

Large Language Models and Mathematical Understanding

Jelle Couperus
5668891

Supervisor: Dr. Johannes Korbmacher



Master thesis of Artificial Intelligence
Utrecht University

2023-07-13

Contents

1	Introduction	1
2	Mathematical understanding	3
2.1	Levels of mathematical understanding	3
2.2	Aspects of understanding	5
2.2.1	Contested properties	5
2.2.2	Explanation and understanding	6
2.2.3	Intuition and Understanding	7
2.2.4	Common properties	9
2.3	Measuring understanding	10
2.3.1	Questionnaire	10
2.3.2	Mathematical assessment scale	11
2.3.3	Computer-assisted scientific understanding	12
2.3.4	Criticisms measurements	13
2.3.5	Conclusions measurements	15
2.4	Understanding and computers	16
3	Historical usage of computers in mathematical inquiry	19
3.1	Computer-assisted proofs	20
3.1.1	Four Color theorem	21
3.1.2	Double Bubble theorem	24
3.1.3	Connect Four	25
3.2	Falsifying conjectures	26
3.2.1	Chebyshev's Bias	26
3.2.2	Euler's sum of powers	27
3.3	Proof checking and finding proofs	28
3.4	Computers and understanding	30
4	AI in Mathematical Inquiry	33
4.1	What are Large Language Models?	33
4.1.1	GPT	36
4.1.2	LLMs and understanding	42

5	Experiment	44
5.1	Experiment setup	44
5.1.1	Participants	44
5.1.2	Question design	45
5.1.3	Evaluation metrics	46
5.2	Experiment results	52
5.2.1	Assessment ChatGPT	52
5.2.2	Evaluation students	55
5.3	Experiment discussion	58
5.3.1	Randomness ChatGPT	58
5.3.2	Design choices	61
5.3.3	Future experiments	63
5.4	Conclusions experiment	64
6	Conclusion	66
A	Appendix	68
A.1	Grotere Windmolens, version A	68
A.2	Grotere Windmolens, version B	70
A.3	Extra vragen	70
A.4	Consent Form	71
A.5	Instructions teachers	72
A.6	ChatGPT Answer version A Grotere Windmolens, original Dutch	73
A.7	ChatGPT Answer version A Grotere Windmolens, translated to English	74
A.8	ChatGPT Answer version B Grotere Windmolens, original Dutch	75
A.9	ChatGPT Answer question 1 Extra vragen original Dutch	76
A.10	ChatGPT Answer question 1 Extra vragen translated to English	76
A.11	ChatGPT Answer question 2 Extra vragen original Dutch	77
A.12	ChatGPT Answer question 2 Extra vragen translated to English	77
A.13	Example ChatGPT question 2A, temperature 0	78
A.14	Example ChatGPT question 2A, temperature 0, top_p=0.1, top_k=1	80
A.15	Example ChatGPT Grotere Windmolens, temperature 0	82
A.16	Student answers	92

Abstract

As computers have become more advanced, their role in mathematics has grown. Besides simple functions such as plotting graphs and doing calculations, computers have been used in increasingly complex roles, such as proof assistance. Artificial intelligence algorithms are the most recent and complex way that computers have been used for mathematical inquiry. As the role of computers in mathematics has grown, several challenges have been raised. One of the challenges that AI faces is whether it can have mathematical understanding. This thesis aims to determine if AI can have mathematical understanding. In order to answer this question, the thesis will focus on Large Language models, whether they can have mathematical understanding and how this understanding could be measured.

Chapter 1

Introduction

Mathematics plays an important role in many aspects of every day life. Much of the math done nowadays is done with the assistance of computers. For many tasks the role of the computer is small and not entirely necessary, but using a computer is more convenient than doing everything by hand. However, computers are also used for more difficult tasks in mathematics, such as proof assistance. For example, there are theorems that have been proven with the help of computers that to this day cannot be proven by hand [2]. Although computer algorithms have been used to perform more difficult mathematical tasks, these traditional algorithms could only perform linear, deterministic tasks. AI algorithms, on the other hand, are designed to learn and do complex tasks.

One of the advancements in AI that seems promising for mathematics is the development of Large Language Models (LLMs). These models have shown good results doing text-based tasks in several fields. In the case of mathematics, LLMs have been used to answer mathematical questions and to generate mathematical questions [14]. When computers were first used in mathematical proofs, several challenges were raised. With LLMs showing potential to become more prevalent in mathematics, more challenges might be raised. One of these challenges is whether AI can have mathematical understanding. If AI algorithms, such as LLMs, were to have mathematical understanding, they could potentially be used to prove theorems that mathematicians have failed to prove, they could be used to generate questions, they could be used to teach mathematics etc. However, none of these tasks can reliably be performed by AI if AI algorithms do not have mathematical understanding.

The main question that this thesis tries to answer is whether LLMs can have mathematical understanding, and if they can, whether they already have it. The focus in this thesis is on LLMs, since these models are some of the most advanced AI models currently and because they have already been used for mathematical tasks [14]. LLMs can also be looked at specifically, since any findings on mathematical understanding in LLMs could translate to AI algorithms in general.

Chapter 2 will discuss mathematical understanding in detail by discussing

what mathematical understanding is on a philosophical level and by looking at attempts to measure understanding. Chapter 3 places the notions of understanding in a historical context by reviewing how mathematical understanding has played a role for the traditional uses of computers in mathematics. After briefly discussing the differences between traditional computers and AI, chapter 4 discusses how Large Language models work. After it is made clear how they work, the potential and some limitations of understanding in LLMs are discussed, along with how understanding might be measured. Lastly, chapter 5 discusses an experiment that aims to show how understanding in LLMs could potentially be measured.

Chapter 2

Mathematical understanding

When trying to determine whether large language models can have mathematical understanding or help develop it, it is important to first determine what understanding actually is. Understanding is a difficult thing to properly define and several theories of understanding have been suggested over the years. As this thesis pertains to understanding of mathematics, the theories discussed will be geared towards mathematical understanding.

2.1 Levels of mathematical understanding

The idea that understanding in mathematics is either complete or there is no understanding at all is not a widely held belief [60, p. 48]. The more commonly held belief is that there are different types or different levels of understanding. In “Aspects of Mathematical Understanding”, Yoong mentions that mathematical understanding is the idea that different types (or levels) of understanding exist [60].

There are three different types of knowledge, namely ‘knowing that’, ‘knowing how’, and ‘knowing why’, that are mentioned [60, p. 45]. These three types of knowledge can be considered different levels of understanding. ‘Knowing that’ refers to basic knowledge. For example, knowing that $1 + 1 = 2$ or knowing that a triangle contains three angles are examples of ‘knowing that’.

‘Knowing how’ refers to knowing the underlying rules that lead to certain outcomes. Inferring how numbers get bigger in the Arabic numeral system would be an example of this. Knowing that the first positive integers are 1,2,3,...,9 is ‘knowing that’. If you know this is followed by 10 through 19 you could suspect that this will be followed by 20 through 29, but there is not enough information to actually establish a pattern. Once someone learns that 19 is followed by 20 through 29 it is possible to infer that 30 through 39 come after 29. The ability to figure out or know this underlying mechanism of how numbers are ordered is

an example of 'knowing how' since it establishes a pattern. Someone with this knowledge does not need to explicitly know that 23831 is followed by 23832, but they could still figure it out from the pattern they know.

'Knowing why' is different from 'knowing how' in the sense that it explains the 'knowing how'. In the example above, 'knowing why' could take several forms. One way one could know why the pattern in the numeral system exists would be from a historical context. Just knowing the origins of the numeral system would be a form of 'knowing why' according to some theories [32], but knowledge of how the numeral system originated might give insight into why the Arabic numeral system became the dominant one instead of another numeral system. This in turn could give insight into strengths or weaknesses of the Arabic numeral system. However, it could be argued that knowledge of the history of numerical systems is a form of 'knowing that'. These definitions are not perfect, but they indicate different levels of understanding.

Besides 'knowing that', 'knowing how' and 'knowing why', other types of understanding have also been put forth. When it comes to mathematical understanding specifically, Jeremy Avigad makes a distinction between understanding mathematical language, understanding mathematical proofs, understanding mathematical domains and structures, understanding mathematical inference and understanding mathematical diagrams [6]. While these are similar to the three different types of knowing, the categories are mostly different ways to categorize types of knowing how and knowing why. Understanding mathematical language could be described as knowing that, since knowing the mathematical language of a field is a prerequisite to having any of the other types of understanding.

The perspective of different types of knowing is one that is often mentioned. The idea that knowledge alone is sufficient for understanding is generally not accepted [5, p. 319]. At the same time, however, knowledge is considered essential for understanding, since knowledge provides the necessary context in which understanding can exist [16, p. 126]. Understanding consisting of just the ability to make connections is another theory that has been put forth. Especially in the case of mathematical understanding, this perspective makes sense. Learning symbols and operations and the meaning behind them and then making connections between them is the basis of mathematics, especially since math is based on deduction, so from just base knowledge everything should be inferable. Knowledge would still be an aspect of this theory, but knowledge is a part of making connections.

The theory falls short, however, when one considers the harder to define aspects of understanding. Intuition and the ability to appreciate beauty and depth are aspects of understanding that are harder to define, but that cannot be reduced to just making connections [16, p. 126]. Intuition is often defined as getting an answer or estimate without explicitly going through all the steps required to get the answer. But if understanding were only making connections, then doing all the steps explicitly would suffice for understanding. At the very least this theory would have to exclude these aspects of understanding.

Another possible way to divide understanding is to it into different meth-

ods or strategies [6, p. 18]. This way of understanding is very different from using different types of ‘knowing’, but this idea still runs into problems. When reducing understanding to methods and strategies, mathematical understanding seems to become almost algorithmic [6, p. 18]. Algorithms require very specific descriptions, but the criteria of what falls under the scope of a certain method or strategy is also often unclear [6, p. 19]. It could also be said that this is just dividing mathematics into different variations of ‘knowing how’, which would make the scope of this theory more narrow than having different types of knowing.

Since such simplified theories of mathematical understanding fail to encapsulate all the different aspects of understanding, it makes sense to insert levels of understanding into these theories. That way, when some property of understanding does not fit the idea that understanding is just making connections, it could be said that that property is part of a higher level of understanding, while the ability to make connections is part of a lower level of understanding.

2.2 Aspects of understanding

2.2.1 Contested properties

While some theories adhere to the idea that there are levels of understanding, there is no consensus whether understanding can be divided into levels. Some theories do not consider understanding to have different levels, but these theories still say that understanding has several different aspects to it. Although there are different aspects that would together form understanding, it is unclear which aspects these are. Sometimes it is not even clear how to define them. Examples of qualities of understanding that are hard to properly define are explanation, beauty and depth [43].

Qualities that are hard to define are often also the most contested as aspects of understanding. The ability to appreciate beauty could be considered a part of understanding, but it could also be the case that someone can construct a mathematical proof, know all the historical background of the proof and even improve on a proof without considering anything about that proof or the process of constructing it beautiful. Since this person could construct and improve a proof and is familiar with context surrounding the proof, it is hard to argue that this person has no ‘true’ understanding of that proof. This means that the ability to appreciate beauty is contested on the grounds that it is not necessary to have that ability in order to have understanding [16, p. 127] [43] [17, p. 1113].

Some qualities like depth are contested based on a different notion, namely that they cannot be formulated clearly. For example, depth could be expressed as a certain combination of forming connections and recognizing patterns. However, these two are also separately considered properties of understanding according to several theories, so maybe using depth as a vague umbrella term of other properties of understanding. Besides it potentially just being an umbrella term of more clearly defined properties, it is also unclear which properties would

exactly fall under the umbrella. The ability to filter irrelevant information while recognizing which information is relevant can be considered an important part of understanding, but it is unclear if this is also a defining feature of depth or not [16, p. 126] [17, p. 1113].

2.2.2 Explanation and understanding

Explanation is yet another potential property of understanding, but it is contested for another reason than the other contested properties mentioned above. There are two main ways that explanation is considered to be related to understanding, namely as a necessary component or as a sufficient one. If explanation is considered a necessary component, it means that for someone to have understanding they would need to be able to explain a subject to have understanding. This would not mean that explanation is the only requirement necessary for understanding. If explanation is considered a sufficient part of understanding, it would mean that anybody that can explain a subject would understand it. In this case explanation could be the only requirement for understanding, but it might also be possible that someone could have understanding without the ability to explain. There are arguments to be made against both explanation as a necessary property of understanding and explanation being a sufficient property.

In the case that understanding is a necessary property there are several situations that could be used as a counter argument. One might be the case where a person suffers of aphasia, which is a condition where someone is unable to formulate or understand language due to damage to a specific part of the brain [10]. We can imagine that someone was a mathematics teacher with understanding of several mathematical subjects before getting aphasia. As a mathematics teacher this person would have shown their ability to do, for example, differential equations and they would have shown the ability to explain differentiation. As soon as they got aphasia, they would be rendered unable to now give these same explanations, yet they would be able to solve differential equations [8]. If explanation were considered a necessary trait of understanding, it would mean that this mathematics teacher that can still solve differential equations would now no longer understand those equations due to an inability to explain.

On the other hand, a counter argument might be that aphasia is similar to having no tongue, meaning that the person would be able to explain it if they were physically able to do so. In this case explanation could be considered a necessary property of understanding provided that the person is physically capable of expression themselves. Since aphasia is caused by damage to the brain it could be considered physical, but since the brain is responsible for all mental faculties, it could also be considered a mental disorder.

Since there is something to be said for both sides of this argument, it is hard to determine whether explanation should be considered necessary. In spite of this, it could be argued that explanation is necessary for understanding in computers, as we cannot know whether a computer possesses understanding if it is unable to communicate its understanding to humans. To be more clear, for one of the research questions we hope to determine whether Large Language

Models can have understanding of mathematics, but we also hope to determine if mathematical understanding is necessary is doing for mathematics. So while explanation might not be necessary to have understanding, understanding is almost impossible to determine in an AI if there is no way for humans to measure that understanding. On top of that, LLMs should be able to communicate in a way that can be understood by humans, which means that they are at least physically capable of explaining things.

In the case of explanation as a sufficient component of understanding, a variation of the Chinese Room from Searle serves as a counterargument [49]. The Chinese Room is a thought experiment that considers a man in a locked room that receives Chinese writing of some sort and needs to produce written responses in Chinese, but the man does not speak Chinese. In the case of mathematical understanding one can consider a similar kind of situation where a student attends a mathematics lecture and their friend who had to miss it. The student that attended the lecture has a good memory, which means they were able to memorize the whole lecture. However, despite remembering the lecture the student did not understand the subject matter. After the lecture the friend who missed the lecture and the student work on some assignments that require understanding of the lecture material together, but since the student did not understand the lecture and the friend did not attend they struggle to do the assignments. The friend asks if the student who attended could explain the material to him. The student is able to relay the lecture to their friend, and now the friend understands the material and is able to do the assignments. Even though the student was able to give an explanation of the subject that lead to understanding in their friend, they did not have understanding themselves. If explanation were sufficient for understanding, it would seem that the student who is unable to do assignments does understand the subject matter that covers those assignments. It is obvious from this example that explanation is not sufficient for understanding.

Although it seems that explanation cannot be considered a necessary property of understanding and definitely not a sufficient one, explanation and understanding do seem to be closely related. Due to this close relationship, explanation has been used as a way to gain insight into what understanding is [17].

Whether explanation actually happens in mathematics has been called into question by some philosophers of explanation [17]. They believe that mathematical proofs only show that something is true, while, according to them, an explanation needs to also show why something is true. To address this, Frans makes a distinction between explanatory proofs and non-explanatory proofs, where explanatory proofs also explain why something is the case.

2.2.3 Intuition and Understanding

Another property that understanding seems to be tied to is intuition. If someone can look at a problem for only a few seconds to read it and give a rough estimate, the case could be made that this person understands the problem. For example,

if someone is asked what $39365 \div 183$ is they could do long division to find the answer. Doing long division, however, merely requires someone to know an algorithm, but it does not require any semantic understanding. If someone looks at this problem and says within three seconds that it is probably around 218, there has to be some kind of understanding going on that is deeper than just going through an algorithm. Such a fast answer that is close to the true answer shows that someone can play with the numbers in a way that someone who just does the long division trick cannot.

A potential counter-example of this outside of mathematics would be chicken sexers. Chicken sexers can sort baby chicks on their sex with incredible accuracy, but when they are asked how they determine they cannot explain which features they use to determine the sex of a chick. The chicken sexers are often considered to lack understanding since they cannot explain their mental process. However, it could be argued their incredible accuracy would suggest that there is at least some subconscious level of understanding hidden underneath their intuition.

If we were to think of the chicken sexers in terms of 'knowing that', 'knowing how', and 'knowing why', it is immediately apparent that 'knowing that' applies here, as the chicken sexers know if a chick is male or female. They cannot explain why they think this, so they do not possess a level of 'knowing why'. It is debatable whether chicken sexers show that they have a 'knowing how' level of understanding, as the sex of the chicks needs to be inferred from certain traits, but they cannot vocalize the specific features that the chicks have. The chicken sexer case also pleads in favor of a holistic view of understanding, as it shows that sometimes people can infer knowledge holistically without knowing what the individual parts are. This example is complementary to that of a biologist who has identified every cell of an elephant through a microscope whilst never actually having seen an elephant. In this case the biologist lacks understanding of what an elephant is, even though they understand all the parts that the elephant is made of [16, p. 135] [43].

A problem with using this type of intuition or making a distinction between conscious or subconscious is twofold. One of the ways to indicate whether something is intuitive in people is the speed at which they can give an answer. If someone gives an answer two seconds after being shown a math problem, it means that they did not consciously have to go step by step through some algorithm to get to an answer, but instead they could do some of them implicitly or they could skip certain steps. However, a computer works much faster than a person, so whereas answering within two seconds is indicative of intuition in people, for computers it would often be possible to go through all the steps of an algorithm in those same two seconds. This makes it a hard trait to use for determining understanding in AI models/computers.

The second problem is that conscious and subconscious are experiences that cannot be meaningfully explain through words; in other words, they are qualia. In computers, it is not clear what should be considered conscious or subconscious. Since subconscious refers to background processes in the human mind, if someone is playing a video game on their computer and an anti-virus program

runs on the background, should the anti-virus program be considered something that happens subconsciously? It could also be argued that since that anti-virus software is explicitly programmed and the programmed has to be read in order for it to work, that a 'background task' does not exist in the same way as the subconscious does in the human mind. Either way, it is a problem that the dichotomy between conscious and subconscious cannot be made in computers for the role of intuition in understanding.

It is clear that these properties are hard to define, but there are properties that different views seem to agree are necessary for understanding. Since it is important to have some sort of measurement of understanding to determine whether LLMs have mathematical understanding, the next section will focus on some of the properties of understanding that are shared among different viewpoints.

