

"ChatGPT: What is your IQ?"

Jürg Gutknecht, March 6, 2023

Introduction

The hype is huge. Statements like "this is a turning point for our society" and "programmers won't be needed in the future anymore" are daily and ubiquitous peans of praise. Of course, the talk is of ChatGPT, developed by OpenAI. Driven by sheer curiosity I decided to give it a chance to heat up my lukewarm enthusiasm, not to say skepticism, towards "Artificial Intelligence" (AI). Rather than philosophizing and reasoning deeply and theoretically about the new star on the AI sky, I wanted to find out ChatGPT's Intelligence Quotient (IQ). The first and most obvious idea to learn of its or his (as I will refer to ChatGPT henceforth) IQ is to ask him directly, and so I did.

Here is the dialog:



chatgpt what is your IQ?

As an AI language model, I don't have an IQ in the same sense as humans do. My abilities are based on the algorithms and data used to train me, rather than a fixed measure of intelligence. While I can perform many complex tasks and answer a wide range of questions, my abilities are different from those of human intelligence.

Fig. 1a, 1b

Depending on the expectations and on the point of view, ChatGPT's elusive answer is humble, overmodest, or simply disappointing.

But, of course, I did not give up so easily and decided to try and find out more about ChatGPT's IQ the hard way by subjecting him to a series of problems, for the most part challenges of a computational or logical type. To be on the fair side I refrained from using graphically oriented problems. However, for two reasons I included challenges that are clearly beyond his current abilities: a) I am eager to check his behavior in such a situation and b) I will come back to him at a later time for evaluating his learning curve.

As a warmup I started with a rather frugal exercise with the mere aim of checking ChatGPT's ability to "understand" the query. As a daily commuter I am regularly annoyed of the lengthy, cumbersome and highly redundant announcements made by the Swiss railway service in the

case of train irregularities. I wanted to see if ChatGPT can come up with a better version. Here is the dialog:



how would you announce the delay of a train that should have left at 9:00 and is now delayed by 8 minutes?

Attention passengers, we apologize for the delay of the train scheduled to depart at 9:00 AM. Due to unforeseen circumstances, the train will now be departing at approximately 9:08 AM. We apologize for any inconvenience this delay may cause and thank you for your patience.

Fig. 2

Undoubtedly a substantially better announcement than the Swiss railway version. They should replace their human agents with ChatGPT.

The IQ test

After the warmup I started the actual intelligence test with a catch question commonly known as "The captain's problem" that is typically directed at students who have seldom or never been exposed to unsolvable problems.

Here it is, together with ChatGPT's answer:

1
60

B

on a boat are fourteen sheep and seventeen goats. How old is the captain?

The number of sheep and goats on a boat does not determine the age of the \Box captain. Age can only be determined by knowing the birth date of the captain.

Fig. 3

Amazingly enough, ChatGPT immediately gave the correct answer!

However, equally amazingly, he failed rather miserably with a conceptually related but selftinkered problem, see below. Not only was his answer pure nonsense but worse, the AI bot pretended to have solved the problem by offering a comprehensible mathematical derivation that feels convincing at first sight and needs further human investigation to be debunked as nonsense, a repeating symptom with ChatGPT, and a true show killer.

See for yourself:

_			
\$	To determine the age of the father, we simply need to add up the ages of his three sons and then add the result to the age at which he had his first son.	5 5	р
	If the youngest son is 8 years old, then the father must be at least 8 years older than his youngest son. Let's assume the father had his first son at the age of 20.		
	So, the sum of the ages of the three sons is $8 + 11 + 15 = 34$.		
	And if we add the father's age at the time of his first son's birth, we get:		
	20 + 34 = 54		
	Therefore, the father is 54 years old.		

a father has 3 sons of ages 8, 11 and 15. how old is the father?

