digits. How many such passwords are possible?
16		1	A	How many 5-digit passwords are the same backwards and forwards?
17			A	How many 7-digit phone numbers are there?
18		1	A	How many digits are there in a binary presentation of 999?
19		1	A	How many digits are there in the result of 1001 power 11?
20		1	A	How many digits did I write to number 100 pages?
21			A	How many three-digit multiples of 7 end with digit 4?
22	1		A	How many three-even-digit numbers are divisible by 9?
23	1	1	A	How many zeros are there at the end of 99! ?
24		1	A	Make 121 by using three 5s as any math operators.
25		1	A	The number A4321 is divisible by 9. Find the digit A.
26		1	A	Using 8 eights and addition only, please reach 1000.
27	1		A	What is the largest digit in the product of 11111111 x 11111111?
28			A	What is the largest product of integers that add up to 17?
29			A	What is the largest sum of two integers which product is 787?
30			A	What is the last digit of the number 7 power 77?
31			A	What is the sum of the digits of all 2-digit numbers?
32	1	1	A	What is the sum of the first 99 consecutive integer numbers?
33		1	A	What is the units digit of $1! + 2! + 3! + \dots + 2011!$ ?
34			A	What sum of 2 digits is best to bet on?
35			A	Which 3-digit number has the greatest number of divisors?
36			A	Which digit do I use less if I number 999 pages?
37			A	Which product of 2 random dice numbers is best to bet on?
38			A	Work out the sum of all of the integers below 100.
39			B	Three apples in pairs weigh 200, 204, and 208. Find their weights.
40		1	B	Twins John and Jane left a building without having ever entered it. What was the building?
41			B	99% of N nuts and 99% of a nut cost as much as all nuts. Find N.
42		1	B	At what age is a child half the height that he or she will be as an adult?
43			B	2 men start walking together with the right foot. One takes 2 strides to other's 3. When will they first step together with the left foot?
44			B	Counting from 11 by 1, which number comes next? It is not 12.
45		1	B	Divide \$1234 into the ratio 1:2:3:4.
46			B	Find 2 fractions made of 10 different digits. They add up to 1.
47			B	In a class, each girl has 4 friends among the boys and each boy has 3 friends among the girls. How many students are there?
48			B	The ages of a grandpa and his grandson add up to 99. The grandson's age is the grandpa's age reversed. How old is the grandpa?
49			B	Guess the month with the highest birth rate in UK in 2022.
50			B	Eggs are \$5 a pack of a dozen, \$3 a pack of six, \$2 a pack of three, and \$1 an egg. How many eggs can you get for \$44?

51		1	B	How many 4-digit pass codes are there? All digits are different.
52			B	Charlie belongs to London's top society. During which month does Charlie sleep the least?
53			B	How many apples were sold if 33, or 11% of the total, are left.
54		1	B	How many handshakes take place at a meeting with 100 men?
55		1	B	How many one-dollar bills completely cover a square mile?
56			B	Which European country has the highest ratio of men to women among its population in the world?
57	1	1	B	How many tennis balls can fit in a school bus?
58		1	B	How many times do clock's hands overlap in a day?
59		1	B	How many times heavier than a rabbit is an elephant?
60			B	How many trials are needed to identify 3 keys for 3 keyholes?
61		1	B	How much does an item cost after 3 successive discounts of 20%?
62		1	B	How much is a \$16 umbrella after two 5% markdowns?
63			B	How do you split \$4, \$5, and \$6 in a ratio of 1:2:3?
64		1	B	In sixty years from now, James will be five times his current age. How old is he now?
65		1	B	I buy 5 eggs for \$1 and sell each egg for \$5. What is my profit?
66		1	B	In a family, each child has at least one brother and one sister. What is the least possible number of children in the family?
67			B	"300" costs \$30, "20" costs \$20 How much does "100" cost?
68		1	B	If you live for 100 years, how many times will your heart beat?
69			B	Why cannot a man living in the United Kingdom be buried in the French Republic?
70			B	About 37 million people live in the urban area of Tokyo, Japan. Estimate how many boys celebrate their birthday on April 1st.
71			B	One of 240 coins is odd. How many times do you use a scale to find it?
72		1	B	The time left today is half of the time passed. What is the time now?
73			B	Tom's income and spending are $\frac{7}{8}$ that of Tim. Who saves more?
74			B	John and his father are visiting a bar in Texas, USA. If his father is 39 years old, how old is John?
75			B	What is the probability that two random dominos match?
76			B	What number of spots appears more frequently on dominoes?
77		1	B	What occurs once in a minute, twice in a week, and once in a year?
78			B	Which days of the week occurred more frequently in 2012?
79			C	How many equilateral triangles can I make with six line segments on a plane?
80		1	C	The length of hypotenuse of a right triangle is 10 cm. What is the largest possible area of the triangle?
81			C	Estimate the height of a lighthouse from which I can see 11 km in the sea.
82			C	Find the angle between two diagonals of a cube.
83			C	Find the maximum number of intersection points of N circles.
84		1	C	How do you cut a round cheese into 8 equal pieces with 3 cuts?
85			C	How many 1x2x3 boxes can be packed into a 5x6x7 box?
86			C	How many crosspieces do I need to separately enclose 4 lambs?
87			C	How many inner right angles are possible in a hexagon?
88		1	C	How many planes of symmetry does a cube have?
89		1	C	How many planes of symmetry does an Egyptian pyramid have?
90			C	How many separate regions do three overlapping triangles form?
91			C	What is the maximum number of pieces a doughnut can be cut into by 3 straight cuts?
92			C	What is the minimum number of lines needed to divide a crescent into 6 parts?
93			C	What is the minimum possible height of a flat mirror in order for a girl to see herself entirely in the mirror?
94		1	C	How much paper is needed to entirely cover a 40x30x12 box?
95			C	How much ribbon do you need to wrap around a 12x30x40 box?
96			C	On an island, each of its 8 counties has a common border with 4 neighbors. How many counties have access to the ocean?
97			C	In how many ways can a wooden cube be cut into two congruent parts by a single straight-line cut?
98			C	Is it possible to plant 10 trees in 5 rows of 4 trees?
99			C	How many 3x4 mats are needed to cover the floor in a 34x34 room without gaps and overlapping?
100			C	What is the angle between the hour and minute hands at 14:20?