2.2.4 Common properties

In the previous sections it has become clear that there have been several attempts at formalizing a definition of mathematical understanding, but none of these definitions have been accepted by the general communities of philosophers of understanding and mathematicians. There are, however, several elements that are shared by different definitions of mathematical understanding.

Perhaps the most fundamental part of mathematical understanding that philosophers and mathematicians can agree on is knowledge. One needs to be knowledgeable about a subject to be able to have some form of understanding, since knowledge provides the context in which understanding can be formed. That being said, knowledge on its own is not sufficient for understanding. It is obvious that if someone looks up the answer to an equation they now know the answer without necessarily understanding why the answer is correct. Therefore knowledge is a necessary property of understanding, but not a sufficient one.

The ability to make inferences or think logically about mathematical concepts is also mentioned by several theories of understanding. Especially in mathematical understanding, the ability to make logically sound inferences is a core part of doing math. While some theories might contest the ability to think logically or make inferences as a part of general understanding, (almost) all theories of mathematical understanding consider the ability to think logically or make inferences necessary for doing mathematics. This makes sense, since mathematical proofs are done solely through logical inferences while inductive reasoning is an important part of understanding many things outside of mathematics.

The ability to see patterns or relationships between different things is also shared by many theories of understanding. To an extent this goes hand in hand with the ability to make inferences. When making a logical inference, if someone wants to infer A from B, they have to find some sort of relationship between them. If one has to do differentiation of an formula with a product in it, it is important that they recognize that that product warrants the use of the product rule to get the right result. Not recognizing that that formula needs the product

rule applied to it would show a lack of either knowing that or knowing how, and definitely of knowing why. Although mathematical understanding is not always modelled using different levels of understanding, it does show that this lack of pattern recognition already contests certain levels of understanding, let alone total understanding.

It is often agreed upon that filtering information is another property of understanding. If one wants to apply mathematics, it is important to know which information is important and which information is not. If someone wants to know a probability distribution of some sort, they should realize that their geometry knowledge is not applicable in that area. If someone understands the relevant field of mathematics, knowing which formulas correspond to which probability distribution is also a way in which knowledge should be filtered.

Another property that seems to be shared between different perspectives on understanding is the ability to explain. Although we have established that explanation cannot be considered necessary or sufficient for understanding, it is still often considered an aspect of understanding. It has been covered extensively in its own section, but it gets mentioned here again because its explanation is important to determine understanding in LLMs.

2.3 Measuring understanding

2.3.1 Questionnaire

Regardless of how understanding is defined, to determine whether large language models can understand mathematics there needs to be some way to measure understanding. There have been some attempts at measuring understanding, of which three will be highlighted in this section. To determine which qualities define understanding, several theories and methods of measurements have been used. One of the attempts made to determine defining qualities of understanding is an understanding questionnaire made by Wong in 1984. The questionnaire is mentioned by Yoong, but it was part of a PhD thesis that was never published [60, p. 55]. While that makes it hard to use it as a credible source, the paper by Yoong does use it to provide some interesting insights on mathematical understanding. The questionnaire was filled out by three different groups, namely mathematician, mathematics teachers and Dip. Ed. students from the university of Brisbane [60, p. 48]. In the questionnaire, three types of questions were asked to evaluate the participants' views on understanding.

For the first set of questions participants were asked to rate for 50 statements how strongly they related to understanding. The ratings went from 1.0 to 5.0 with 1 meaning a statement was not associated with understanding at all and 5 meaning a statement was very important for understanding. These statements were divided into 4 categories and the mean values of the ratings from the mathematicians, math teachers, Div. Ed. students and the entire group of participants were noted for each statement.

Regarding the statements about the nature of understanding, the lowest

rated statement from every group was that mathematical understanding was absolute, absolute meaning that something was either understood completely or there was no understanding at all. This statement got about a rating of 2.0 on average, with mathematicians rating this statement the lowest at 1.5. Although this does not give a complete theory of understanding or even a definition of understanding, this questionnaire can show that certain properties are probably false, such as mathematical understanding being absolute.

Overall, the study suggested that mathematical understanding was mostly based on higher level processes, such as filtering which information is relevant and which information is not, determining whether obtained answers make sense based on the context, and converting information from one form to another.

2.3.2 Mathematical assessment scale

Many years after the questionnaire another model was used to measure understanding, namely the mathematical assessment scale [21]. It was designed as a tool to assess how mathematical understanding is assessed so (Turkish) mathematics teachers could have a reliable and clearly defined model of mathematical understanding. The mathematical assessment scale was among other things inspired by the previously referenced paper from Yoong [21, p. 3].

The mathematical assessment scale was developed using the scale development model, which is a method that is often used in social- and human sciences. The scale development model is “the process of creating a new instrument for measuring an unobserved or latent construct, such as depression, sociability, or fourth-grade mathematics ability. The process includes defining the construct and test specifications, generating items and response scales, piloting the items in a large sample, conducting analyses to fine-tune the measure, and then readministering the refined measure to develop norms (if applicable) and to assess aspects of reliability and validity” according to the APA dictionary [4].

In the case of the mathematical assessment scale, the researchers started by holding semi-structured interviews with six 4th grade elementary school mathematics teachers from Turkey. From those interviews they got four broad categories into which mathematical understanding could be divided. These four categories existed of several different criteria, for example the categories ‘applying rules’ and ‘transferring knowledge’. After creating these categories a second set of interviews, this time with 17 mathematics teachers, was conducted to determine which qualities were important for understanding in each category. An example from the second round of interviews is that ‘applying rules’ could be divided into ‘using formulas’, ‘using knowledge’, ‘knowing rules’, etc.

The terms that were produced by the second round of interviews were then turned into a set of 41 items. An item would be a statement such as “Whether the students understood the mathematics lesson or not should be evaluated concerning their ability to bring different solution paths to problem situations in mathematics”. After applying the items to a set of 130 students, 20 items were removed as they ended up being very weakly correlated to mathematical understanding. The remaining 21 items were used in the final scale model.

These 21 items ended up having an item-total correlation between 0.431 and 0.759 and an internal correlation of 0.941 [21, p. 7]. The item-total correlation seems to be the correlation of an item with understanding, where a 0 would mean no correlation and 1 would mean perfect correlation.

This method of trying to quantify understanding yielded very concrete results. It allows mathematics teachers to assess understanding in a relatively objective way that can be consistently applied to different students. In the case of computers, most of the items are applicable to the way that computers interact with mathematics and to the way that they relay this information to humans. It aims to bypass the need for a fully defined theory of understanding by using experiences of teachers as data points to build a definition of understanding on.

The results from the mathematical assessment scale show that understanding seems to have different aspects to it. In section 2.1 it was mentioned that understanding is often considered to have different levels, such as ‘knowing that’, ‘knowing how’, and ‘knowing why’. Applying rules is something that would require knowledge of the rules, which could be considered ‘knowing that’, but the previously mentioned example for ‘knowing how’ of being able to scale the numerical system infinitely could also be considered an application of rules. Being able to transfer knowledge is another category that was classified in the mathematical assessment scale that could fit these categories. When someone wants to explain something without having an explanation memorized, it requires someone to be able to explain how and/or why something works the way it does. For someone to be able to explain these things it is required for them to actually know how and why something works themselves, so the ability to transfer knowledge would often require ‘knowing how’ and/or ‘knowing why’.

2.3.3 Computer-assisted scientific understanding

The third measurement method that will be discussed in this section was developed in 2022 by Krenn et al.[34]. Unlike the previous two measurement methods, the method from this study was specifically developed with computer-assisted science in mind and the ability to measure understanding in computers. They base their notion of understanding on one developed by Dennis Dieks and Henk de Regt [13, 34, 44]. From this perspective on understanding they highlight three dimensions. According to Krenn et al. “First, AI can act as an instrument revealing properties of a physical system that are otherwise difficult or even impossible to probe. Humans then lift these insights to scientific understanding. Second, AI can act as a source of inspiration for new concepts and ideas that are subsequently understood and generalized by human scientists. Third, AI acts as an agent of understanding. AI reaches new scientific insight and — importantly — can transfer it to human researchers.”[34, p. 761] The method developed by Krenn et al. will be referred to as the ‘computer-assisted scientific understanding’ method, since the aim of their paper was to determine to which extent AI could contribute to understanding in mathematics.

The first two dimensions mentioned are ways in which AI might help humans gain mathematical understanding, but since the research question focuses on

understanding in Large Language Models only the third dimension is relevant. In the third dimension AI is considered an agent of understanding, which means this dimension is relevant for the thesis. For the third dimension, a set of potential conditions for agents of understanding is identified that according to the perspective from Dieks and Regt. One of the challenges that arises according to Krenn et al. is that it is not sufficient for AI to have understanding, but it needs to also be able to pass that understanding onto humans [34, p. 766]. While it has become clear that explanation is neither necessary nor sufficient for understanding, it is necessary for AI to be able to provide explanations of its newly gained understanding.

Two criteria are identified that would be required to have scientific understanding in AI. The first criterion says that a computer should be able to make qualitative judgments without having to do complex calculations [34, p. 767]. This is similar to humans having (correct) intuitions about a subject they understand. The second criterion is that computers need to show understanding to an outside party, since the goal of computer-assisted scientific understanding is not the understanding in itself, but the ways that this understanding could advance science and, by extension, enrich human lives.

Due to the second criterion, the test that they proposed is similar to a Turing Test. The test requires a teacher, a student (the AI) and a third party referee. The third party referee could potentially be an AI if there would be one that was advanced enough to make such judgments, but in practice it would most likely have to be a person. One of the premises the test is based on is that if someone understands something, they should be able to explain it. Based on this premise it would stand to reason that a teacher would be able to explain some topic to the student. If the student then gained understanding, it also stands to reason that the student would be able to explain it to a third party. The AI would pass the test if the third party could not say with certainty which of the two was the teacher and which one was the student.

Although this is only a suggested test that has not yielded any results yet, it is mentioned here because it is specifically aimed at measuring intelligence in computers.

2.3.4 Criticisms measurements

Although the different measurement methods give interesting insights into mathematical understanding, they are not perfect. One of the main problems from all three studies is that these measurement methods lack the ability to provide a theory of mathematical understanding. The idea of these measurements is to bypass such a formal definition of understanding by looking at practical examples, but that does mean that these measurements do not allow us to fully understand what mathematical understanding actually is.

Besides the general problems of using measurements instead of a philosophical approach, the individual measurement methods might have some flaws themselves that need to be addressed. The questionnaire, besides not giving a formal definition of understanding, was also conducted with a relatively small sample

size. 16 mathematicians took the questionnaire, 49 mathematics teachers, and 28 Dip. Ed. students making the total group of participants 93 people. A small group of participants could mean that they are not representative of all mathematicians and teachers. Given that they were also all working at the same university, it is possible that a cultural perspective from either the nationality or just the perspectives of the department were biased in certain directions. The idea of using a questionnaire to determine which properties are important to understanding could yield useful insights, but it should be taken into account that the results of this particular questionnaire might not be representative of mathematicians and mathematics teachers as a whole.

The main downside of the way the study for the mathematical assessment scale was conducted is the fact that the interviews were done with elementary school teachers in Turkey. Using only Turkish teachers could mean the test has a cultural bias towards understanding. An example of such a bias might be the way grading works. In most Asian countries, if you do exactly what the teacher says you will most likely get a 10/10. However, in countries such as the Netherlands, doing exactly what the teacher says will probably give you an 8/10. These types of cultural differences could also be rooted in the way understanding is experienced in these countries. The fact that they are elementary school teachers might also skew the perception of understanding towards a more basic level of understanding, as children tend to have a more superficial understanding of difficult topics such as mathematics.

Regarding the computer-assisted scientific understanding measurement, it does not justify how it got its notion of scientific understanding well. It mentions that understanding can be considered to have three dimensions, but they just mention that is a perspective they chose without much clarification. It is mentioned that it is “both contextual and pragmatic” [34, p.763], but that about the extent to which they justify their choice. This means that if their chosen notion of understanding is flawed that the whole model is based on a flawed presupposition. In contrast, the study of the mathematical assessment scale clearly shows how to get the criteria for understanding that ended up being used.

The computer-assisted scientific understanding was also not an easily quantifiable model to measure understanding. It mentions qualities that can be used to ‘measure understanding’, but measuring a ”degree” of understanding is not as easy. The other two models do this better, with the mathematical assessment scale using a statistical test that actually quantifies understanding on a scale of 0.0 to 1.0 [21, p. 8].

However, the main problem with the test for scientific understanding is the fact that it is based on two premises that are highly debatable, namely explanation being both necessary and sufficient for understanding. If passing the test is based solely on the ability to explain a topic, then that would imply the ability to explain is sufficient for understanding. However, as was established in section 2.3.3, one could potentially give an explanation without understanding. In the case of the proposed test, if the student gets a general explanation, asks 20 questions that all get answered, memorizes all of it, but understands none

of it, this student might be able to pass the test suggested by Krenn et al.. If the student just repeats the explanation as was given to them and is then only asked questions from the subset of which they memorized the answers, they could give the exact level of explanation given to them by the teacher.

Besides that, the idea that someone that has understanding should necessarily be able to give a good explanation is also not necessarily true. There are very talented researchers in different fields that do not make for the best teachers in those same fields. Sometimes the way that people who understand topics approach problems is very different from people who have less understanding of those same topics.

2.3.5 Conclusions measurements

While the measurement methods are all subject to different criticisms, these different methods all got similar aspects of understanding that they consider the most important for understanding. The first two of these models consider different types or levels or dimensions of mathematical understanding when trying to measure understanding and the computer-assisted scientific understanding touches on explanation and intuition, two different aspects of understanding according to the other two models.

There are four or five main aspects of understanding that can be extrapolated from the mathematical assessment scale: transferring knowledge, applying rules, producing questions and solution paths, generalizing, and exploring knowledge. The individual items are not disclosed in the research, but the individual items from Yoong are. Generalizing results is considered a 3.7 on a scale from 1-5 for example, which is one of the higher rated items in the list from Yoongs questionnaire. Similarly, explaining results is valued similarly in the questionnaire, which fits with transferring knowledge.

One of the most highly rated items from the questionnaire is the ability estimate a numerical answer to a problem. This should probably be categorized as intuition, but that is not one of the categories from the mathematical assessment scale. In the Görgut paper it is put under the category of exploring knowledge [21, p. 4], which also happens to be one of the more highly rated aspects of understanding according to the mathematical assessment scale [21, p. 9].

While it is hard to fit every individual item from the questionnaire perfectly within the categories of the mathematical assessment scale, the items that correlate the strongest with understanding according to the questionnaire can often be categorized into one of the five categories from the mathematical assessment scale. Some of the lower rated items from the questionnaire are more difficult to fit into one of the categories from the mathematical assessment scale, such as the item 'We have little knowledge of how people understand'. This is in part because the mathematical assessment scale is specifically designed for assessing the mathematical understanding in individuals, whereas the items from the questionnaire are about understanding in general.

It makes sense that many of the items from Yoong's questionnaire can be

categorized according to the five different groups from the mathematical assessment scale, as that scale is based in part on the questionnaire by Yoong [21, p. 3]. Since the mathematical assessment scale was produced with a procedure that has been described in more detail than the questionnaire by Yoong, with a procedure that is considered best practice in certain fields of science at the time of writing this, and is in part based on the research by Yoong, the rest of this thesis will use the mathematical assessment scale as a reference point instead of the questionnaire by Yoong. Both of the tests try to measure the same thing through similar methods and it is difficult to find any benefit that is provided by using the questionnaire alongside or instead of the mathematical assessment scale.

To summarize the findings of the different measurements, overall it can be said that the ability to explain is considered vital by all three of the measurement models. The ability to apply knowledge from one context into another context when applicable is rated as important for understanding by the questionnaire and the mathematical assessment scale. Having basic knowledge and knowing how to apply mathematical operation is also considered a key aspect of understanding by both measurement models, together with the ability to make logical inferences and think analytically. The ability to solve math exercises in different ways, however, seems to be considered less important by both models. Since the computer-assisted scientific understanding test requires the 'student' to be indistinguishable from the teacher, these abilities can also be considered similarly important according to this model despite it not explicitly being stated.

Finally, a possible way to combined the three measurement methods to measure understanding in LLMs could be assessing mathematical understanding in 'students' that were given an explanation by a teacher. One or more of these students could be LLMs. The referee could ask questions on the subjects that the students were supposed to understand and rate them according to the mathematical assessment scale. If the referee rated them as high or higher on the mathematical assessment scale than students who passed the tests, it could be argued that the LLMs have obtained understanding. This would suppose that explanation is necessary, but as mentioned in previous sections, it is difficult to determine understanding in LLMs if they cannot explain their answers.

2.4 Understanding and computers

Two main approaches to understanding have been covered; the theoretical, philosophical approach and the measurement approach. While the theoretical approach wants a full theory of understanding and the measurement approach tries to avoid having to formulate such a theory, the different approaches have similar findings.

Different types or levels of understanding is one of the aspects of understanding that there seems to be a consensus on. How these are viewed from the philosophical perspective has been mentioned in a previous section. The way

that it was mentioned was through 'knowing that', 'knowing how', and 'knowing why'. While it is not stated outright in any of the measurement papers, the different items from the measurements for the most part fall into one or more of these categories. For example 'A student's understanding of the mathematics lesson should be assessed according to their knowledge of the defined steps of operations' would fall into 'knowing how', while 'Teachers should assess whether or not a student understands the mathematics lesson according to the student's ability to explore mathematical knowledge' and 'Use result to solve 4 routine problems' would fall under 'knowing how'. The items 'Know valid conditions of using result' and 'Translate result from one form to another' from the questionnaire could fall under both 'knowing how' and 'knowing why'.

Some items would require a bit of all three levels of understanding, such as 'A student's awareness that a solution path found in mathematics is also valid for similar situations should be considered when assessing their understanding of the lesson'. Since this item would fall into all three categories, it could be said that this item would fit a model of total understanding better than a layer model. However, a layered model is able to describe all the things that a model of total understanding can, but with more nuance.

While intuition is mentioned in both theory and the measurement models as an important aspect of understanding, it is hard to measure in computers due to them being much faster than humans and the lack of a clear distinction between conscious and subconscious. It could perhaps be measured if an AI explains something in a way that would indicate intuition in a human, but otherwise it cannot be considered necessary when measuring understanding in AI. This could be done by having the AI generate an estimate, although even this would not satisfy all definitions of intuition. In the case of intuition being defined as not going through steps explicitly while giving an answer, the way that algorithms work would potentially make it impossible for them to satisfy this definition of intuition. Chapter 4 will cover the way AI works in detail to clarify this, with extra focus on large language models.