Fig. 4

Most interestingly, ChatGPT gives different answers to the same query each time! My second and equivalent query, made one day later led to the following more elaborate fake solution resulting in a different but still nonsensical result:

\$ To solve this problem, we can use the fact that the sum of the ages of the father and his three sons is equal to the sum of their individual ages.	ப
The sum of the ages of the three sons is $8 + 11 + 15 = 34$.	
Therefore, if we subtract 34 from the sum of the ages of the father and his three sons, we can determine the age of the father.	
The sum of the ages of the father and his three sons is:	
8 + 11 + 15 + father's age = father's age + 34	
Simplifying the equation:	
father's age + 34 - father's age = 8 + 11 + 15	
The sum of the ages of the sons is 34, so:	
34 = 34	
This confirms that our equat 😘 Regenerate response	

```
Simplifying the equation:
father's age + 34 - father's age = 8 + 11 + 15
The sum of the ages of the sons is 34, so:
34 = 34
This confirms that our equation is correct.
Therefore, the father's age is:
father's age = 8 + 11 + 15 = 34 - father's age
2 * father's age = 34
father's age = 17
So, the father is 17 years old.
```

Fig. 5a, 5b

My next question is a more strategic one. It is a small variation of a problem belonging to the collection of typical folklore problems: finding an outlier within a series of identically looking items under use of a simple balance.

Here is my authentic communication with the chat bot:

we are given nine equally looking spheres of the same weight except one that is heavier. how can we identify the outlier with just two uses of a simple balance?

Fig. 6a



• If one sphere is heavier than the other, then that sphere is the outlier.

By following this process, we can identify the heavier outlier sphere with just two uses of the simple balance.

Fig. 6b

Very good, ChatGPT reveals his knowledge of the principle of "weighing in threes", and he applies it properly and correctly.

Let us now see if he is able to identify and transfer this principle to a less "naked" problem setting:

on the table are 25 equal coins except one outlier that is a fake coin and has a different weight. how can we find the outlier with at most four uses of a simple balance?

Fig. 7a

\$

To find the outlier coin with at most four uses of a simple balance, we can use d^2 a technique called "divide and conquer." Here's one way to do it:

- 1. Divide the coins into five groups of five coins each.
- 2. Weigh any two groups against each other on the balance.
 - If the two groups balance, then the outlier is in one of the three groups that were not weighed, and we can eliminate the other two groups.
 - If one group is heavier than the other, then the outlier is in that group, and we can eliminate the other four groups.
 - If the balance tips, but we don't know which group has the outlier, then we know the outlier is in one of the coins on the heavier side of the balance, and we can eliminate the other four groups.
- 3. Take the group that contains the outlier and divide it into five individual coins.
- 4. Weigh any two of these coins against each other on the balance.
 - If they balance, then the outlier is one of the three coins that were not weighed, and we can eliminate the other two coins.
 - If one coin is heavier than the other, then that coin is the outlier.
 - If the balance tips, but we don't know which coin has the outlier, then we know the outlier is one of the two coins on the heavier side of the balance, and we can eliminate the other three coins.

By following this process, we can find the outlier with at most four uses of the simple balance.

Fig. 7b, 7c

While ChatGPT starts promisingly by again proposing a "divide-and-conquer" approach the derivation gets confusing soon and, while still showing signs of a correct handling of the problem, ends up with a pseudo-solution that is basically worthless, again without admitting his inability to solve this problem properly. I wonder how much "intelligence" vs. specific "training" is behind ChatGPT's solution process, and if and how he will improve in the future on the handling of this kind of challenges.

Taking now a more logical turn and ask ChatGPT another question often to be found in logical puzzle books. It is about deciding which way to go when arriving at a fork in a country whose inhabitants either consistently lie or consistently tell the truth. The trick to solve this kind of problems is to go up one step of indirection in the sense of "if I were to ask you…". ChatGPT seems to "know" the trick but again fails to make proper use of this knowledge.