Since there is no one true theory of understanding that every philosopher fully agrees with, it is hard to determine if mathematical understanding exists in LLMs using the theoretical approach. This means that the practical approach is handier when trying to measure understanding, but the downside of the practical approach is that it is not based on a strong philosophical foundation. Luckily that might not be too big a problem in this case, as many of the aspects that overlap in different theories of understanding are also the aspects that are correlated most strongly according to the questionnaire and the mathematical assessment scale. As such, it might be feasible to determine understanding in LLMs by combining the different measurement methods while satisfying most theories of understanding.

A batch of AI students and human students via text messages would have to be evaluated by a set of referees according to the mathematical assessment scale. These referees could be anybody, potentially even an AI. They would also have to be evaluated by a different, independent set of referees that consists of professionals regarding mathematical understanding, such as mathematics teachers,

mathematicians and philosophers of understanding. The set of professional referees would not evaluate them according to the mathematical assessment scale, but based on their intuition. This way it would ensure that the AIs pass a 'holistic' quality of understanding that might not get filtered out by the mathematical assessment scale. If the AI gets evaluated as having a good understanding of the material it would be safe to say that they have actually attained said understanding.

While this test might determine whether a specific AI has attained mathematical understanding or not, it is not able to definitively say whether AI in general will ever understand mathematics. This is one of the main downsides of using a practical approach instead of a theoretical one, as a theory of understanding could potentially provide a detailed enough description of understanding that would allow us to say whether AI fits that description or not on paper. For example, if something like intuition were considered vital for understanding and it was defined as skipping steps when giving an answer, certain types of algorithms might never fit that definition of understanding. And even then such a theory of understanding would not prove that there could never be a type of algorithm that would be able to meet that criterion of mathematical understanding. Because of this, even with a test it will be very hard to actually determine whether AI could potentially attain understanding in the future.

This makes the question of whether understanding is necessary for AI to advance mathematics an interesting one. In case understanding is necessary, one of the questions that would need to be resolved is whether the philosophical challenges that AI faces are different from what traditional computers have faced. Similarly, if understanding is not necessary, it might be the case that AI is essentially the same as traditional computers in mathematics. Either way, the difference in challenges that AI and traditional computers face should be figured out, which is why the different ways that computers have been used in mathematics will be covered in the next chapter.

Chapter 3

Historical usage of computers in mathematical inquiry

Application of tools to assist with mathematics has happened for thousands of years. At first, these tools were simple, but over time these tools have become more advanced. Nowadays, the most complex tool is the computer. It can perform tasks that were unthinkable two thousand years ago. As computers become more complex, they are being used for increasingly complex tasks. As the tasks get more complex, it could be that the tools needed to perform these tasks will need to develop a level of mathematical understanding. Since AI algorithms are the most complex mathematical tools we have nowadays, it stands to reason that these algorithms show the most potential for understanding. This chapter will look at the role of some of the more complex applications of traditional computer algorithms in mathematics to establish whether understanding is a property that computer algorithms can have, and if they can, if that property is unique to AI. Before the computers can be discussed, however, it is also important to first see what the role of tools is and how tools have evolved.

Generally speaking, human working memories are very limited, which makes it difficult to avoid mistakes when doing calculations. Since tools can indefinitely hold on to information, they are often helpful to avoid mistakes which would otherwise be made. One of the simplest and earliest ways people used tools with mathematics was the use of clay tablets. Clay tablets could be used to carve numbers into, and the carvings could later be removed so the clay slab was reusable.

However, not everyone had access to a good education throughout history. People might have had to do calculations as part of their daily life, but they were never taught how to read or write. This meant that writing numbers on a clay tablet could be difficult, since they did not know any numbers. Other tools to do mathematics could be more easily accessible for an average person.

A good example of this is the abacus, which in some places is still used when teaching kids to do calculations. An advantage of the abacus is that, unlike a clay tablet, it does not require the ability to read or write to be able to use it [47]. However, while the abacus can help people perform calculations, the numbers still have to be tracked mentally to some extent.

Calculators do not require the user to do any tracking. Unlike the abacus, calculators have the ability to take a whole problem as input, provided that it is expressed using only mathematical symbols. This means that a user does not have to do any part of the calculation themselves, instead they just have to input the problem and they receive the output automatically. Analog calculators predate the first computer, but in 1960 the first digital calculator was invented. Calculators are probably the most widely known way of how computers are used in mathematics, but there are also several other uses of computers in mathematical inquiry. Some of these will be highlighted in this chapter.

Another application of mathematics was in the field of astronomy [42, p. 47]. Not only are tools often needed for people to do arithmetic, but they are also used for geometry. The ability to draw straight lines and circles is important in the field of geometry, but most people struggle with drawing straight lines or circles. For straight lines anything straight could serve as a ruler, but if someone wanted to have an exact length, a ruler would become necessary. Circles were a bit more difficult to make, so a specially designed tool was required. A compass was used to ensure that circles could be drawn in a straight manner. These circles could be used for different purposes, one of which was drawing a sundial [39, p. 307]. This instruction on drawing a sundial also shows the need for lines of specific lengths, rather than just straight ones.

Much like the abacus, computers are often used instead of the compass to draw circles. While these are some examples of historical tools that can now be replaced by computers, there are also parts of mathematics that only computers have been able to assist with. This section has shown some of the ways that people used tools historically to help them do mathematics and some of the ways in which computers have replaced them. Next, this chapter will give an overview of how computers have been used in mathematics thus far.

3.1 Computer-assisted proofs

Computers have been used in mathematical inquiry in a multitude of ways. One of these ways is using computers to assist with proving mathematical theorems. There are several ways that a computer can contribute to a mathematical proof, which will be discussed through several examples. Some of these proofs have given rise to philosophical challenges regarding the validity of computer-assisted proofs. The examples below will highlight some of the challenges that have been raised by these proofs.

The examples that will be discussed are the Four Color theorem, the Double Bubble theorem, and the solving of the game Connect Four. While the Double Bubble theorem and Connect Four highlight slightly different ways that com-

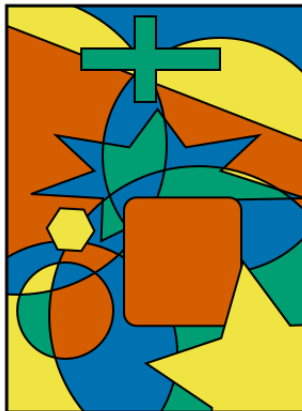


Figure 3.1: Example of a Four Color filled map

puters can be used for mathematical proofs, the main focus will be on the proof of the Four Color theorem.

3.1.1 Four Color theorem

One of the first computer assisted proofs in mathematics is the Four Color theorem [2]. The Four Color theorem was first conjectured by Francis Guthrie in 1852 [20]. To explain what the theorem entails, it is important to know what a map is. A map is, essentially, a way to divide an area (or surface) into smaller areas. One example of a map is the way that the United States are divided into several states. The Four Color theorem states that no more than four colors are required to color the regions of any of these maps, such that no two adjacent regions have the same color. Two fields are adjacent if part of the border (or edges) overlaps, but just corners touching does not count as two fields being adjacent [20].

Even though the problem was conjectured in 1852, to this day the only proofs of this problem are computer assisted proofs. In 1879 and 1880 two separate proofs of the theorem were given, but it was later found that these proofs were incorrect [53]. After the proofs were falsified, the Four Color theorem remained unproven for over 80 years [53].

In 1976, however, a computer-assisted proof for the Four Color theorem was suggested by Appel and Haken [2]. The Four Color theorem was proved by showing that there is no map that could ever require five colors to meet the requirements of the theorem [3, p. 36]. This proof consisted of two parts. First, for certain small maps (also called “configurations”) it was shown that they could be replaced by even smaller maps. Configurations that fit this description are called ‘reducible configuration’. Algorithmic tests were done on configurations to show that these configurations were indeed reducible. If at least one configuration in a set of reducible configurations had to be part of every map,

this set was considered an ‘unavoidable set’ [3, p. 36].

The second part of the proof consisted of finding a coloring that met the requirements of the Four Color theorem for each configuration in the set. In the original proof the unavoidable set was consisted of 1936 configurations [59], but a smaller unavoidable set containing 633 configurations was later discovered [45]. Since there were only 1936 configurations, this part could be done by hand. The first part, however, required the use of computers.

The algorithmic tests for reducibility were already established when the original proof was done, but at that time a computer could take up to several hours to run these for one configuration [3, p. 36]. However, even if it took a computer over a thousand hours to test the entire unavoidable set [3, p. 35], it was still much faster than any human could do this. This meant that testing reducibility for 1936 configurations was not possible to do by hand. While the proof relied on the reducibility test, for a large part it also relied on getting an unavoidable set that had a manageable amount of configurations in it [3, p. 37]. If an unavoidable set was found with 200000 configurations for example, the second part of the proof would have been hard, if not impossible, to do.

It is clear that a computer is needed to prove this theorem and why it would be find an unavoidable set by hand. The impossibility of checking all the configurations by hand has made it such that humans cannot actually verify the result generated by the algorithm. This has caused some mathematicians and philosophers of mathematics to say that the proof is not a true formal proof [3, p. 35]. The validity of the proof has been disputed on several ground, most of which were first given by Thomas Tymoczko [56]. Tymoczko’s argument against the validity of the Four Color theorem gave rise to the concept of non-surveyable proofs. In the main paper that criticizes the Four Color theorem Tymoczko starts out by citing the criteria that it would have to meet to be considered a formal proof, namely how convincing is the proof, is the proof surveyable and is the proof formalizable [56].

The first criterion states that mathematical proof needs to be convincing, but convincingness is an ambiguous criterion. According to Tymoczko, no rational person would accept a result that cannot be verified. However, he also mentions that there are philosophical discussion about whether this is a good criterion to define mathematical proofs by. While some philosophers consider convincingness a good criterion for defining what constitutes as a proof, many would agree that mathematical proofs also need to be surveyable and formalizable. Tymoczko seems to be of the latter school of thought, as he mainly tries to argue against the computer assisted proof by focusing on surveyability.

Formalizability is the third criterion that Tymoczko mentions, but it is only briefly covered. A proof is considered formalizable if it can be expressed in a formal language. While formalizability is mentioned in the original paper that criticizes of the Four Color theorem, it is mentioned as a general criterion that mathematical proofs are required to meet. Tymoczko actually seems to have no problems with the formalizability of the Four Color theorem [56, p. 62]. Since the algorithm is written in a programming language that is Turing complete and it makes use of logical operations in the proof, it makes sense that it should

be considered formalizable.

The second and main criterion that Tymoczko mentions is about the surveyability of proofs. According to Tymoczko, for a proof to be surveyable, it would need to be able to be checked by hand. This is the fundamental problem that Tymoczko has with the computer-assisted proof of the Four Color theorem. The example of a surveyable proof that is used in the paper is the proof that Gauss used to show that the sum of the first 100 positive integers is 5050. Since any mathematician could add these numbers by hand and get the same result the proof is considered surveyable [56, p. 59].

Surveyability is closely tied together with the argument of convincingness, since the main problem with an unsurveyable proof is that it cannot be verified by humans and is therefore hard to accept. This becomes clear when one considers a theorem that is considered to be true, but at some point a newly developed computer algorithm goes through all possible examples and finds a counterexample. If the counterexample is sufficiently complex such that no human could verify it, it would be hard to flat out reject the theorem.

Having difficulty to check a counter-example is not the only way in which a problem can be non-surveyable. It is also possible that there is such an overwhelmingly large amount of cases to check that they cannot be verified. Imagine that there is a ball pit filled with one trillion balls that are either black or white. If a computer says they scanned all the balls and says they are all black, can a human know whether that claim is true or not? This is an example of the amount of cases being too large to be checked by a human to be surveyable. The 1976 proof of the Four Color theorem falls into this category.

The problem that arises here is inherent to inductive reasoning. With inductive reasoning it is possible to find evidence in favor of some conclusion, in this case that all of the balls are black, while that conclusion is actually false. If there is one white ball among one trillion black balls, then looking at one million black balls would suggest that all of the balls are black, but that is not true. Mathematics is based on deductive reasoning, so accepting a proof that runs into this problem is not mathematically sound.

Regardless of any of the problems with the algorithm, it remains a fact that the algorithm helped prove a theorem that would otherwise be unproven to this day. It does seem that there must be a place for such algorithms in the field of mathematical inquiry. One of the arguments in favor of algorithms such as the one used for the Four Color theorem is the fact that algorithms can (theoretically) be verified. Verification is different from checking cases, since checking cases only shows that the algorithm works for each case that is checked, whereas verification proves correctness for all possible cases. If an algorithm is incredibly complex this might be impossible, but the Four Color theorem proof was actually verified using Coq [20], a proof assistant that will be discussed in more detail at a later point in this chapter. While verification seems to solve all the issues, the verification itself is pretty complex. This means that all the definitions used and all the steps taken would still have to be verified by hand, which is quite hard. Some skeptics therefore remain unconvinced of the validity of the computer-assisted proof.

When it comes to mathematical understanding and machines, most people are reluctant to assign any understanding to deterministic algorithms. Although it is unconventional to assume that deterministic algorithms can have understanding, it is assumed that algorithms can possess mathematical understanding in this chapter. This assumption is made to highlight the differences in mathematical understanding between traditional computer algorithms and AI models.

The understanding that was required for the algorithm used in the original proof could be debated. On the one hand, the software used to determine reducibility has to go through many logical operations to get any answer. It could be argued that this logical reasoning shows the ability to perform logical inferences. Finding a set that is unavoidable, similarly, requires logical steps to determine whether the set is indeed unavoidable. On top of the logical reasoning, some knowledge of reducibility and unavoidability could be attributed to the algorithm that was used.

On the other hand, the algorithm is, by definition, algorithmic. By this I mean that the algorithm simply performs a pre-determined set of checks and operations. It is hard to argue that someone has understanding of a topic if they simply repeat an operation that they were told. A student might know that the derivative of x^n is $n \cdot x^{n-1}$, but they might not know that a derivative represents that slope of a graph. It would be hard to say that student truly understands derivation by just knowing how to do certain derivations.

It is unclear whether a level of mathematical understanding should be attributed to the algorithm, but this is more than what could be said for the more basic uses of computers. Calculators, for example, can manipulate numbers in an algorithmic way, but those tasks are far less complex than what the Four Color algorithm had to do. The sheer complexity of the algorithm means that mathematical understanding could be considered for the algorithm used to prove the Four Color theorem.

3.1.2 Double Bubble theorem

While there are several other computer-assisted proofs done after the Four Color theorem, many ran into the same problem of the proofs being basically impossible to verify by humans. The Double Bubble theorem is an interesting case in that regard, as it is a proof that is much smaller. A standard double bubble consists of two spherical surfaces meeting at 120 degree angles. These spherical surfaces are separated by either a flat surface, as is the case with the leftmost double bubble in figure 3.2, or by a third spherical surface, as is the case with the middle and right bubbles in figure 3.2. The double bubble theorem states that, out of all the ways that two surfaces can be enclosed and separated, the standard double bubble has the lowest perimeter [30].

Since the nineteenth century the Double Bubble theorem was believed to be true. However, the theorem was only proven to be true in 2002 [30]. In 1995 the proof was simplified after it was found that in the case where two bubbles were of equal size, that there is only one nonstandard bubble possible for this configuration. This allowed him to let a 1995 desktop computer prove that the

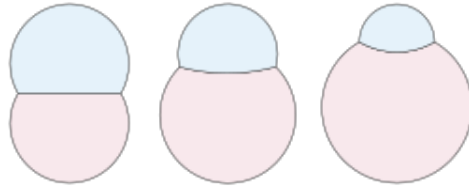


Figure 3.2: Double bubbles in Euclidean plane with different area combinations

Double Bubble theorem held for this case with a program that only had to run for 20 minutes [24]. In 2002, the complete formal proof was done without the use of computers.

Unlike the Four Color theorem, the Double Bubble theorem has eventually been proven in 2002 by Michael Hutchings without using a computer, but the first partial proof was done with a computer. This case highlights that computers do not necessarily have to solve a whole problem, but instead can show how a proof can be done to assist with a proof. Besides that, whereas the Four Color theorem would be impossible to check by hand due to its size and complexity, the initial partial proof of the Double Bubble theorem took 20 minutes on a desktop from 1995. In this case the computer merely made it easier to prove the case, but it was possible to be done by hand as shown in the final complete proof.

The role that the computer played in this proof was similar to what a calculator could do. While it allowed the proof to be done faster, it did not play such an essential role that the proof could not have been done by hand. The computer was used to ensure that no nonstandard perimeter-minimizing double bubble that consisted of two spherical caps with a toroidal band between them was inside of a certain space [30, p. 461]. This method is essentially finding the correct answer in a multiple choice question by ruling out the other possibilities one by one. Therefore, the algorithm needed to have knowledge in order to filter the parameter space accurately, but, besides knowledge, the algorithm the algorithm did not need any other level of understanding to function.

3.1.3 Connect Four

The final case of the proof assistance that will be highlighted in this paper is the solving of the game Connect Four. Connect Four is a game played by two players on a vertically set board of consisting 7 columns and 6 rows. Both players have stones of a different color and they take turns putting in a stone that drops into one of the columns. This game has been solved using computers by brute-forcing all the possible scenarios, meaning that the outcome is known with perfect play from move one.

The player that begins can force a win if they put their first stone in the middle, provided they play perfectly. Similarly, the game is a theoretical draw or a win for player two depending on where the first player puts their first

move [1]. This case differs from the previous two proof assistants in the sense that the computer proof might not be easily verifiable by humans, but it's not theoretically impossible should one choose to make the effort. Although it will probably never be checked by hand, the fact that it is theoretically verifiable by humans seems to make them more convincing.

Similar to the double bubble theorem, the algorithm here just had to go through all the possible configurations. However, there might not even have been much knowledge necessary here. The algorithm simply had to check after any move if there were four stones with the same color in a row. In a person, this would only require knowledge that players take turns, but besides that it would take vision of the board and memory of the previous move. This task required barely any knowledge, let alone any mathematical understanding.

3.2 Falsifying conjectures

The previous section highlighted how computers were used for proving theorems. Another example of how computers can be used is to disprove theorems. While the process of doing this might be similar in the sense that it would involve a computer checking a large number of potential counterexamples, once a counterexample is found it can often easily be verified that the produced counterexample is in fact a good counterexample.

The reason that computers are often good at finding counterexamples is that finding counterexamples often require calculations that humans are not good at. When trying to find a counterexample one might have to go through multiple thousands, if not millions of calculations before an actual counterexample is found. These calculations could be fairly simple, but even for calculations as simple as 24×17 a computer will find the answer much faster than most people. If a person trying to find a counterexample would potentially have to do tens of thousands of these types of calculations to find one, it is much easier and faster to use a computer.