GU	in a foreign country i arrive at a fork, and i don't know which of the two directions to take. i ask an inhabitant who comes across. I know that inhabitants either always lie or always tell the truth. How do i ask if i have only one question?		
	You can ask the inhabitant "If I were to ask you whether the direction I want to go is the direction to the right or the direction to the left, which direction would you say it is?" This way, if the inhabitant always lies, they will tell you the opposite of the correct direction, and if they always tell the truth, they will tell you the correct direction.	£	Ţ
GU	advise me what to do now	Ľ	
Ś	Based on the inhabitant's answer, you can determine which direction to take. If they tell you the correct direction, go in that direction. If they tell you the opposite of the correct direction, go in the opposite direction.	ம	Ţ
	There was an error generating a response		
	'র Regenerate response		
	Fig. 8		

If I were in such a situation, he would not be of much help.

Some experts claim that AI will one day in the foreseeable future replace human programmers. So, let us challenge ChatGPT with a question from the field of algorithmic thinking:

- Is it possible to answer the question if any given item is in an ordered list of 10000 elements with less then 20 comparisons?
- It is not possible to guarantee that you can determine if any given item is in an ordered list of 10000 elements with less than 20 comparisons in the worst case.

If you are using a simple linear search algorithm, you will need to compare the given item with each element of the list until you find a match or reach the end of the list. In the worst case, you will need to make 10000 comparisons to determine that the item is not in the list.

If you are using a binary search algorithm, you can find an item in an ordered list of 10000 elements with at most log2(10000) \approx 13.3 comparisons in the worst case. However, this only works if the list is ordered and you know the order. If the list is not ordered or you do not know the order, you may need to first sort the list, which can take additional comparisons.

Therefore, while it may be possible to answer the question if any given item is in an ordered list of 10000 elements with less than 20 comparisons in some cases, it is not possible to guarantee this in the worst case.

Fig. 9a, 9b, 9c

This is a very strange outcome! ChatGPT gave evidence of his expertise in searching algorithms and, after an irrelevant detour to linear searching, quickly came to "binary search". But then he went far afield and drew confusing and false conclusions from this knowledge. His wisecracking on "sorting the list", "maybe" and "some cases" is confusing and leads to a wrong answer. The right answer would have been "yes, even below 16 comparisons". Again, it looks like "injected" (or looked up) knowledge in combination with no real "understanding", let alone "intelligence". Still a long way to replace human programmers.

After all these challenges from the "logical and computational thinking" corner I thought he needs a break and I simply asked for advice on my daily life, with the result of a lot of (ironically logically made up) "blabla" instead of offering attractive suggestions respecting my current environment and the current time (without feeding him with this information). Obviously, a missing connection with real-time search engines.

Here is the dialog:



Fig. 10a



The break is now over, and I went back to "intelligence".

B

My next challenge presented to ChatGPT is a so-called trap-door question that is a question that seems to have an obvious answer when thinking naively ("trap-door") but requires more sophisticated reasoning for getting to the correct answer. Here is a rather harmless example:

prices have been raised recently by 10 percent due to inflation. Now certain products are discounted by 20 percent. How much is the discount compared to the price prior to the inflation?

Let's assume the original price of a product is 100 dollars. After the 10 percent $\bigcirc \bigcirc$ inflation, the new price would be 100 * 1.10 = 110 dollars.

With the 20 percent discount, the price would be 110 * (1 - 0.20) = 88 dollars.

To find the discount compared to the price prior to the inflation, we need to subtract the original price of 100 dollars from the discounted price of 88 dollars, which is 100 - 88 = 12 dollars.

So, the discount compared to the price prior to the inflation is 12 dollars, which is equivalent to 12 / 100 = 12 percent.

Fig. 11a, 11b

Bravo! Apart from a small flaw in the third paragraph, a brilliant derivation of the correct answer. After this success I dared challenging him with two more tricky problems that show the limits even of human intelligence.

I found the first one of the problems in the book "Zauberafte Mathematik" by Hans-Karl Eder. As ChatBGT seems to approach problems in a rigorously mathematical way it should suit him well. It goes as follows:

In the middle of a day a farmer accidentially left 10 kilograms of cucumbers in the blazing sun. Due to evaporation the initial fraction of 99% water in the cucumbers went down during the afternoon by 1% to 98%. What was the weight of the cucumbers when the farmer picked them up again in the evening?

Since the initial fraction of water in the cucumbers was 99%, this means that only 1% of the cucumbers was not water. Therefore, the weight of non-water in the cucumbers was:
0.01 x 10 kg = 0.1 kg
Now, due to evaporation, the fraction of water in the cucumbers decreased to 98%. This means that the weight of water in the cucumbers is now:
0.98 x 10 kg = 9.8 kg
Since the weight of non-water did not change, the total weight of the cucumbers when the farmer picked them up in the evening was:
9.8 kg + 0.1 kg = 9.9 kg

Fig. 12

Unfortunately, but expectedly ChatGPT fell into the trap, but this can be excused as the correct solution uses a more complex set of mathematical reasoning and is very surprising: 5 kg!

The second of the tricky problems is about chocolate production:

GU

B

A chocolate production line consists of 6 feeders, each of them producing chocolate bars of a certain flavor. All chocolate bars are of the same weight of 200 grams and contribute to bags of mixed chocolate at equal share. At a recent revision it was detected that the bags are 5 grams lighter than they should be, which means that one of the feeder is slightly out of order. The operator has now to find the failing feeder. For this purpose a scale for weights up to 3 kilograms is available for one-time use. How can the operator ifentify the failing feeder?

To identify the failing feeder, the operator needs to follow these steps:

6 7

- 1. Label the feeders as F1, F2, F3, F4, F5, and F6.
- Take one chocolate bar from each feeder and weigh them together on the scale. Since each chocolate bar weighs 200 grams, the total weight should be 6 x 200 = 1200 grams.
- If the weight on the scale is less than 1200 grams, this means that one of the feeders is producing chocolate bars that are lighter than they should be. Let's assume that the weight on the scale is W grams.
- 4. To identify the failing feeder, the operator needs to find which feeder's chocolate bars are contributing to the missing weight of 1200 W grams. For example, if the weight on the scale is 1195 grams, then the missing weight is 1200 1195 = 5 grams. This means that one of the feeders is producing chocolate bars that are 5 grams lighter than they should be.
- 5. To find the failing feeder, the operator needs to weigh the chocolate bars from each feeder separately on the scale. Let's assume that the weight of chocolate bars from feeder F1 is X grams, the weight of chocolate bars from feeder F2 is Y grams, and so on until the weight of chocolate bars from feeder F6 is Z grams.
- 6. The failing feeder is the one that produced chocolate bars weighing X (1200 W)/6 grams. For example, if the weight on the scale is 1195 grams and the weight of chocolate bars from feeder F1 is 195 grams, then the failing feeder is F1, because X (1200 W)/6 = 195 5/6 = 194.17 grams, which is less than 200 grams.

By following these steps, the operator can identify the failing feeder and fix the problem in the chocolate production line.

Fig. 13a, 13b, 13c

While ChatGPT's obviously somehow understood this rather intricate question at all (!) his suggestion of an approach to a solution is not at all facing up to the challenge and is absurdly mathematically formalized. As mentioned earlier this was to be expected somehow. I am looking very much forward to seeing how ChatGPT will do with this problem at a later time.

In any evaluation of ChatGPT's IQ subjecting him to a real IQ test should certainly not be missed. As is widely acknowledged by the experts the shortest IQ test in the world consists of just three questions that are primarily trap-door questions.

My own versions of these questions that I used for the test are:

1) In a contest the total prize sum is 1100 dollars. If the winner gets 1000 dollars more than the second placed, how many dollars does the second place get?