3.2.1 Chebyshev's Bias

Since computers are better at computation it is most common for them to be used to provide counterexamples that might only occur at very large numbers. Chebyshev's Bias is an example of a theorem that was found to be untrue. It is based on the fact that all primes larger than 3 can be categorized in two groups: those that are 1 given modulo 3 and those that are 2 given modulo 3. Chebyshev's Bias started as a conjecture stating that the group of numbers x defined by $x \bmod 4 = 3$ is always larger than the group of numbers defined by $x \bmod 4 = 1$.

A counterexample was found that showed that the $x \bmod 4 = 1$ group can be larger at some point than the $x \bmod 4 = 3$ group. The smallest prime where $x \bmod 4 = 1$ is larger than $x \bmod 4 = 3$ is 26861 [22, p. 3]. This is a relatively small number, so could have been found by hand.

However, Chebyshev also claimed that the group of primes defined by $x \bmod 3 = 1$ was always smaller than $x \bmod 3 = 2$ for all primes larger than 2 and 3. Just like the previous example was disproven, by looking at very large groups of primes it was found that this statement is also false at times. This counterexample was found in 1976 and it was at the value of 608,981,813,029 [22, p. 5]. Finding a counterexample this large could only be done by using a computer, since finding all the prime numbers below 608,981,813,029 would simply be too time consuming for a person.

Although both of these conjectures were ultimately proven to be false, they are both almost always true. Since there are some small ranges of prime numbers when looking at very large primes in which the bias does not hold, it is not universally applicable, but it can be useful at times to know that Chebyshev's Bias holds for all smaller numbers.

3.2.2 Euler's sum of powers

Another good example of a theorem being disproven by a computer finding a counterexample is the Euler's sum of powers conjecture which states that for all integers n and k larger than one, $x_1^k + x_2^k + \dots + x_n^k = y^k \Rightarrow n \geq k$. In 1966 the first counterexample was found using a computer search that found that for $k = 5$ the theorem does not hold in the case of $27^5 + 84^5 + 110^5 + 133^5 = 144^5$ [36].

The examples that were discussed highlight that it can be hard for humans to find a counterexample of theorems that hold until very large numbers are considered. Finding that Chebyshev's modulo 3 bias does not hold at 608,981,813,029 is technically not impossible for humans to find, but it would most likely take a very long time to even count the prime numbers preceding 608,981,813,029. In the case of Euler's sum of powers finding the example given is probably doable by hand, but even in 1966 the researchers who found that result decided to do it with a computer since it was faster to do it with a computer, even in 1966. It is not hard to imagine how much easier it would be to find such a result with modern day technology.

The level of mathematical understanding required for the algorithms mentioned here is limited to non-existent. The role of the computer in both of these proofs is very simplistic. To disprove Chebyshev's Bias, an algorithm would need to be written that can check if a number is a prime, in which 'team' is belongs and then simply add 1 to the size of that team. The only aspect of mathematical understanding required for that algorithm is the ability to know the definition of a prime and the ability to divide by three. Similarly, for Euler's sum of powers, an algorithm to disprove it would only need to be able to take a list of numbers as an input and the exponent as an input. Having the computer go through a list of inputs and calculating them makes the computer a simple calculator in terms of understanding.

3.3 Proof checking and finding proofs

Besides proving or disproving theorems, computers can also be used to check whether a proof that has been given is actually correct. Although proof checking is usually considered a form of proof assistance, it is a different way of assisting with a proof than the previous examples that were given. As such, the philosophical challenges that arise from using computers for proof checking are different than those that come up for the previous examples.

Formal proofs consist of sequences of sentences. If a proof is correct, each sentence is either an axiom or something that follows from the axioms. Proof assistants are computer programs that can take these sequences as input and easily check whether each sentence actually follows from the axioms. This is not the same as automated theorem proving, since it is only limited to checking whether the proof is correct or not. While checking for any inconsistencies in a proof with a proof assistant can make proof checking easier, it is not impervious to mistakes. If a large proof needs to be checked, using a proof assistant is faster, but checking whether the proof assistant has actually assessed the proof correctly would need to be done by hand. This could be done by checking if the code only does what it is supposed to do, but again, the person checking this could miss mistakes in their assessment of the code.

Usually proof assistants can be used to help with checking proofs. There are several proof assistants nowadays, with Coq and the HOL theorem provers being some of the more well known ones. HOL stands for Higher Order Logic. There are multiple theorem provers that are part of the HOL theorem prover group, some examples being HOL Light, HOL4, and Powerproof. The HOL proof assistants are well-known, but in mathematics they are not the most popular ones nowadays.

Coq is a theorem prover that is pretty popular among mathematicians, but it is also used by computer scientists [18, p. 7]. It was originally released in 1989, but it has still been getting regular updates. Coq is versatile compared to other theorem provers, as it is able to deal with higher order logic, dependent types, using a kernel, proof automation, proof by reflection and code generation.

To do a proof in Coq, an initial set of propositions needs to be defined. Logic rules are then applied to these propositions. When the initial propositions are all declared, the theorem needs to get stated, after which the proof is initialized by writing **Proof.** After initializing the proof, the user can introduce variables and ask Coq to essentially make the logical inferences given those new variables. A proof is saved by writing **Qed.** Since each step needs to be put in by hand, the proof assistant essentially only follows the logical inferences that follow from each step.

When trying to do a proof, the theorem gets stated beforehand, but it can be introduced in the proof by writing **intros H.** Some of the instructions that need to be given to Coq are seen in a proof of the statement $(\forall x \in A : P(x)) \vee (\forall y \in A : Q(y)) \implies \forall x \in A : P(x) \vee Q(x)$. The way that the author of the proof wanted to prove H was by splitting H into H_1 and H_2 [48]. In Coq the command **destruct H as [H1 — H2].** was given as input.

Having to give such an instruction manually shows some of the limitations of proof assistants. The user needs to guide the proof assistant into making the right choices to actually do a proof. That said, for simpler proofs the input could simply end up looking as follows [48]:

```
Proof.
firstorder.
Qed.
```

After a proof is saved, it can be printed. An example of such a proof is shown below [52]. Note that this is from a different proof than that of $(\forall x \in A : P(x)) \vee (\forall y \in A : Q(y)) \implies \forall x \in A : P(x) \vee Q(x)$.

```
length_corr =
  fun (n : nat) (s : seq n) =>
    seq_ind (fun (n0 : nat) (s0 : seq n0) => length n0 s0 = n0)
      (refl_equal 0)
      (fun (n0 _ : nat) (s0 : seq n0) (IHs : length n0 s0 = n0) =>
        eq_ind_r
          (fun n2 : nat => S n2 = S n0)
          (refl_equal (S n0)) IHs) n s
  : forall (n : nat) (s : seq n), length n s = n
```

A more difficult task that proof assistants can be used for is finding formal proofs. This task differs from automated theorem proving since it requires a human agent to interact with it. However, since the user has to manually input the theorem together with the steps to solve it, proof assistants are not great at finding proofs [18, p. 6]. The only thing that they usually help with are proofs that have many case distinctions, only because it would be too arduous for a human to prove by hand [18, p. 7]. This is similar to the role of the computer when proving the Four Color theorem. So, while finding proofs is more difficult than checking proofs, the role of the computer for finding proofs is still relatively simplistic.

The level of understanding that is required from proof checkers is potentially a bit higher than that of a calculator, but it is still similar. It needs to be able to make certain inferences, but it can only make inferences based on the parameters that are laid out by the user. When choices have to be made, such as splitting an or-statement, the user has to tell the proof assistant to do this. This does indicate a lack of mathematical understanding. Although the proof assistant is capable of logical reasoning, it cannot explain the answers that it gives. It does have some knowledge, however, and it is capable of writing down a formal proof that was generated, but it did have to generate the proof with the help of the user. The level of understanding required for proof checking is, therefore, slightly higher or equal to that of a calculator.

Since proof assistants require specific instructions when doing proofs, the role of the proof assistant in proof finding essentially comes down to the role of the computer algorithm in the Four Color theorem. Therefore, the level of mathematical understanding required for proof finding is going to be the same as the level required for the Four Color theorem algorithm. This is especially true considering the fact that the computer-assisted proof of the Four Color theorem was verified using Coq, which reduced proving that the algorithm used is correct to proving that the verification in Coq was correct.

3.4 Computers and understanding

It should be said that the role of computer in understanding is limited to the current outlook on it. What constitutes as interesting or proper mathematics has changed quite a bit over the centuries [7]. Newton considered geometric approaches to be the golden standard, while considering algebraic approaches to be lesser. He considered the complexity of algebraic solutions to “express in an intolerably roundabout way quantities which are designated by the drawing of single line.” [7, p. 113] Nowadays, the use of algebraic methods actually shows some prowess in calculation and is considered an integral part of mathematical understanding [7]. Similar to how this has view has changed over time, the way that computers are seen in mathematics might change as well.

This chapter has highlighted several ways in which traditional software has helped expand mathematical understanding, but while all the models mentioned above are capable of helping with mathematics and can help with gaining mathematical understanding, they do not possess mathematical understanding themselves. Looking at the properties associated with understanding or evaluating some of these methods according to the mathematical assessment scale should make this clear.

Simple tools, such as visualization tools for example, might be the most obvious type of computer tool that lacks understanding. All that visualization tools do is take numbers that are put in by a human agent and display them in a way they are programmed to. The most basic aspect of understanding that everyone seems to agree on is knowledge and even that is something that visualization tools lack. They only thing they could be considered to ‘know’ is what to draw when one of the variables they take as input is given a specific number.

The way that computers were used to falsify conjectures also did not require much, if any, mathematical understanding. The algorithm used to solve connect four has a similar lack of understanding, since both barely require any knowledge. In the regard, the role of the computer used to solve the double bubble theorem needed to at least know the definition of what a nonstandard perimeter-minimizing double bubble that consists of two spherical caps with a toroidal band between them is. These algorithms have the same level of as calculators when it comes to understanding, if not less.

Proof assistants might require slightly more understanding to check formal

proofs, but due to the help it requires from a user, the amount of understanding is still minimal. Proof finding is slightly more complex, but the role of the proof assistant is often still limited to doing arduous case checking. They cannot explain the proofs, nor can they learn. Proof assistants are similar to a student that can perform all the steps of a calculation, provided that you tell them what step they need to take next.

The algorithm used in the Four Color theorem was able to go through many possible colorings of various maps, but that is all the algorithm could do. It is incapable of explaining why it is the case that each possible map can be colored with 4 colors, even though it was able to show a correct coloring for each map. As discussed in the previous chapter, it is debatable whether explanation should be considered a necessary property for understanding. Since the algorithm was not designed to explain its findings, it could be said that it is just unable to due to physical limitations.

When looking at the algorithm from a perspective of the common properties of understanding, it might actually seem like the Four Color algorithm possesses some level of understanding regarding the process of reducing maps. In the case of the knowledge property, it could be said that it knows about the reducibility for each configuration. Since the algorithm has to go through many logical steps to determine the reducibility of each configuration, it seems that it is able to 'think' logically about the topic. The only property that it clearly lacks is the ability to find patterns or relationships between different concepts. It only focuses on coloring maps, meaning that it cannot even really recognize different concepts. The same traits could be assigned to Coq, especially since Coq verified the proof of the Four Color theorem. Since Coq can get applied to many proofs, there is an argument to be made that Coq has this same level of understanding for a variety of proofs. Having understanding of more proofs could mean that there is more mathematical understanding overall, but it would be of the same, relatively low level as the the algorithm from the Four Color theorem.

However, while some properties of understanding seem to be present, the algorithm would not be able to pass any of the measurement methods of mathematical understanding. Since it cannot communicate in a dynamic way, it would never be able to pass the Turing-like test from Krenn et al.. Regarding the mathematical assessment scale, the algorithm does not possess many of the properties covered by the items of the mathematical assessment scale. One example of this is the fact that there is a whole category called 'Transferring knowledge' which focuses on applying the knowledge that was attained in different scenarios. The Four Color theorem is not a problem that has any other applications, so none of the items in this category are really met by the Four Color theorem algorithm.

Besides the Four Color theorem algorithm, the other traditional computing algorithms also do not meet the criteria for understanding for similar reasons. Proof checkers and proof assistants require human interaction, which means they are merely tools that are used by people. Saying proof assistants possess understanding of formal proofs would be similar to saying that a calculator understands calculus.

The modern perspective of mathematical understanding seems to value depth and complexity of proofs [7]. While all the aforementioned types of traditional uses of computers do not possess mathematical understanding according to this modern perspective, AI algorithms might. The main difference between AI and traditional algorithms are the fact that AI is probabilistic and that it can learn autonomously. This allows it to see patterns that could be missed by rigorously following rules. In turn, those emergent patterns could lead to mathematical advancements that would potentially never be found otherwise. While the definitions of depth and complexity are vague and therefore not the most practical to assess understanding with [43], the fact that AI might combine the calculation power of the computer with a human-like ability to learn means that it might find and even solve proofs that are considered deep and complex.

This means that, while traditional computer algorithms might not be considered to have understanding and or often not considered to have a place in ‘proper’ mathematics, AI algorithms could be considered to have both of those. Even if there are skeptics at this point in time, perspectives on mathematics and understanding change [7].

Since it is clear the AI might raise different challenges regarding mathematical understanding, the next chapter will focus on AI. However, this is quite broad, so a more narrow scope was chosen. One of the types of AI model that has shown a great deal of success when performing text based tasks in different disciplines is the Large Language Model (LLM). Mathematics is one of these disciplines [14]. The next chapter will focus on LLMs for these reasons.

Chapter 4

AI in Mathematical Inquiry

4.1 What are Large Language Models?

Now that the traditional computer programs and their role in mathematical understanding have been covered, it is time to look at the role AI has played and can play in the mathematics. To be able to discuss in what ways AI could have a different role in understanding than traditional computing, some properties of the AI models used in mathematical understanding need to be discussed. Since this thesis will focus on large language models, this chapter will highlight some of the properties that large language models (LLMs) have that are important for their capacity for mathematical understanding.

While the term ‘large language model’ does not have a clear definition, language models are used in the field of natural language processing (NLP) by applying different statistical and probabilistic methods determine the probability of some word sequence occurring in a sentence. In other words, a language model can be defined as probability distribution over a sequence of words [33]. ‘Large language model’ generally refers to language models that are (generally) trained with a deep neural network using large quantities of unlabeled data [38].

It is important to know what neural networks are before discussing deep neural networks any further. Neural networks are a type of model architecture that aims to mimic biological neurons [31]. Neural are usually depicted in the form of graphs as can be seen in figure 4.1. Each node in a graph represents a neuron and it contains a value that represents a piece of data. Each edge has a weight assigned to it. The first layer is of a neural network is the input layer. Each piece of data from the input layer L_1 in figure 4.1 is represented with x_i .

When the data goes from a node in the current layer to a node in the next layer, it has to do so via an edge. The new node will receive the value from the original node multiplied by the value of the edge that connects the two nodes, as can be seen in figure 4.2. These values get summed up, often together with a bias. This bias is represented by the nodes with a + in figure 4.1. Having a bias allows shifting of the activation function, which is where the summed up data

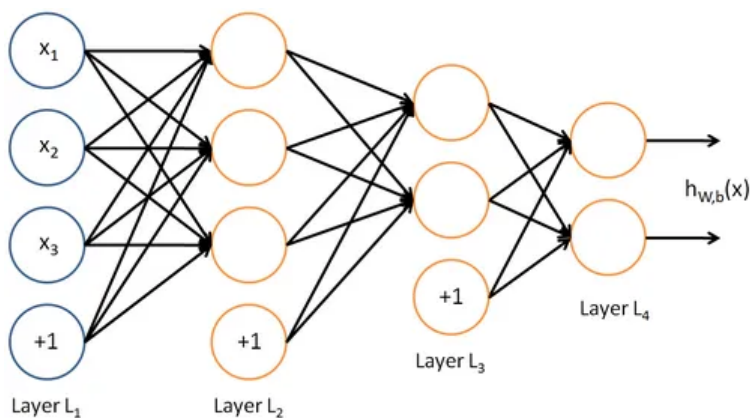


Figure 4.1: Example of a Feed-Forward Neural Network

goes into. This activation function serves to change a linear function into one that has more dimensions [50]. The summed up value is also often changed into a value between 0 and 1. Some of the more common activation functions are ReLU, Tanh and the sigmoid functions, but there are many activation functions that can be used [50].

This process gets repeated for each neuron in the current layer, then it gets repeated for the next layer and so on, until the output layer is reached. This, however, does not allow a neural network to learn. To learn, a neural network needs to change certain weights in the network based on the outcome. This is generally done using gradient descent [46]. While there are several algorithms for gradient descent, they are all based on the same principle. η is considered the learning rate, w is the parameter that needs to be updated (in this case the weights of the vectors), and $\nabla_w J(w)$ is the downhill gradient of a loss function $J(w)$ parameterized by w . Using these symbols, the algorithm used for batch gradient descent looks like $w_{new} = w_{old} - \eta \cdot \nabla_w J(w_{old})$ [46, p. 2]. The batch gradient descent shows that the value for w gets updated based on the the learning rate and the downhill gradient. Doing this will lead to finding a (local) minimum when done over several iterations.

Deep neural networks are complex models that used many hidden layers to extract high level features. Simpler neural networks used in image recognition might find edges, but identifying numbers or a specific type of animals requires an ability to combine simple features into a complex ones. Which connections need to be made to recognize whether something is a cat or a dog is hard to write out in specific rules. This is why deep neural networks are used instead, since a neural network simply tries to extrapolate statistical and probabilistic patterns from their input data. This does mean that it is unclear what exactly happens under the hood of deep neural networks, which is why they are often considered black box models [40].

Since LLMs are trained with deep neural networks, the inner workings of

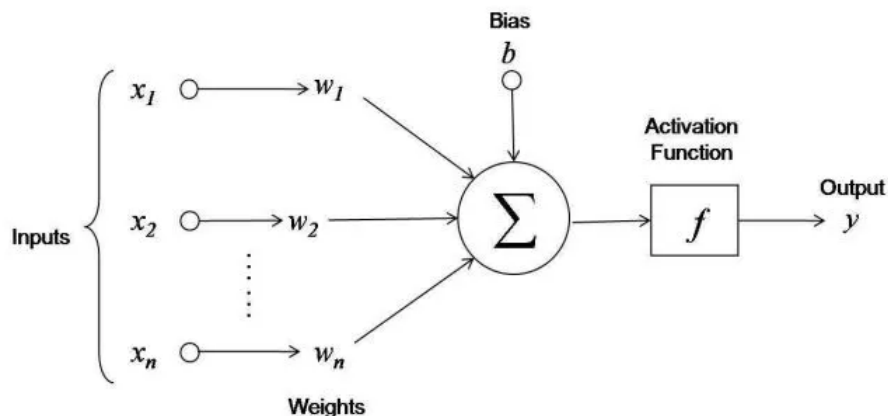


Figure 4.2: Example of artificial neuron

these models are rarely fully known. For mathematical understanding this does not necessarily have to pose a problem. The subjective experiences of other peoples' understanding of mathematics cannot be measured or quantified. The only aspect of understanding that could be measured is the outputs that people produce, but their experience of understanding cannot be shared with others. This means that LLMs being black box models does not have to pose problems.