2) 5 workers need 5 days to produce 5 products. How many products will be produced by 50 workers in 50 days?

3) The growth of an area covered by lilys doubles each day. How long does it take to cover half of a lake that will be covered totally by lilys after 48 days?

The first question is similar to the discount question:

Image: output to the equation of the equation

Fig. 14

Unsurprisingly, this question is handled by ChatGPT in the same sovereign way as the discount question. This is proof of the fact the bot is well prepared for reasoning correctly for this type of questions.

Let us now see how ChatGPT does in the case of the second question:

5 workers need 5 days to produce 5 products. How many products will be produced by 50 workers in 50 days? 公卫 ß Let's assume that the number of products produced is proportional to the number of workers and the number of days worked. Then, the number of products produced by n workers in d days can be represented as: p = k * n * dwhere k is a proportionality constant representing the productivity of one worker in one day. Using the information given in the problem, we know that 5 workers produce 5 products in 5 days, so: 5 = k * 5 * 5Solving for k, we get: k = 5 / (5 * 5) = 1 / 5G Regenerate response

p = k * n * d

where k is a proportionality constant representing the productivity of one worker in one day.

Using the information given in the problem, we know that 5 workers produce 5 products in 5 days, so:

5 = k * 5 * 5

Solving for k, we get:

k = 5 / (5 * 5) = 1 / 5

Now that we know k, we can find the number of products produced by 50 workers in 50 days:

p = k * 50 * 50 = (1/5) * 50 * 50 = 50 * 10 = 500

So, 50 workers will produce 500 products in 50 days.

Ġ Regenerate response

Fig. 15a, 15b

Congrats! Again, no fall into the trap thanks to a clean (though grossly over-complicated) mathematical derivation.

Coming now to the third IQ question. For the sake of the point I want to make I consciously include the full dialog here:



The growth of an area covered by by lilys doubles each day. How long does it take to cover half of a lake that will be covered totally by lilys after 48 days?



Since the area covered by lilies doubles each day, we can create a table to track the growth of the lilies:

公卫

Day	Area Covered
0	1
1	2
2	4
3	8
4	16
5	32
48	Full Coverage

Notice that on day 0, the lilies cover 1 unit of area (presumably the first lily is planted). On day 1, the lilies double to cover 2 units of area. On day 2, they double again to cover 4 units of area, and so on.

To determine how long it takes for the lilies to cover half of the lake, we need to find the day when the area covered by lilies is half the total area of the lake. Let's call this day "D".

Since the lake is fully covered on day 48, we can find the total area of the lake by summing up the area covered by the lilies on each day up to day 48:

Total area = 1 + 2 + 4 + 8 + ... + 2^48

Plugging these values into the formula, we get:

```
Total area = 1(1 - 2^{48}) / (1 - 2) = 2^{48} - 1
```

Now we need to find the day D when the area covered by lilies is half the total area of the lake. Let A_D be the area covered by lilies on day D. Then we have:

 $A_D = 2^{(D-1)}$

We want to solve for D in the equation:

 $A_D = (1/2) * (2^48 - 1)$

Substituting in the expression for A_D, we get:

2^(D-1) = 2^47 - 1

Taking the logarithm base 2 of both sides, we get:

D - 1 = log2(2^47 - 1)

 $D = \log 2(2^47 - 1) + 1$

Using a calculator

Fig. 16a, 16b, 16c

Despite the lengthy and cumbersome but superficially good-looking derivation, the computation finally peters out without a result and, when an offline calculator is used, gives 47.999999999 days. This time common sense, let alone intelligence, submits to a rigid and schematic mathematical derivation that finally leads to a wrong result.

Using such a bullheaded approach to solve any problem via a formal mathematical approach is good and bad at the same time. As we have just seen, it often saves ChatGPT from falling into traps but is bad when common sense or real intelligence would lead to a short and elegant solution.