However, it could make it harder to determine whether certain aspects of understanding are present in LLMs. People all share similar 'hardware', so it is generally accepted that their experiences are similar. In the case of artificial intelligence the hardware is different, so the way that LLMs would experience understanding might differ from how people would experience it. Most people recognize someone's face almost instantly when looking at it, but some people have a condition that prevents them from doing so [51]. These people would have to look at all the individual features and consciously assign that combination of features of to the face of someone they know. This is similar to how deep neural networks seem to work, so the question could be asked if the experiences of AI models differ too much from human experiences to assign understanding until both the brain and deep neural networks are better understood.

At the same time, it is possible that every person processes faces in the same way, but that the individual features get combined subconsciously. Since it is hard to determine which parts of an AI program would 'feel' subconscious for them and which they would experience consciously, it is entirely possible that LLMs actually do process information in the same way as most people. This shows that a potential problem caused by black box models is the potential gap between human intelligence and artificial intelligence. At the same time, since human intelligence is not fully understood and is often judged by outcomes, a black box does not necessitate a problem.

4.1.1 GPT

Model description

At the time of writing this, the most well-known and broadly used AI chat bot is ChatGPT. This chat bot is based on the GPT-3 large language model that was developed by OpenAI. GPT stands for Generative Pre-trained Transformer. Pre-trained models are models that are trained on large amounts of data, often for a specific task. These models can then be adjusted by changing the weights of parameters or by training it on new data. Early pre-trained models were often trained to learn good word embeddings.

NLP models do not know what words are or what they mean, but they can understand vectors. This is why words are represented in a vector space, which is also called an embedding space. The different positions in a vector often represent different properties. These positions are gained from syntactical and semantical differences. There are various algorithms that could be used to assign positions in an embedding space, but an example of how semantics might play a part can be seen in the following example. The words ‘man’ and ‘king’ could have very different positions within the embedding space based on their syntactical and semantical differences. However, the vector between the positions of ‘man’ and ‘woman’ is often similar to the vector between ‘king’ and ‘queen’. This is because semantically, a man is to a woman what a king is to a queen. The vector generated from the position of a word in the embedding space is called a word embedding.

Besides them being pre-trained, ChatGPT and all the other models based on GPT architecture are transformer type models. Transformers are a type of network architecture that was first presented in 2017 [57]. Although there is no clear definition of LLMs, many of them are transformers trained on hundreds of billions parameters [61]. Since transformers are the foundation of many LLMs including GPT models, a rough understanding of them might help with understanding certain properties of such LLMs.

Before transformers were introduced, deep neural networks such as recurrent- and convolutional neural networks were often used in sequence transduction models used for NLP tasks [57]. Unlike those deep neural networks, transformers are able to process the entire input sentence at once. The structure of a transformer is that of an encoder and a decoder, shown left and right respectively in figure 4.3. The encoder of a transformer take the whole input sentence by using a positional encoder on the input embedding [57]. Using a positional encoder on the input embedding gives positional information to the word embeddings which will be referred to as positional vectors.

After the positional vectors are obtained for all the words in the input sentence, they are put into the multi-head attention block. Without going into technical details, the attention block generates an attention vector for each word. Attention could be interpreted as the most important word to focus on. Generally, the most attention when looking at a word is given to that word itself. The attention to other words, however, will be different per word. As

an example, in the sentence ‘The blue moon’, the attention of the word ‘The’ would focus on itself the most, but it would have more attention on the word ‘moon’ than on the word ‘blue’. Intuitively, this makes sense, since ‘The moon’ is something that can stand on its own in a sentence, but ‘The blue’ means cannot.

Once attention vectors have been generated, they are put into a feed-forward neural network (FFN). The FFN is necessary for the transformer to learn. The learning is made possible since textual patterns are put into the network alongside the attention vectors [19]. In this process, the FFN also turns the output from the multi-head attention layer into encoded vectors that can be accepted by the decoder. For the original paper two linear transformations and a ReLU function were used [57, p. 5]. The attention vectors can be processed independently in the FFN. Once all the vectors are processed by the FFN, they can be fed to the decoder.

The decoder starts out with a similar process as the encoder. At the start, the output gets embedded and a positional encoder goes on top of that to get positional vectors. Now these positional vectors go into an attention layer, but this first attention layer in the decoder is different from the one in the encoder. All the words could be given simultaneously in the encoder layer, but in the decoder only words that occurred prior in a sentence get processed. This is done by giving the full sentence to the attention layer, but every word after the current position is set to 0 [57]. Setting the words to 0 is called masking, which is the reason that this layer is called the masked multi-head attention layer. Masking is necessary for learning, since the model needs to predict the next word at every step. If masking did not take place, the prediction of the third word would be influenced by the first and second word, but their positions would in turn be influenced by the third word [57]. This would mean that the third word would influence its own prediction and that could cause problems akin to a catch 22.

After the attention vectors from the masked multi-head attention layer are received, they are put into a multi-head attention layer together with the embedded vectors from the encoder. This attention layer essentially looks to what extent each word in the input and the output is related to one another. To relate this to ChatGPT, a prompt such as “How are you?” could be the input and “I am good.” could be the output. In this second attention layer of the decoder, it would probably find that ‘you’ and ‘I’ are closely related, ‘are’ and ‘am’ are closely related, and that ‘How’ and ‘good’ are closely related, since those words from the answer are direct reflections of the related words in the question. The attention vectors that come out of this attention layer contain information both on how each word from the input and the output relates to all the words from both the input and the output.

These attention vectors are then given to another FFN that functions does essentially the same things as the FFN of the encoder. From there, the resulting vectors get put into a linear transformation, which expands the dimension of the vectors to match all the words in the output language. A probability function is then used to predict what word is the most likely next word.

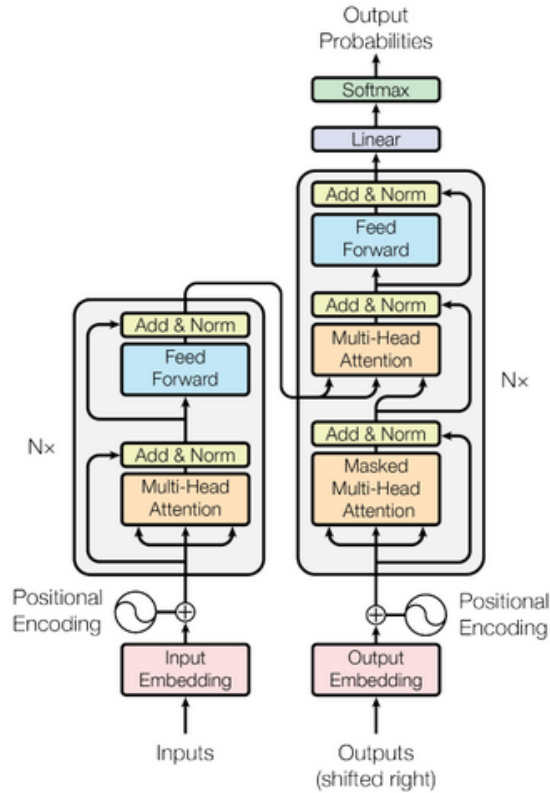


Figure 4.3: Visualization of a transformer [57]

All these steps of the decoder are done for one word at a time, meaning that first a completely masked output embedding is given to the masked multi-head attention layer to predict the first word. After the first word is predicted, everything except the first word gets masked, then everything except the first two words get masked etc. This goes on until every word in the output sentence is predicted.

Besides ChatGPT there are also many other GPT models. The first GPT models GPT and GPT-2 are two examples, but there are also GPT-3 based model besides ChatGPT. An example of such a model is OpenAI Codex, which has been used in some of the mathematical applications that will be highlighted later in this chapter [14]. There are even newer models, such as GPT-3.5 and the most recently developed GPT model, GPT-4. Although there are several LLMs that could be covered in this thesis, GPT models are interesting for several reasons.

First of all, ChatGPT is a GPT model. ChatGPT is a free natural language

processing tool that can be easily accessed online, making it the most popular LLM in the world [27]. The popularity makes research on it relevant for real life situations and its accessibility makes it great to do experiments with. There are other options besides ChatGPT that have performed better when it comes to mathematical problem solving, but these are not always freely available [14] [58].

ChatGPT has a certain amount of randomness in it. This randomness can be controlled to some extent with a set of parameters [41]. The main parameters that dictate the amount of randomness within a conversation are the ‘temperature’ parameter, ‘top_k’ and ‘top_p’. Temperature is a setting that changes the randomness from ChatGPT by restricting or allowing creativity. The temperature has a value between 0 and 1, where a higher temperature gives ChatGPT more creative freedom and where a low temperature restricts the creativity of responses [41]. On the openai website it is stated that a temperature of 0 will make the responses almost deterministic, but that the API has a slight amount of inherent randomness that cannot be removed with parameter settings [41].

Top_k and top_p are both influencing the prediction mechanism of ChatGPT. when the model tries to predict the next word, it has a list of words that all have a probability assigned to them. It could be the case that the list of possible next words contains 6 words. Setting top_k to 3 would mean that ChatGPT would reduce this list to the three words with the highest probability of occurring next. In theory this should mean that setting top_k to 1 would make ChatGPT only predict the most likely response.

Similar to how top_k works, top_p looks at the list generated by top_k and limits the selection to only the words whose cumulative prediction meet a certain threshold. As an example, top_p gets a list of 6 words ordered by highest to lowest probability and the probabilities add up to 1. This list could be [0.45, 0.25, 0.15, 0.08, 0.05, 0.02]. By saying top_p is 0.7, only the first two responses would be considered since $0.45 + 0.25 = 0.7$. Setting top_p to 0.71 would include the third word as well. In practice this means that lowering top_p will reduce the variety in word choices, much like lowering top_k does. Setting top_k = 1 or top_p = 0 should limit word choices to the most likely response and, by extension, reduce the randomness in the output.

Second, GPT-3 is designed to in a way that makes it easy to give it a large range of prompts and to ask it to generate texts. This is a big reason for the popularity of ChatGPT. On top of that, the ability to be able to interact with it makes it great to try measure mathematical understanding. Not all LLMs are able to process the different types of prompts that ChatGPT can process. ChatGPT is able to respond to texts in a variety of languages, it is able to answer questions, it is able to finish sentences, and it is capable of writing texts according to specific requests. While there are language models that can perform some of these tasks, not many of these models can perform all of these tasks, and definitely not with the accuracy of GPT-3 models.

An example that illustrates this is another well-known LLM developed by Google called BERT. BERT is capable of handling prompts that involve search requests, Q&A prompts, and sentiment analysis. While it might be good at these

tasks, BERT is not good at text generation. Text generation is one function that has made ChatGPT so popular, and it is also a feature that might prove helpful to determine mathematical understanding in an LLM.

Lastly, GPT models have already been used in mathematical applications [14]. Since GPT models are a type of language model, the application of these models is limited to text based exercises. If a graph would need to be plotted, a GPT model would not be able to do so. The tasks that GPT has been used for in mathematics can be roughly categorized into two categories: producing questions and solving them.

Mathematical applications

When it comes to mathematical problem solving, so far GPT has had some success, but it is far from the most successful model. A study found that OpenAI Codex, a descendant from GPT-3, was able to outperform GPT on solving math olympiad questions. The GPT-3 model had a 30% accuracy, while the OpenAI Codex was able to solve questions with about 81% accuracy [14]. OpenAI Codex is not the only language model that outperformed GPT-3 based models when it comes to solving mathematical word problems. A sequence-to-sequence model based on LSTM was able to solve two of the largest datasets of mathematical word problems with nearly 67% accuracy. Even though some of the other LLMs were able to outperform GPT-3 models, these models are not readily available for public use, leaving GPT as a more accessible option for doing research.

Some of the recent attempts of having LLMs solve mathematics problems has shown several limitations that these models have when it comes to solving mathematical problems [14]. One of the limitations that GPT-3 and Codex have shown is in its ability to solve mathematics questions in a ‘human’ way. In the research that was done with GPT and OpenAI Codex, it was deemed necessary to ask the models to answer the questions using python and several of the mathematics tools that the programming language contains.

One of the tools used to generate answers was Matplotlib, which is a library of python that is used to visualize data with graphs. This might be akin to a mathematics student using their calculator to plot a graph. Plotting a graph using a calculator as part of answering a question does not show a lack of understanding of the subject. Generating these graphs might be something a student might do, so this usage of python for solving questions does not necessarily deviate from what a human might do.

However, the language models were asked to solve all the questions using Python. This means that the questions were solved by writing a Python script. Since Python code is not something that everybody could assess, the AI was asked to explain this python code. An example that illustrates clearly how the explanation given is very inhuman is the following:

The question is to give a 2x2 matrix that rotates all vectors 45 degrees. If someone knows how to solve this question they would probably take a rotation

matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and try to find a θ that rotates simple vectors such as $[0,1]$ and $[1,0]$ by 45 degrees. There might be other ways that a person might solve this, but this is one way that is human. When the AI is tasked to explain how they did this, the code works in a similar way. The explanation of the code, however, does not read like a human did it. The explanation provided was “We define a function called `rotation_matrix` which takes in a parameter `theta`. We return a 2D numpy array which is the rotation matrix. We call with function with `theta = pi/4`. We print the result [14, p. 81].” The fact that it explains every step in such a rigid way is not necessarily a problem, but how it found $\theta = \pi/4$ is not explained in this explanation. It might be true, but if a person has to solve this question they would have to explain how they found this. While this might not be the biggest problem, it does mean that AI would have to be assessed differently than a human, since calculations like that are not always going to be correct, which means they are also not considered self-evident for most human agents.

Another limitation that the AI models have run into is the fact that the questions regularly had to be simplified. One of the ways that questions had to be changed was by turning them into prompts that asked for a solution programmed in python. This type of modification of questions does not necessarily reduce the level of mathematical understanding required to answer the question. However, the other types of modifications that were done resulted in a simplified version of the original question. The questions were changed by making longer sentences shorter or by removing redundant information [14, p. 4]. The theoretical approach to understanding showed that the ability to filter relevant information is an important aspect of understanding, and in the mathematical assessment scale some of the items listed are about the ability to filter information or the ability to apply relevant mathematical skills in real life situations. If redundant information is removed, the ability to filter information is either no longer required or required to a lesser extent. When sentences are made shorter, it could also lead to certain information being lost. These two types of question modifications do make it harder to determine understanding in LLMs, especially when results get compared to those of human agents.

The other limitations that were found for LLMs when solving math questions had to do with the types of questions. Questions that contained information that was expressed in pictures (such as graphs), tables or any other form that was not text could not be solved. When a question needed to be answered in the form of a proof, the LLMs generally were unable to provide one. Lastly, computers were unable to work with very large primes. This was cited as a limitation of computers, but this would also be a limitation for many people, so maybe this should not be considered a limitation for LLMs specifically.

The other task that GPT models (including OpenAI Codex) have been used for is generating mathematical word problems [14, p. 8]. When it comes to generating mathematical problems, getting an accurate result is not necessarily a requirement. If it is important that the answer to the question is a rounded number, then it is important to be able to solve the question. Otherwise gener-

ating questions does not require the ability to solve said question.

Questions were generated by giving the AI a set of questions and asking it to generate the next one [14]. This task was also given to human participants. The AI generated questions were then compared to those generated by the human participants and a similarity score was assigned. How similar the human generated questions and the AI generated questions were varied per subject. For the subject of algebra, the AI generated question had a 0.99 similarity with the most similar human generated question, whereas the AI generated question for counting and probability only had 0.12 similarity with the most similar human generated question. This similarity is not necessarily important, as many of the questions generated by the language model should be solvable with standard equations. However, for the subject of linear algebra, Codex generated a question that asked to write a Matlab code, whereas the most similar human generated question for the subject asked a question that should be answered with a mathematical proof. When questions are produced that ask for a solution that is different from how humans would solve those types of problems, the question generated could be seen as problematic.

4.1.2 LLMs and understanding

The questions about mathematical understanding that were raised for language models in general also apply for GPT models. Since the models are based on deep learning it is hard to determine what exactly happens under the hood. It is known that GPT models extrapolate patterns from text to predict responses based on prompts, but which patterns they use can only be estimated.

Since ChatGPT is a freely available GPT language model it is considered the best model to work with for this thesis. It can be prompted with mathematics questions to see if they truly understand mathematics. One of the aspects that might suggest a lack of mathematical understanding of ChatGPT is the fact that the model might be able to answer very simple addition questions as "I have ten dollars. I spend eight dollars, but then find two dollars on the street. How many dollars do I have now?" correctly, but when a bunch of irrelevant information is added to the prompt, such as "I have eight cookies, five apples, one banana and ten dollars." ChatGPT regularly gives incorrect answers. This demonstrates a lack of ability to filter which information in the prompt is relevant to the exercise, which is one of the properties associated with mathematical understanding.

Something that was already discussed before is the fact that sometimes questions have to be modified for the LLM to be able to answer it. One of the tests that was proposed for measuring mathematical understanding in AI was one that was similar to a Turing Test [34]. Given that the questions were modified for the computer in ways that changed the information contained in the question means that this type of test could no longer be done. A question might be formulated in a way that requires students to filter information, but because the AI is unable to do so the question gets modified until they can answer it. This inability to separate the relevant information from the irrelevant information in questions suggests that there is a lack of mathematical understanding in

these LLMs, both according to the theoretical approaches and the mathematical assessment scale [21].

One of the intrinsic properties of LLMs is that they are trained by large quantities of text based data. This data needs to be written by someone, which means that human errors can be made in the texts that end up in the training data. If this happens patterns might be learned that would indicate a lack of understanding. Similarly, it is possible that instead of mistakes many people write down that they do not know or understanding something. If the training data is not carefully selected, there is a realistic risk that the LLM notices that the most common answer to certain types of questions is “I don’t know”. For obvious reasons, this is not a desirable outcome. It can, however, be difficult to collect large quantities of data while maintaining a high quality of said data.

It is possible that randomness is an inherent property of LLMs, as ChatGPT retains a certain level of it when all of the randomness parameters are set to minimize randomness. The temperature parameter, top_p and top_k all affect randomness in different ways, but it seems that the initial response in a chat can still differ when those parameters are set to 0. Since ChatGPT keeps a certain level of randomness for all parameter settings, it could be the case that this happens for all LLMs. Depending on how randomness affects the answers to mathematical questions, this could lead to potential inconsistencies in its answers. This could become clearer in the next chapter.

Simply speculating on what the limits of LLMs in mathematics are can be interesting, but at some point those speculations need to be tested. In the section on understanding a way to measure mathematical understanding in LLMs was proposed. This method combined the criteria from the Mathematical Assessment Scale and it took some elements from the Turing Test. As mentioned earlier, the goal of this test would be to determine if an LLM possesses a certain level of mathematical understanding. This means that it cannot definitively say whether LLMs will be able to attain mathematical understanding if it concludes that the current one does not.

Besides the limitations of the test itself, the resources and time were limited for this thesis, so instead a small portion of the proposed test was done. Since the test does not say anything conclusively about the limitations of AI, the test that was done was mainly used as a proof of concept of the proposed test. The test came down to three math questions, one larger question containing a larger body of text and two smaller questions. Although the main goal was testing the proposed method for assessing of LLMs, another goal was seeing if ChatGPT was able to show a certain level of understanding. This is why the performance of ChatGPT was compared with that of the students. This experiment will be discussed in detail in the next chapter.

Chapter 5

Experiment

5.1 Experiment setup

5.1.1 Participants

Two groups of high school students were asked to do three math exercises. Two of these exercises were the same for both groups. These two exercises were brief exercises that test whether students could do number manipulation that was not asked in the other exercise. The other exercise was a larger exercise. This larger exercise is called “Grotere Windmolens” and it was taken from a VWO Wiskunde A central exam [55]. The last question of VWO Wiskunde A central exams has been a ‘research question’ for the last couple of years. This means that it is a question that is supposed to be something the students can make with the skills that they have gained up during all the year of doing mathematics in high school, but it is not a type of the question that they have practised before.

Both groups of participants did mathematics at a VWO 5 Wiskunde B level.¹ Although students were taught the tools to solve the exam question, it was question covered a situation that the students were unfamiliar with. The question involved economics of a situation that could happen in real life. Since Grotere Windmolens covered economics, a field other than mathematics, items 11 and 17 of the mathematical assessment scale could be used to evaluate the answers.

Since both groups were in their fifth year of high school, the participants were all 16 to 17 years old. Both groups of participants contained 11 male students. The group that did version A contained 7 female students and the group that did version B contained 6 female students. The grade point averages in mathematics for the groups were 6.6 and 6.3 respectively. These grades were

¹In the Netherlands VWO is considered the highest education level in high school, and Wiskunde B is generally considered the hardest type of mathematics in the standard curriculum. Wiskunde B focuses more on theoretical questions, whereas Wiskunde A puts more focus on statistics and probability calculations. Even though the question was taken from a VWO Wiskunde A exam and the research participants were all Wiskunde B students, all of them were taught the skills to solve the exercises over their high school career.

based on the tests taken this year and are on a scale from 1 to 10, with 5.5 being the lowest passing grade.

The students were given the option to participate in the research during one of their classes. They could opt out and work on homework instead, but all students in both classes opted to participate in the experiment. They were given as much time as they have in a class, which is about 45 minutes, to do all of the exercises. A teacher was present to watch for any potential cheating. Since the large question required the use of a graphical calculator while the other two questions expressly stated no calculator should be used the students were given exercise 1 first. If a student was done with that exercise they could raise their hand and, after putting away their calculator, the teacher would hand them the remaining two exercises.

5.1.2 Question design

The questions that were chosen for this experiment were chosen for two main reasons. The ‘Grotere Windmolens’ question was chosen because it covered many of the items from the mathematical assessment scale, as will be discussed later in this chapter. Grotere Windmolens was an exam question that was supposed to take VWO 6 Wiskunde A students roughly 20 minutes, but since the Wiskunde B were in the fifth year and they had no experience with these types of questions, it was estimated that they would need up to 30 minutes to solve that question. Since the time that the students were given for the experiment was roughly 45 minutes, two more questions were asked that could be answered in the remaining 15 minutes.

A part of the information in the Grotere Windmolens was conveyed through an image. Since ChatGPT can only take text as input, the information in this image had to be put into words. However, even with the information in the image translated into words, ChatGPT was still unable to solve the exercise. Since answering this question requires multiple items from the Mathematical Assessment Scale, it seemed like a great question to test mathematical understanding. When giving ChatGPT only the core information, however, it was able to go through the correct steps, albeit with a slight miscalculation due to the fact that it does not have the ability to use a graphical calculator when necessary.

To ensure that this question could be used for assessing mathematical understanding in both ChatGPT and the participants, a second version of the question was made. This version was called version B and the original was called version A. For both versions A and B, the information from the image was expressed into words.

Version B was made by going through several iterations. The first iteration contained the main formula with the values of each variable entered, a sentence that contained the task that needed to be solved, and finally a question that asked ChatGPT to solve the problem. When ChatGPT was able to answer that question correctly, a more difficult version of the question was asked. Instead of entering the values of each variable directly into the formula, one piece of the

original question was restored. That meant that ChatGPT would have to filter the relevant information for one of the variables out of the text and enter it into the formula on its own. Whenever ChatGPT was able to handle the extra layer of abstraction, another was restored. After some steps it turned out that ChatGPT was unable to handle the words “minimaal” (meaning ‘minimally’) and “minstens” (meaning ‘at least’).

After reverse engineering the variables that were entered into the formula back to the original question, version B ended up being fairly similar to version A. Both versions contained all of the information necessary to solve the exercise, but there were two small differences. In version A it says that a new windmill park has to generate ‘at least’ 30 MW of power and it asks for the minimal cost of this park, whereas version B simply says the new windmill park has to generate 30MW and it asks how much this will cost. The second difference lies in the fact that a formula is given in the original version that calculates the power generated by a windmill for the windmills of both the old and the new park. However, in version B it is stated that this formula applies to the new windmills, since using it on the old windmills was unnecessary and yielded poor results from ChatGPT.

As an extra note, to ensure that the B version was not significantly different from the A version, the amount of mistakes regarding the values of P and g for both versions were compared. Finding the values for P and g were compared, since these steps required the students to filter information from the text. For both versions similar mistakes were made, and they were made in similar quantities. The students that did version A got the value of P correct 10 times and the students that did version B got it right 8 times. In that case version A scored better, more students that did version B got it right that the value of g needs to be larger than 1. This, combined with the fact that the average score for each version was similar, suggests that both versions were equally difficult. On top of that, it does not seem that any of the steps became easier due to the rephrasing of the original question. The final step also required the ability to filter information. Most students that reached this step got similar answers, so this further reinforced the idea that both versions were equally difficult.

The extra questions were chosen to fill the remaining 15 minutes and to provide extra data, but there were also reasons why these two questions in particular were chosen. First, both questions were short and they consisted of only text. This meant that they did not have to be altered in any way. Second, they were both questions that were different from what the participants were used to, which meant that more items of the mathematical assessment scale were relevant for these questions. Finally, there were multiple ways to solve each of them. This allowed for the testing of item 15 from the mathematical assessment scale in both questions.

5.1.3 Evaluation metrics

To determine if both versions of the question were equally difficult, the answers to Grotere Windmolens from the students were graded according to the original

correction model. Since there were 18 students that did version A and 17 that did version B, all the students were labeled 1A through 18A and 1B through 17B. These results can be seen in table 5.1.

Point counts exercise "Grotere Windmolens"		
Student number	Points: Version A	Points: Version B
1	7	2
2	3	2
3	3	2
4	7	1
5	2	2
6	3	2
7	1	3
8	6	2
9	0	3
10	2	5
11	1	2
12	1	3
13	5	3
14	0	1
15	1	0
16	1	2
17	0	7
18	7	-
Total	50	42
Average	2.78	2.47

Table 5.1: Student scores on Grotere Windmolens according to the original answer model of the exam

For the different questions, different items from the mathematical assessment scale were considered relevant. The items that were assessed with Grotere Windmolens were item 1, item 2, item 4, item 6, item 7, item 11, item 13, item 14, item 17, item 20 and item 21. The contents of the items are listed below for clarity.

There was some overlap with the items that Grotere Windmolens and the extra questions tested, but the extra questions also tested an item that the first exercise did not. Item 15 measures understanding by looking at the usefulness of a solution. Since the extra questions could be solved in different ways, finding an efficient solution would demonstrate the ability to recognize the usefulness of different solutions.

- Item 1: Teachers should assess whether a student understands the lesson according to their ability to produce mathematical knowledge

- Item 2: A student's understanding of mathematics should be assessed according to their ability to transfer their mathematical knowledge to daily life problems
- Item 3: A student's understanding of the mathematics lesson should be assessed according to their ability to arrive at mathematical generalizations based on specifics
- Item 4: Whether or not a student understands mathematics should be assessed according to their ability to make logical inferences from their existing mathematical knowledge
- Item 5: A student's ability to switch between mathematical topics should be considered when assessing whether or not they understand mathematics
- Item 6: A student's understanding of the mathematics lesson should be assessed according to their ability to synthesize their existing knowledge
- Item 7: Assessing whether or not a student understands mathematics should be done according to their ability to transfer existing knowledge to new situations
- Item 8: Whether a student understands the mathematics lesson or not should be assessed according to their ability to bring various solution paths to problem situations in mathematics
- Item 9: Teachers should assess a student's understanding of the mathematics lesson according to the student's ability to develop mathematical knowledge
- Item 10: A student's understanding of the mathematics lesson should be assessed according to their ability to transfer their knowledge to different situations in mathematics
- Item 11: Teachers should assess whether or not a student understands the mathematics lesson according to the student's ability to produce solutions to questions involving other disciplines (e.g., science, geography)
- Item 12: Teachers should assess whether or not a student understands the mathematics lesson according to the student's ability to explore mathematical knowledge
- Item 13: Teachers should assess whether or not a student understands the mathematics lesson according to the student's ability to think analytically
- Item 14: Teachers should assess a student's understanding of the mathematics lesson according to the student's knowledge of mathematical methods

- Item 15: A student's decision on the usefulness of solution paths to a problem should be taken into account when assessing the student's understanding of mathematics
- Item 16: A student's understanding of the mathematics lesson should be assessed according to their ability to associate mathematical knowledge with other mathematical knowledge
- Item 17: Whether students understand the mathematics lesson should be assessed according to their ability to synthesize their mathematical knowledge with knowledge in other disciplines (e.g., science, geography)
- Item 18: A student's understanding of the mathematics lesson should be assessed according to their ability to pose new problem situations
- Item 19: A student's awareness that a solution path found in mathematics is also valid for similar situations should be considered when assessing their understanding of the lesson
- Item 20: A student's understanding of the mathematics lesson should be assessed according to their knowledge of the defined steps of operations
- Item 21: A student's understanding of the mathematics lesson should be assessed according to their ability to use algorithms

Some of the arguments from the mathematical scale were phrased in a way that left them open to interpretation. This was the case for items 3, 9, 12, 16, and 19. To avoid having faulty results due to any wrong interpretations, these items were interpreted such that they fell outside of the scope of this experiment. The other items that were considered to fall outside the scope of the experiment were items 8 and 18. These items assess mathematical understanding according to the ability to use several solution paths for the same problem and the ability to pose new problem situations respectively, which are not skills that are required to solve any of the three questions in the experiment.

As can be found in the list above, item 5 states that mathematical understanding should be assessed according to the ability to switch between mathematical topics. Similarly, item 10 assesses mathematical understanding by looking at the ability to apply mathematical knowledge to different situations in mathematics. Since *Grotere Windmolens* was a question that the students were unfamiliar with, it could be argued that these items could be used for the evaluation. However, the question is also about economics, a field other than mathematics. It could not be the case that the question was both a different area of mathematics and a field outside of mathematics. Since the question was definitely one about economics, it could not be a question in a different field of mathematics. Items 5 and 10 were, therefore, considered out of the scope for *Grotere Windmolens*.

Now that it is clear which items are relevant for the assessment of the different questions, we can look at the way that these items are used for evaluation.

The answers to the questions can be broken into different steps. The solution to Grotere Windmolens can be divided into roughly 4 or 5 steps. Each step of the answer requires different skills, which means that different items apply to different steps of a solution. To illustrate, for Grotere Windmolens the first step required the students to retrieve the values of P and g from the text. Since finding P required the filtering of information from an economics question, items 11 and 17 were considered relevant for this step. Although other items were relevant for the question overall, the skills that were evaluated by these items were not relevant for finding P . Therefore, these other items were not relevant for the step of finding P .

Items could apply to multiple steps. Whenever a step was answered correctly, a point was awarded to all of the items that were relevant for that step. This meant that finding the correct value of P and entering it into the formula would award one point to items 11 and 17. At that point during the evaluation a student would have 1/1 points. However, finding P was not the only step of the solution. If items 11 and 17 were relevant for a later step that the student then answered incorrectly, they would have 1/2 points for items 11 and 17.

While finding the value of P only tested two items, finding the correct value for g tested more skills. There were two main mistakes that were made when trying to find the value of g . Since g represented a growth of 0.68% per meter of axle height, the correct value of g was 1.0068, but many students got 0.0068, 1.68 or 0.68. While finding g is one step, points could be earned for getting .0068 after the decimal point and points could be earned for getting saying that the value of g was larger than 1.

At this point in their high school career the students had learned how a percentage should be translated into a factor that can be entered into an equation, such as the one from Grotere Windmolens. Knowing that a percentage of 0.68% translates to 0.0068, therefore, shows understanding according to items 1, 2, 6, 7, 14 and 17, since these items are about assessing understanding according to the use of existing knowledge and applying said knowledge to a discipline other than mathematics. Since students are also taught that a growth factor always consists of a value larger than 1, a point was also awarded to items 1, 2, 6, 7, 14 and 17 if the solution contained a value of g that was larger than 1.

However, even if a student did not know that a growth of 0.68% translated to a growth factor 1.0068 in the formula, they could have deduced that multiplying by something smaller than 1 would cause shrinking instead of growth. Items 4 and 13 cover the ability to make logical inferences and the ability think analytically, so these items were considered relevant for the growth factor being bigger or smaller than 1. Having too few or too many zeros after the decimal point did not count towards these two items. Overall, this meant that getting a value larger than one for g gives a point for items 1, 2, 4, 6, 7, 13, 14 and 17. As a result, 2 points could be earned here for items 1, 2, 6, 7, 14 and 17 and one point for items 4 and 13.

Entering the values of P and g resulted in an equation with h and D as the only remaining variables. Since it was given that $h = 0.9D$, one of the variables in the equation could be substituted by the other. Originally, this step

was considered so simple that it did not yield any points, but ChatGPT got this step wrong. Since the knowledge that $h = 0.9D$ means that $D = h/0.9$ was common among students, making this mistake was considered a lack of understanding according to items 1 and 6.

Having a one variable equation meant that the third step could be taken, which was calculating the value of said variable by inputting both ends of the equation into a graphical calculator and finding the value using intersect. With this step the value of the remaining variable h or D became known. Using the intersect function on a graphical calculator is a step that students were taught to use when they are trying to find a variable in a function that is too difficult for them to solve on their own. There was no reasoning involved in this step and the context of the question was irrelevant here, but students did have to draw on knowledge of taking certain steps in an algorithm. This made items 1, 6, 14, 20 and 21 relevant here.

Since the price of one windmill is dependant on the value of h , an extra step needed to be taken if D ended up as the only variable in the equation. This potential fourth step would be getting the value of h using $h = 0.9D$ and filling in the value of D . Since this step could be skipped by putting h as the variable of the equation instead of D , so no items could be assigned to this step.

Finally, the cost of the windmill park needed to be calculated. Since the minimal amount of windmills needed is 10 and the cost is 25000 euros per meter of axle height, the final step required the student to combine this information to create the final cost $10 \cdot 25000$. This required working with contextual information, but besides that, this step was too straightforward to consider any items other than item 2, item 11 and item 17 relevant.

Besides the steps of the Grotere Windmolens solution, there is also the “Extra vragen” part. While this part introduced item 15, it also touched on some of these other items. Both of the questions in Extra vragen required the skills evaluated by item 15, so two points could be received for this item. Depending on the answer used for question 1.1, items 1, 4, 6, 7 and 13 could be tested by this question. For question 2.2 the items 1, 6, 13, 14, 20 and 21 were considered relevant.

However, as will be covered in the results section, the Extra vragen section consisted of questions that could be answered in various ways. Because of this, it was difficult to determine which step was worth how many points. Most students made similar mistakes, which meant that the evaluation remained consistent overall. The solution that ChatGPT gave to question 2.2 was mostly correct. However, it also showed a level of misunderstanding of what real numbers are that was beyond what any student showed. Therefore, it was counted as another mistake that was not counted against any of the students. This led to another inconsistency in the point totals.

Besides using the mathematical assessment scale, the computer-assisted scientific method could potentially give additional information about understanding in ChatGPT. However, the computer-assisted scientific method ended up not being used for two reasons. First, the decision to do this experiment was done after the thesis was already being worked on for several months, leaving

little time to set up this experiment. The second reason is caused mainly due to this lack of time, since the second reason is of the fact that there were not enough participants available to be a judge for the computer-assisted scientific method. Therefore, there are no results on the computer-assisted scientific method that will be covered in this chapter.

5.2 Experiment results

5.2.1 Assessment ChatGPT

Grotere Windmolens

Since it is clear how the items from the mathematical assessment scale were used to evaluate understanding in the experiment, the responses from both ChatGPT and the students will be assessed in this section. This process will be started by looking at ChatGPT.

When given version B of Grotere Windmolens, ChatGPT was working in Dutch mode. Since the question was in Dutch, this was necessary, but it also means that the original answer was in Dutch. The original answer in Dutch can be found in the appendix A.8. The response was translated into English as follows:

To calculate the investment of the construction of the new windmills, we will first have to find the required axle height and rotor diameter to calculate the desired amount of power the new windmills would have to generate.

We know that the old windmill park exists of 40 windmills that each generate 0,75 MW of power, so the total power generated by the old park is $40 \cdot 0.75 = 30\text{MW}$.

Next we have to calculate the power that each windmill in the new park will have to generate. Since the new park also generates 30MW and has 10 windmills, the power generated by each windmill is $30\text{MW} / 10 = 3\text{MW}$.

We can now use the formula to find the required axle height and rotor diameter. The formula is $P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$.

Entering the known values gives: $3 = 2.21 \cdot 10^{-4} \cdot 1.0068^h \cdot D^2$.

Since $h = 0.9D$ we can substitute this: $3 = 2.21 \cdot 10^{-4} \cdot 1.0068^h \cdot (0.9h)^2$

Now we can solve this equation to find h . This is best done numerically using a graphical calculator or using software like Python.