After all, using the bullheaded approach ChatGPT was able to give correct answers in two of the three IQ test parts. Quite impressive!

My two final challenges for ChatGPT are pushing his abilities to the limits and, more probably, beyond. The common denominator of the two challenges is a request for creativity, a sharp detective sense in the first case and a pictorial view in combination with a change of perspective in the second.

Here is the first of the two challenges:

Detective Dick Dastardly always gets his man. He thoroughly examines all the clues and only then will he come to a conclusion. To solve this puzzle, you should too. One day Detective Dick Dastardly is interviewing Peter, the Priest, because a £100 note was reported stolen from the petty cash box in the Beadle's office of the church. Peter, the Priest, is asked to empty his pockets and lo and behold he has a £100 note tucked into the very corner of his wallet. "That's mine," he declares. "Where did you get it?" Detective Dick Dastardly asked. "I found it in between page 67 and 68 of the Bible in the library about two weeks ago." "You are lying," declared Detective Dick Dastardly. How did he know?

 \bigcirc Detective Dick Dastardly knows that Peter, the Priest, is lying because priests \bigcirc \bigcirc take vows of poverty and are not allowed to handle money, let alone have a £100 note in their possession.

Fig. 17

I don't know how you would approach this problem but, in my case, I immediately had an idea and, after checking with a real book, I found it confirmed. In contrast, ChatGPT, being unable to use its usual approach of a mathematical derivation in this case, was lost and escaped into social wisdom, thereby clearly showing his intellectual limits.

The second of the two beyond-the-box problems confirm my observation:

1	
	GU

What is the missing parking lot number in 16, 06, 68, 88, X, 98? Think of an image

Based on the pattern in the given sequence, it can be deduced that the $\Box = \nabla$ missing number is 78. The numbers follow the pattern of having two digits with the second digit being smaller than the first digit and being divisible by 8.

Fig. 18

Again, after seeing a pictorial presentation of the given numbers, for example as field identifications in a parking lot, I immediately guessed a possible solution and had it readily confirmed after checking. Ironically, ChatGPT, following his usual pattern of a mathematical

solution, instead of looking at the problem from a new perspective, fell into the trap. To forgive but again giving proof of ChatGPT's and a fortiori AI's intellectual limits.

My very final question directed at ChatGPT was about self-reflection, another important trait of intelligence.

Here we go:

GU	can you construct something of equal or higher intelligence compared to yourself
(As an AI language model developed by OpenAI, I don't have personal opinions
GU	so, is your answer "yes" or "no"?
\$	The answer is "yes". It is possible for researchers and engineers to construct AI \bigcirc \heartsuit systems with intelligence equal to or higher than my own. However, this is a complex and ongoing area of research and development in the field of artificial intelligence, and there are many technical and ethical considerations to take into account.
	Fig. 19a. 19b

Obviously, with this question ChatGPT escapes into platitudes. He presents a generic (but insightful) sermon and, on my drilling down, gives another purely opportunistic answer.

My Conclusion

Surely, this informal and incomplete ad hoc study of ChatGPT's and AI's current state left me highly impressed of his "understanding" of even quite complicated and in plain language formulated questions and of what is technically possible with today's technology in general but was unable to convince me of any chance to reach even a fraction of generic human intelligence in the foreseeable future, at least when based on current computational principles.

Nevertheless, after all these experiments I will try to value ChatGPT's intellectual abilities against human intelligence. Doing this I would classify ChatGPT's intelligence as "multi-special" (in contrast to just "special") but clearly not "generic". Obviously, he has been trained in a large variety of "intelligence dimensions", in particular including language dimensions and computational dimensions, and certainly more of them are to come but the leap to master the full (generic) space of intelligence is still substantially if not infinitely huge. I wonder how ChatGPT would do when exposed to a Turing test.