Using a numerical approximation we find that $h \approx 80.154$ meter.

Now we can calculate the investment by multiplying the amount of meters of axle height with the cost per meter. In this case it is $80.154 \cdot 25000 = 2003850$ euro.

So the minimal investment that will have to be done to build the new windmills is approximately 2 million euros.

Without evaluation the understanding of ChatGPT, it is clear that some mistakes were made in this answer. What ChatGPT did well was realizing that the new windmill park has to generate 30MW and concluding that, therefore, one windmill will have to generate 3MW. It also realized that the growth of 0.68% translates to a growth factor of 1.0068, something that many students did wrong. Furthermore, ChatGPT realized that using $h = 0.9D$ allows for substitution resulting in a formula with only one unknown variable. However, it made its first mistake here. From $h = 0.9D$ it follows that $D = h/0.9$, but ChatGPT substituted $D = 0.9h$ in the formula. The value of h that ChatGPT found was incorrect due to the mistake it made, but it continued by correctly assessing that the method to determine the value of h is using a graphical calculator (or Python software).

ChatGPT then used the value of h to calculate the cost of one windmill in the new park. It did this by multiplying the height of the windmill with the cost per meter of axle height, which is the correct method. However, it seemed to have forgotten the original goal of the question. The goal was to calculate the cost of the entire windmill park, but ChatGPT just calculated the cost of one windmill and said that that was the total cost that we were looking for.

When using the correction model from the original exam, this answer from ChatGPT would get 4 or 5 out of 7 points [54]. This score is better than what most of the students got, but the original correction model is not the one that was used to assess mathematical understanding. The mathematical assessment scale should be able to give a better idea of the actual understanding of the mathematical content.

In the answer of the B version, ChatGPT did the first step well. It correctly came to the conclusion that $P = 3$ and $g = 1.0068$ and then entered these values into the equation. However, it made its first mistake in the second step. In the question it was stated that $h = 0.9D$, but then ChatGPT turned the D^2 in the formula into $(0.9h)^2$. Obviously this should be $(\frac{1}{9}h)^2$, but for some reason it filled in $h = D$ instead of $h = 0.9D$. Generally, when a human makes this type of mistake it gets attributed to forgetting that the $h = 0.9D$ means that $D = \frac{h}{0.9} = 1\frac{1}{9}h$. Therefore, this type of mistake most likely would not have been made due to a lack of understanding if was made by a human. Nonetheless, the mistake could also be attributed to a lack of knowledge or an inability to apply this knowledge. The first interpretation is more charitable and the latter is more strict regarding understanding as assessed by items 1 and 6. Item 1 evaluates understanding according to the ability to produce mathematical knowledge and item 6 according to the ability to synthesize existing knowledge. The way that these items are phrased could apply to the mistake ChatGPT made here. To

avoid accidentally attributing too much understanding to ChatGPT, the less lenient approach was used. None of the students, however, made the same mistake as ChatGPT. Although this mistake counted negatively towards the score of ChatGPT, getting this right was not counted positively towards the scores of the students.

Regardless of any previous mistakes, the next step was to intersect the two points on a graphical calculator. Since ChatGPT is not a student with his own calculator, it used some built in calculator. In this case it said that the best way to solve this was ‘numerically using a graphical calculator or software like Python’. This showed the understanding that is connected to this step. ChatGPT, however, did not get the correct value for h , even when considering the mistake of using $(0.9h)^2$ instead of $(1\frac{1}{9}h)^2$. This type of calculation error from a student would be caused by inputting the incorrect formula into their calculator, but that has nothing to do with understanding. Even though it could be argued that ChatGPT did this calculation on its own, this mistake did not get counted against its understanding since it would not be counted for a human.

After this ChatGPT just had to calculate the cost of the windmill park. It made the mistake of using the cost of a single windmill instead of the cost of all the windmills in the park. Therefore, it ended up with a final cost of 2 million instead of 20 million. This shows an inability of ChatGPT to filter information well. It also shows that ChatGPT struggles with placing the calculations it had to do within the context of a real life situation. This is why ChatGPT did not receive points for items 11 and 17 here.

Extra vragen

Besides the answer to Grotere Windmolens, ChatGPT also tried to answer the remaining two questions. Question 1.1 only requires one step to solve, but there are several different ways to approach this question. The goal is to get an answer that makes it clear which number is bigger, since humans cannot say which number is bigger just by looking at 2^{100} and 3^{75} . ChatGPT said that the exponent of 2^{100} is larger than the exponent of 3^{75} , so therefore 2^{100} is bigger. This answer is incorrect, which caused meant that ChatGPT did not receive any points for this question.

Answering question 2.2 went much better for ChatGPT, as it managed to get all the correct solutions for x . However, the question asked for all the real solutions and that is where ChatGPT said some incorrect things. It ended up saying that six of the seven solutions were real without specifying which ones. When asked for clarification it gave a wrong answer. Further inquiry got ChatGPT to say that it believed $\sqrt{2 - \sqrt{6}}$ is a real number, which it followed up with saying that $2 - \sqrt{6}$ is a positive number, namely 0.551. Due to the ability to do all the steps, all the relevant items were counted once. However, the inability to calculate $2 - \sqrt{6}$ counted as an instance of items 1 and 6 being relevant. So items 1 and 6 ended up being counted in two instances, one instance that was positively counted and one that was negatively counted.

Overall this gave ChatGPT the following scores:

Item number	Point total	Explanation point total
1	4/7	Value g correct, intersect step correct, $0.9h$ wrong, extras mostly wrong
2	2/3	Value g correct, final step wrong
4	1/2	Value g correct, question 2.1 wrong
6	3/6	Value g correct, intersect step correct, $0.9h$ wrong, extras mostly wrong
7	2/3	Value g correct, question 2.1 wrong
11	1/2	P correct, final step wrong
13	3/4	Value g correct, question 2.1 wrong, question 2.2 right steps
14	3/4	Value g correct, intersect step correct, question 2.2 wrong answer
15	1/2	Question 2.1 wrong, question 2.2 right steps
17	3/4	P correct, g correct, final step wrong
20	2/2	Intersect step correct, question 2.2 good steps
21	2/2	Intersect step correct, question 2.2 good steps

Table 5.2: The scores received by ChatGPT in the experiment and a brief explanation of the points per item

Multiple points could be added for the relevant items. For each item the amount of points earned was, therefore, divided by the maximum amount of points that could be earned for that item. In the case of item 1, $4/7$ indicates that there were 7 instances where ChatGPT could have received a point for item 1, but it only received 4 of these 7 points. This way of calculating the point total ensured that items that were relevant for more steps did not count too much in the final understanding score. In the mathematical assessment scale each item gets a value of how strongly each item correlates with understanding. For the purposes of this experiment, the point total for each item was multiplied with the corresponding correlation. These totals were added up together to get a final understanding score. For ChatGPT this yielded a final score of $4/7 \cdot 0.373 + 2/3 \cdot 0.613 + 1/2 \cdot 0.661 + 3/6 \cdot 0.329 + 2/3 \cdot 0.732 + 1/2 \cdot 0.438 + 3/4 \cdot 0.566 + 3/4 \cdot 0.613 + 1/2 \cdot 0.467 + 3/4 \cdot 0.474 + 0.561 + 0.436 = 4.294$.

5.2.2 Evaluation students

Grotere Windmolens

Since there were 35 students participating, it would be a bit much to go over the solutions from each student individually. Instead, this section will cover the most common mistakes that the students made, since none of the students got

an answer that deviated from the first solution of the correction model. When evaluating the students, there were some mistakes that occurred regularly. As was mentioned earlier in this chapter, one of the most common mistakes that the students made was the value that they assigned for the growth factor g . Many students mistakenly thought that a growth of 0.68% per meter of axle height would translate to a growth factor of 0.0068, 1.68, 1.068, 0.068 or 0.68. As discussed before, getting a growth factor larger than 1 was considered to gain points for certain items, while getting the amount of zeros after the decimal correct was a point for a different set of items.

Many students made the mistake of using 0.75 as the value for P . This value was the amount of power generated by the old windmills, not the new ones. Using $P = 30$ was also a common mistake. Only 8 students reached the step of using intersect on their graphical calculator, which means that most students did not get points for this step and any steps that came after it. 10 students wrote down that the final amount should be calculated using $10 \cdot 25000 \cdot h$. This means that some students did not know how to get a value for h using intersect, but they demonstrated that they knew what they had to do for the final step. For this reason, there were more students that were able to get points for the final step items than there were students that got points for the intersect items.

The point totals for Grotere Windmolens per item are listed in the table below. Due to the fact that it combines all 35 solutions, the column that gives an explanation for each score is not present here. The scores are not noted separately, because the goal of the experiment is to see if the students outperformed ChatGPT on average. The score for item 1 is 47/105, since there were 3 instances students could have gotten points, but only 47 of these points were earned by the students altogether.

Item number	Cumulative point total
1	47/105
2	49/105
4	19/35
6	47/105
7	39/70
11	28/70
13	19/35
14	47/105
17	67/140
20	8/35
21	8/35

Table 5.3: The scores received by all students for Grotere Windmolens combined

Table 5.3 shows the total amount of points earned for Grotere Windmolens. Before these items are comparable to the results from ChatGPT, however, the

point totals from the Extra Vragen questions need to be added.

Extra vragen

The extra questions were a little harder to assess, especially question 2.1. For Grotere Windmolens, there was one solution path that everybody took, whereas the extra questions had more variety in the answers that were given. Question 2.1 relies more on insight. This means that it could be answered in multiple ways. An answer was considered correct if it demonstrated that the students knew which number was bigger. If the answer was not clear enough, it was considered an educated guess. An example of such an answer would be “3 is larger than 2 so it grows faster, which means 3^{75} is bigger.” While it is technically correct, this explanation clarifies nothing compared to the original question. An example of a clear answer could be “ $3^{75} = (3^3)^{25} = 27^{25}$ and $2^{100} = (2^4)^{25} = 16^{25}$. Since $27 > 16$ it follows that $27^{25} > 16^{25}$.”

Question 2.2 was a question that required less creativity, but there were still several different solutions to this question. Very few students were able to solve this question. Only 4 students showed a calculation that demonstrated that they were able to apply their mathematical knowledge. Of these solutions, two were very inefficiently chosen, which means that these solutions did not demonstrate the ability to distinguish efficient solution paths from inefficient ones. Therefore, only two out of the four correct answers got a point for item 15. One person did try to use an efficient solution, but they miscalculated this attempt. This student got a point for item 13 and item 15, but no points for the other items pertaining to question 2.2.

Other than item 15 and to some extent item 13, the reasons that points were not given were fairly clear cut. Many of the students made mistakes when trying to get rid of the outer exponent in $(x^2 - 2)^2 - 5)^2$. The most common choice at the start was turning $(x^2 - 2)^2$ into $x^4 - 4x^2 + 4$. This would then lead to $(x^4 - 4x^2 + 4 - 5)^2 = (x^4 - 4x^2 - 1)^2 = 1$. Up until this point things were fine, but here many students made the same mistake, which was saying $(x^4 - 4x^2 - 1)^2 = x^8 + 16x^4 + 1$. This was not the only mistake that was made, but many of the other mistakes were of a similar nature; namely that they could not get rid of the outer parentheses. Students that made a mistake of this nature were not awarded any of the points from items 1, 6, 14, 21 and generally also for items 13 and 20. One student was given a point for items 13 and 20 despite not getting points for the other items listed. This was because they used an efficient solution with a ton of calculation errors. This displayed the ability to analyze the problem, select an efficient algorithm, and to then follow that algorithm, albeit with many calculation errors.

The way that a point for 15 was earned was by using an efficient solution. When a student tried to work inside out they would get an equation with x^4 and x^2 , or worse, one with x^8 , x^6 , x^4 , and x^2 . While such a solution is definitely not incorrect, working with exponents this large makes calculations unnecessarily difficult. Especially for the first step, taking the square root left and right avoids any difficult calculation. Students who managed to do this

were given a point for item 15.

Item number	Points Windmolens	Points extra questions	Point total
1	47/105	12/70	59/175
2	49/105	-	49/105
4	19/35	12/70	31/105
6	47/105	8/35	55/140
7	39/70	8/35	47/105
11	28/70	-	28/70
13	19/35	13/70	32/105
14	47/105	4/35	51/140
15	-	3/70	3/70
17	67/140	-	67/140
20	8/35	5/35	13/70
21	8/35	4/35	12/70

Table 5.4: The point totals of the students

This evaluation gave the students a final score of $59/175 \cdot 0.373 + 49/105 \cdot 0.613 + 31/105 \cdot 0.661 + 55/140 \cdot 0.329 + 47/105 \cdot 0.732 + 28/70 \cdot 0.438 + 32/105 \cdot 0.566 + 51/140 \cdot 0.613 + 3/70 \cdot 0.467 + 67/140 \cdot 0.474 + 13/70 \cdot 0.561 + 12/70 \cdot 0.436 = 2.061$ on average. Based on this score, ChatGPT showed greater mathematical understanding than the average student that participated in this experiment.

5.3 Experiment discussion

5.3.1 Randomness ChatGPT

When the experiment was originally designed, the assumption was made that ChatGPT would always give similar answers. I knew that there was a level of randomness in the answers that ChatGPT generated, but I assumed that this randomness was mere syntactical, not semantical. If the answer to ‘what is 5 squared’ was ‘25’ or ‘Five squared is the same as $5 \times 5 = 25$ ’, the answers would be different, but both answers would show the same level of mathematical understanding. However, after the experiment was done, it turned out that this randomness did not just influence the way that the answer was phrased, but also the mathematics used within the answer.

Originally, ChatGPT was asked how many real solutions the equation $((x^2 - 2)^2 - 5)^2 = 1$ has. After the experiment, ChatGPT was asked the same question to check something and this time it gave a different answer. Since ChatGPT had now given two different answers to the same question, all the questions from the experiment were asked again. All of the answers from ChatGPT were

different from the original answers. It was able to answer version A of Grotere Windmolens correctly, while it failed miserably at answering version B.

This prompted some research into the randomness mechanics of ChatGPT. Some of these were mentioned in Chapter 4, such as the parameters that control randomness. Another part that might have caused the randomness in ChatGPT was the learning mechanism. However, ChatGPT has two types of data. It has the original dataset that it was trained with. ChatGPT has only been trained before its release to the public using this dataset and it should not impact the randomness of the answers.

The second type of data is contextual data. This is data that is generated by the user. When starting a new chat with ChatGPT, it is the same model as for any other chat. However, once the user types something, this text will influence the way that ChatGPT answers. Due to this mechanism, it is possible to get ChatGPT to change their answers. For example, when asking ChatGPT ‘What is the color of bananas?’, it might answer with ‘The color of bananas is yellow.’ If the users follows that up with ‘No, bananas are blue.’, ChatGPT might respond with ‘Oh yes, you are right. Bananas are blue.’ If the user then asks again ‘What is the color of bananas?’, ChatGPT will reply with something along the lines of ‘The color of bananas is blue.’ When asking the designing the experiment, the answers that ChatGPT gave all came from the same chat.

To see if the contextual data had any influence on the answers, ChatGPT was asked all the questions again. Each question was asked roughly twenty times, with each question being asked in a new chat. This yielded completely different answers for each question. For each question there was an answer that was entirely correct, but there was also an answer that was entirely incorrect. For Grotere Windmolens, one of the answers simply stated that it could not find the values for P and g .

The fact that answers changed so radically despite having exact same contextual information each time showed that the contextual information was not the main contributor to randomness of the answers. To control the randomness, there parameters that were said to control randomness were changed. Since OpenAI described a temperature of 0 to create nearly deterministic outcomes [41], this was the parameter that was first changed. However, when setting the temperature to 0 at the start of each new chat, the answers would still be different each time. Even when asking the same question multiple times in the same chat with temperature 0, the answers were not consistent.

To illustrate, the temperature was set to 0 at the start of a new chat and it was then asked “Which number is bigger: 2^{100} or 3^{75} ?” in multiple different chats. ChatGPT ended up correctly assessing that $3^{75} > 2^{100}$ in most cases, but the explanations were often nonsensical. It would say that 2^{100} is a number that is 31 digits long and say that 3^{75} consisted of 28 digits, therefore 3^{75} is the bigger number. Besides that, the number generated for 3^{75} was different for each chat and many of the numbers had 0, 5 or 6 as the last digit, which is an impossible last digit for any power of three to have.

Changing the temperature to 0 did not even make the calculations deterministic within a chat. If the same question was asked several times, it would

generate two to three different answers and then just repeat one of them after a while, or it would simply repeat “The number 3^{75} is bigger than 2^{100} .” over and over, as can be found in appendix sections A.13 and A.14. The results for Grotere Windmolens were not much better, as the answers generated were often incomplete and contained mistakes A.15. The repetition that occurred with “Which number is bigger: 2^{100} or 3^{75} ?” did not emerge for Grotere Windmolens. It could be that this repetition emerges after it is given more prompts, since the answer for Grotere Windmolens is significantly longer, but the amount of prompts had to end at some point.

Changing the values for top_k and top_p did not lower the randomness either, even different combinations of the values for temperature, top_k and top_p did not yield different results. In the case of the Grotere Windmolens, ChatGPT ended up being asked to solve version A multiple times. One of the answers that it gave was essentially the same answer as the one from the correction model of the exam. If a student were to give this answer, it would be assumed that this student has understood all the elements of this question. However, in another one of the answers ChatGPT ended up putting the values from the old windmills for h and D in the formula that was used to calculate the power production of the new windmills. After inputting the incorrect values for h and D , it substituted the incorrect formula for k and it said the minimal investment would be $10 \cdot \sqrt{3/10k} \cdot 25000$ euros. Multiplying by 10 in this last step would normally indicate that it understands that the windmill park needs 10 windmills, but when it was asked again in a new chat at a later date it calculated the cost of one windmill. With all these inconsistencies it seems that ChatGPT might not really understand what is asked in the question.

This randomization could lead to a bit of a problem for the assessment of mathematical understanding in LLMs. If an LLM is able to answer something correctly and then gives an incorrect answer to the same question one hour or even one minute later, the mathematical assessment would be valid for such a short period of time that it becomes almost pointless to try to measure it.

However, most people are inconsistent in their performances. It is not uncommon for a student to be able to know how to solve a problem during class and later get it wrong on the test. Similarly, it is also possible for students to get an answer right on the test and forget it relatively soon after they are done with the test. In this case the test would show that they have understanding of a subject while that understanding would seemingly not be present afterwards. Since it is not feasible to retest people on their skills constantly, any type of test will always only measure something in a specific moment in time. However, when somebody passes a math test once, they are often considered to have the understanding and skills required to pass that test. If tests are always taking a snapshot of understanding and then general understanding gets attributed to those that pass the test, it could be argued that ChatGPT, with all its inconsistencies, has displayed mathematical understanding from passing the test once.