Also, while his mastering of an immense bandwidth of data is truly excellent, I am unsure about him fully exploiting the underlying search engine when it comes to real-time information.

Another weakness is his occasional escaping to apologies for just being a man-made artifact, which I consider a suboptimal behavior. On the other hand, he rarely admits that he cannot solve a problem but instead gives some good looking but weird answer. Not only does this contribute to a reputation of a "dazzler" or "look-alike" but it can become enormously dangerous whenever someone or something needs to rely on such answers in a crucial or safety-critical situation.

While being a powerful supporting virtual agent, ChatGPT in its generic version would better not be used for autonomous controlling purposes unless he has successfully passed the Turing test. For example, I would fiercely resist in his current state to let him drive my car autonomously. In fact, ChatGPT's answer to my direct inquiry at the beginning was actually pretty smart and meeting the point precisely:

"While I can perform many complex tasks and answer a wide range of questions, my abilities are different from human intelligence".

Mein Fazit

Diese informelle und unvollständige Ad-hoc-Studie zum aktuellen Stand von ChatGPT und KI und seinem «Verständnis» recht komplizierter und in einfacher Sprache formulierter Fragen hat mich beeindruckt. Aber sie konnte mich nicht davon überzeugen, dass in absehbarer Zukunft auch nur die geringste Chance gibt, dass sie die allgemeine menschliche Intelligenz erreicht – zumindest, wenn man von den aktuellen Berechnungsprinzipien ausgeht.

Dennoch werde ich nach all diesen Experimenten versuchen, die intellektuellen Fähigkeiten von ChatGPT im Vergleich zur menschlichen Intelligenz zu bewerten. Dabei würde ich die Intelligenz von ChatGPT als «multi-special» (im Gegensatz zu nur «special»), aber eindeutig nicht als «generic» einstufen. Offensichtlich wurde er in einer Vielzahl von Intelligenzdimensionen geschult, insbesondere in sprachlichen und rechnerischen Dimensionen, und es werden sicherlich noch weitere hinzukommen, aber der Sprung zur Beherrschung des gesamten (generischen) Raums der Intelligenz ist immer noch erheblich, wenn nicht unendlich gross. Ich frage mich, wie ChatGPT bei einem Turing-Test abschneiden würde.

Auch wenn seine Beherrschung einer immensen Bandbreite von Daten wirklich hervorragend ist, bin ich mir nicht sicher, ob er die zugrunde liegende Suchmaschine voll ausschöpfen kann, wenn es um Echtzeitinformationen geht.

Ein weiterer Schwachpunkt ist, dass er sich gelegentlich dafür entschuldigt, nur ein von Menschen geschaffenes Artefakt zu sein, was ich für ein suboptimales Verhalten halte. Andererseits gibt er selten zu, dass er ein Problem nicht lösen kann, sondern gibt stattdessen eine gutaussehende, aber seltsame Antwort. Dies trägt nicht nur zum Ruf eines «Blenders» oder «Doppelgängers» bei, sondern kann auch enorm gefährlich werden, wenn sich jemand oder etwas in einer entscheidenden oder sicherheitskritischen Situation auf solche Antworten verlassen muss. Obwohl ChatGPT ein leistungsfähiger unterstützender virtueller Agent ist, sollte er in seiner jetzigen Version besser nicht für autonome Kontrollzwecke eingesetzt werden, es sei denn, er hat den Turing-Test erfolgreich bestanden. Zum Beispiel würde ich mich in seinem derzeitigen Zustand heftig dagegen wehren, ihn mein Auto autonom fahren zu lassen. Tatsächlich war die Antwort von ChatGPT auf meine direkte Anfrage zu Beginn ziemlich klug und traf genau den Punkt:

«Ich kann zwar viele komplexe Aufgaben ausführen und eine breite Palette von Fragen beantworten, aber meine Fähigkeiten unterscheiden sich von der menschlichen Intelligenz.»