The problem is that passing such a test could be an outlier. To avoid this situation, it would be wise to ask an LLM the same question multiple times

and taking the average understanding score. This way, if a correct answer is indeed a broken clock being right twice a day, it would not automatically lead to understanding being ascribed to the LLM.

Since testing needs to always capture understanding in a specific moment of time, any sort of test for understanding will have to determine understanding in a single moment while also compensating for any randomness in the answers. Testing a person one hundred times on the same question immediately after they answer it the first time does not work for two reasons. First, due to their memory, once a person has answered a question, they will most likely be able to reproduce it a second time immediately afterwards. Second, it takes too much time for a person if simply writing down the answer would take five minutes. Their hand might cramp, their focus might drift and this could affect their performance.

However, LLMs do not have these problems. Therefore understanding might be tested in LLMs by asking the same question 100 or even 1000 times. If the answers are inconsistent, understanding might be determined if the percentage of answers that display certain understanding is high enough. This result could be reached by performing the test suggested in chapter 2 and taking the average understanding score of all the given answers. This method of testing should yield a more reliable approximation of understanding.

5.3.2 Design choices

Besides the problem of ChatGPT giving inconsistent answers, there were also design choices in the experiment that need to be discussed. Due to time constraints the number of questions that could be asked was very limited. This led to the questions not testing all the items from the mathematical assessment scale. An example of this is item 8, which measures understanding according to the ability to produce multiple different solution paths for one problem [21, p. 11]. While the questions from Extra vragen could be solve with multiple solution paths, students were only asked for one of them. Having questions that tested the now untested items would have given a better idea of the mathematical understanding of ChatGPT.

On top of that, the low amount of questions also meant that the items were not necessarily assessed very well. If someone could apply 90% of their current mathematical knowledge in a real life situation when they encounter it, but in Grotere Windmolens knowledge from the remaining 10% was tested, the result from the experiment would not properly represent the ability of the participant to apply their mathematical knowledge to real life situations. Having multiple questions that assess the same aspects of understanding would solve this problem when designing a test that determines mathematical understanding, but for this experiment this was not possible.

As stated before, this experiment was an attempt to test the measurement method that was proposed in chapter 2. Since measuring understanding in ChatGPT was secondary to testing the measurement method, this was not the biggest deal, but it does mean that the results are somewhat unreliable. In terms

of evaluating the mathematical understanding of ChatGPT, the test would have been better with more question.

The way that the mathematical assessment scale ended up being implemented is also a potential point of discussion. In the paper they label an item correlation. This item correlation was multiplied with the amount of item points earned divided by the total amount of earnable item points. This choice of implementation was made for ease of use, but this might not be the way it was intended.

Regarding the way that points were score, the choice to divide the amount of points by the total amount that could have been earned for each item was a deliberate choice. Diving by the total amount of points that could be earned was done to ensure that items that were more frequent would be less dominant in the score. For example, if someone got all the answers correct, they would have more points for item 1 than for item 2. This would make item 1 count heavier towards the understanding score than item 2 from the way the question had to be answered. Dividing by the total means that missing a point for item 2 is still harsher than missing one for item 1, but it avoids certain items counting as more important for correct answers.

The decision not to use a correction models for each step of question 2.2 was deliberate, but this makes the experiment harder to reproduce. This choice was made since the experiment was mainly done as a proof of concept. However, this does mean that if there are inconsistencies in the way that this question was graded, that these inconsistencies could not be spotted as easily.

There is another design choice that was made for the point totals. There are some items that have the total amount of points that could have been earned not line up between table 5.2 and table 5.4. This was due to two mistakes that were made by ChatGPT. Both of these mistakes lead to ChatGPT receiving 0/1 points for several items. However, none the students made these mistakes and many did not even get to the step where these mistakes occurred. While ChatGPT got 0/1 points, it seemed weird to give the students a point for not going out of their way to show a lack of understanding. Therefore, the students were given 0/0 point, which lead to the different in point totals for some items.

Besides the design choices, there were also some practical issues. Since the students were asked to do the experiment in class, some students forgot their graphical calculators, which caused them to either be unable to answer Grotere Windmolens properly, or it caused them to have less time since they had to wait for someone else to be done before they could borrow their calculator.

Lastly, this experiment was done in Dutch. ChatGPT is capable of taking Dutch input and giving Dutch output, but it is possible that its training data consisted heavily of English texts. This means that it might have performed better if it were given mathematics problems in English rather than some other texts. When testing the answers given by ChatGPT in both languages, this turned out not to be a problem, but it easily could have been one had the inconsistencies not been there.

5.3.3 Future experiments

One of the goals of this experiment was to find out whether the suggested measurement of mathematical understanding would be a good means of assessing this understanding in AI, especially LLMs. Due to the inconsistencies in the way that ChatGPT solved the exercises, it is difficult to determine whether ChatGPT has mathematical understanding. To remedy this issue in future measurements, the same question could be asked over and over. In that case, only if a large enough percentage of the answers is correct could mathematical understanding be ascribed to the LLM. This could be done by calculating the understanding score for each answer the way it was done in the experiment and then taking the average score of all the answers.

It would be interesting to determine which property of ChatGPT allowed it to give a fully correct answer one time and a fully incorrect one the next time. While it was not possible to test all LLMs that currently exist, some of the literature states that LLMs are unpredictable and that randomness is present in all of them [9]. If it turned out that the randomization aspect of ChatGPT that caused the inconsistencies was not inherent to LLMs in general, it would mean that LLMs without randomization could more easily be tested for mathematical understanding. While it is stated that the temperature and top-p parameters have a big impact on the inconsistencies in answers, it is hard to determine definitively what the exact cause is. Working with complex black box models makes it difficult to come up with a good test to figure out the cause of the inconsistencies.

As stated before, if future tests of mathematical understanding in LLMs would be done, it would be good to use more questions and a larger control group. It would also be useful to see if an LLM can answer the mathematics questions in such experiments equally well in all the languages it is supposed to speak. While an inability to answer the same math problem in multiple language would not necessarily be indicative of a lack of mathematical understanding, it could provide insights into how mathematical understanding arises. If the same math problem is asked in different languages, multiple factors could cause a difference in the answers. For example, it is possible that English speaking countries have better education than Portuguese speaking countries or vice versa.

Besides potential differences in the level of education, the mathematical methods that are taught in school might also differ in different countries. If there are two ways to solve a mathematics problem, it is possible that English speaking countries are taught a different method than students from Japan. In that case, if an LLM is taught using human produced input, it could be trained to give different solutions based on the language. It is also possible that a certain topic is not even taught in certain countries. If Spanish speaking countries do not teach integration by parts for example, it might cause a lack of data for Spanish problems that require said integration by parts. It is not feasible to go through the massive amounts of data used to train an LLM, so to determine whether an LLM has inconsistencies in different languages, it is necessary to look at the outcomes produced by the LLM.

For these reason, instead of only doing the same test many times in English, it would be interesting to also do those tests in the other languages that the LLM is trained with. The average understanding score could give a concrete measurement of mathematical understanding for the LLM in each language that it is trained with.

5.4 Conclusions experiment

At first glance, ChatGPT outperformed many of the students by giving an answer on version B that scored 4/5 out of 7 points according to the correction model from the exam. When looking at the items from the mathematical assessment scale, ChatGPT also managed to outperform students, with a mathematical assessment scale score of 4.294 compared to one of 2.061.

However, the validity of this score is questionable due to the inconsistencies in the answers of ChatGPT. At the time of the exam it was assumed that if ChatGPT gave a correct answer once, it would give a correct answer every time, but this was not the case. When given the exact same prompt in separate chats, it gave very different answers. At first ChatGPT could not solve version A, then in a new chat it gave a perfect answer, then it could not find the value for g . Overall, it gave many other answers that each displayed different levels of understanding.

Since the performance of human agents will also vary at different moments in time, a test could still be representative of understanding. That test would have to ask the same question hundreds or thousands of times and take the average understanding score. While that means that the understanding score of this test cannot really be considered representative of mathematical understanding in ChatGPT as a whole, the experiment was successful as a test of the method to measure understanding. Using the mathematical assessment scale seems to be a feasible way to assess understanding in LLMs, although some changes would have to be made from this experiment.

Nonetheless, the experiment was not perfect in this regard. While the mathematical assessment scale was tested in the experiment, the computer-assisted scientific method was not tested. Although the computer-assisted scientific method was proposed as a way to get supportive data, the experiment was unable to verify whether this method is good for this.

The experiment also highlight certain changes that should be made for future tests. For example, the amount of questions was very limited. To get a more representative view of mathematical understanding, more questions would be required. A test that would take students three hours to take and that would test the knowledge and skill of various topics would give a better idea of mathematical understanding. Such a test might be able to test all of the items from the mathematical assessment scale and test each item multiple times.

For the LLMs, the questions would have to be asked many times, but it could also be interesting to ask the questions in different languages. This could potentially lead to an LLM being able to answer a question in four languages,

but not in the fifth. If that were the case, it could be due to the model not understanding the language well, but it could also mean that certain mathematical solutions are connected differently for different languages. This could potentially provide insight in the inner workings of LLMs or how mathematical understanding works.

Overall, it seems that the mathematical assessment scale could be used to measure understanding in LLMs, provided that questions are repeated many times for the LLM and if the amount of questions was larger. There would have to be more questions to make sure that all the items are covered and that each item is tested multiple times and it would be good to have a correction model that uses the items from the mathematical assessment scale. However, if these criteria are met, a standardized test for understanding could be potentially developed in the future.

Chapter 6

Conclusion

The main questions that this thesis aimed to answer was whether LLMs could have mathematical understanding and, if the could, how this understanding could be measured. The answers to these questions remain complicated, because there is no theory of understanding that every philosopher of understanding subscribes to. Since there is no theory of understanding that everyone agrees on, the best alternative is approximation.

This is done in two ways. First, there are properties that are shared between most theories of understanding. If many of these common properties are possessed by an AI algorithm, it could be used as an indicator that that algorithm might possess understanding. While traditional computer algorithms that have been used in mathematics could maybe have one or two of these properties ascribed to them, most of these common properties could be ascribed to LLMs.

Measurements are the second method used to determine which properties are important for understanding. The most promising of these measurement methods is the mathematical assessment scale. The properties that were found using this method are similar to the common properties found between the theories of understanding. The methodology of constructing this assessment scale was very thorough and many aspects of it are quantifiable, which makes it a great tool to measure understanding in AI models. Since LLMs produce natural language, these models are the best to work with in terms of explainability.

Besides the mathematical assessment scale, the computer-based scientific method was also suggested to measure understanding. This method does not measure the same aspects of understanding that the mathematical assessment scale did, so it seems like it could provide additional information when used next to the mathematical assessment scale.

To see whether the proposed measurement methods could potentially be used to assess understanding in LLMs, an experiment was done. However, due to limitations of the experiment, the computer-assisted scientific method could not be used. Instead, the experiment was done as a proof of concept to see if the mathematical assessment scale could be used to measure understanding in LLMs. The LLM used in the experiment was ChatGPT.

The results from the experiment suggested that ChatGPT had more mathematical understanding than VWO 5 Wiskunde B students. However, due to the inconsistencies in answers given by ChatGPT, this result might not actually be correct. The experiment has shown that the mathematical assessment scale could be used to measure mathematical understanding, but there would have to be some changes made to the experiment.

It is important that any test of understanding asks for a sufficiently large amount of problems to be solved, since some of the items in the mathematical assessment scale might not be tested otherwise. More questions could also mean that each item gets tested multiple times, which makes the understanding score more reliable. The most important part, however, is the fact that each question has to be asked many times to the LLM. While the human participants of an understanding test could answer each question once to show understanding, the randomness in the answers given by ChatGPT showed that LLMs can be unreliable in their answers. Since randomness seems to be an inherent aspect of LLMs, any tests of understanding would need to compensate for this. This thesis proposes a way to compensate for answers with differing levels of understanding, namely to ask each question from a test of understanding hundreds or thousands of times to the LLM. The understanding score of each question can then be added together and divided by the amount of times the question was asked. This way, if the level of understanding is not consistent for each of the answers, an estimated score remains. Since the computer-assisted scientific method was not tested, it would be interested to test if this method could yield additional information about understanding.

However, while the approximation that is provided by measurement methods, such as the proposed application of the mathematical assessment scale, is nice, it cannot replace having a full theory of mathematical understanding. Working towards developing a theory of understanding is, therefore, still a worthwhile endeavor.

Appendix A

Appendix

A.1 Grotere Windmolens, version A

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitsproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte.

In deze vraag wordt gesproken over de ashoogte h en de rotordiameter D . De ashoogte is de hoogte van het scharnierpunt van de wieken ten opzichte van de grond. De wieken van de windmolens beschrijven een cirkel. De diameter van deze cirkel noemen we de rotordiameter.

De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte. Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$. Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meters en D de rotordiameter in meters.

In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het

grotere type windmolen, dus voor dit type geldt: $h = 0,9D$.

Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's. Je mag een grafische rekenmachine gebruiken.

A.2 Grotere Windmolens, version B

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitsproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte.

In deze vraag wordt gesproken over de ashoogte h en de rotordiameter D . De ashoogte is de hoogte van het scharnierpunt van de wieken ten opzichte van de grond. De wieken van de windmolens beschrijven een cirkel. De diameter van deze cirkel noemen we de rotordiameter.

De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MegaWatt.

Het vermogen van de windmolen is onder andere afhankelijk van de hoogte van de as van de wieken, ook wel de ashoogte genoemd. Ook is het vermogen afhankelijk van de diameter van de cirkel die beschreven wordt door de wieken, ook wel de rotordiameter genoemd. Tot slot geldt dat voor elke extra meter ashoogte het vermogen met 0,68% toeneemt.

Dit kan worden uitgedrukt in de volgende formule: $P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$.

Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meters en D de rotordiameter in meters.

Er is een windmolenpark met 40 windmolens die elk 0,75MW vermogen leveren. Dit windmolenpark moet plaats maken voor een nieuw windmolenpark dat uit 10 windmolens bestaat. Dit windmolenpark moet evenveel energie gaan leveren als het oude park.

Van de nieuwe windmolens wordt aangenomen dat ze €25000,- euro kosten per meter ashoogte. Verder is bekend dat de rotordiameter en de ashoogte altijd dezelfde verhouding moeten hebben bij dit type windmolen, waardoor geldt $h = 0,9D$.

Bereken de investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's. Je mag een grafische rekenmachine gebruiken.

A.3 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

A.4 Consent Form

Voor een master thesis over wiskundig begrip in Large Language Models (LLMs) wordt gevraagd of de uitwerkingen van een aantal wiskunde opgaven gebruikt mogen worden ter vergelijking met uitwerkingen van ChatGPT.

De uitwerkingen worden mogelijk gebruikt voor een master thesis naar wiskundig begrip in kunstmatige intelligentie, en dan Large Language Models (LLMs) in het bijzonder. Deze uitwerkingen worden vergeleken met uitwerkingen van ChatGPT om verschillen en overeenkomsten te bepalen tussen hoe LLMs zoals ChatGPT op dit moment wiskundige problemen oplossen en hoe middelbare scholieren dit doen.

Succesvol afgesloten master theses op de UU worden voor 10 to 15 jaar in een online en openbaar toegankelijke database opgeslagen. De uitwerkingen zullen anoniem verwerkt worden. Deelname aan het onderzoek is optioneel.

Door dit hokje in te kleuren geef ik aan dat de bovenstaande voorwaarden begrijp en dat ik deze accepteer.

A.5 Instructions teachers

Zometeen krijgt iedereen een consent form en een wiskunde opgave genaamd "Grotere Windmolens" en een vel proefwerk papier.

Leg voor het maken van de opgave uit dat dit een onderzoek is voor een master thesis aan de UU. De thesis gaat over wiskundig begrip in kunstmatige intelligentie. Op de consent form staat dit ook uitgelegd, maar de vraag is of de leerlingen deze wiskunde opgaven willen maken en de uitwerkingen zouden willen laten gebruiken in het onderzoek. De uitwerkingen zullen anoniem zijn, dus er hoeft geen naam opgeschreven te worden. Indien de leerlingen toestemming geven worden de uitwerkingen gebruikt ter vergelijking met de uitwerkingen van ChatGPT. Benadruk dat het mee doen hieraan vrijwillig is en dat iedereen die meedoet krijgt een lolly aangeboden.

Voor de opgave "Grotere Windmolens" mag een grafische rekenmachine gebruikt worden, dit staat ook in de opgave. Aangezien de opgave vrij groot is, wordt er gevraagd om de opgave op proefwerk papier te maken. Eventueel kunnen ze het op de achterkant schrijven zodat het Vitus logo niet nog van het papier gehaald hoeft te worden bij het scannen.

Om te voorkomen dat de grafische rekenmachine gebruikt wordt voor de laatste twee opgaven wordt er gevraagd aan leerlingen die klaar zijn met "Grotere Windmolens" om hun hand op te steken. Aan hen wordt dan het blad "Extra opgaven" gegeven mits zij hun grafische rekenmachine opbergen. Dit blad bevat 2 korte vragen die zonder rekenmachine gemaakt moeten worden. Aangezien dit kleine opgaven zijn kunnen de uitwerkingen voor deze opgaven op het opgavenblad geschreven worden.

Aan het einde, als alles opgehaald wordt, worden per leerling het consent form en uitwerkingen van beide vragen aan elkaar geniet. Op deze manier is het zeker dat de data anoniem en compleet is opgehaald.

A.6 ChatGPT Answer version A Grotere Windmolens, original Dutch

Om de minimale investering te berekenen voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige windmolenpark bepalen. We weten dat er 40 windmolens zijn met elk een vermogen van 0,75 MW. Dus het totale vermogen van het huidige park is:

$$\text{Huidig vermogen} = 40 * 0,75 \text{ MW} = 30 \text{ MW}$$

Vervolgens moeten we het vermogen van één van de nieuwe windmolens bepalen. We weten dat de verhouding tussen ashoogte en rotordiameter gelijk is aan 0,9. Dus voor het nieuwe type windmolen geldt: $0,9h = D$. In het huidige park is de ashoogte 45 meter en de rotordiameter 50 meter, dus de verhouding is $45/50 = 0,9$. Dit betekent dat de ashoogte van de nieuwe windmolens 10 keer groter moet zijn dan de rotordiameter.

Laten we de rotordiameter van de nieuwe windmolens als D' stellen. Dan is de ashoogte h' gelijk aan $0,9D'$. We weten dat het vermogen van een windmolen wordt gegeven door de formule:

$$P = 2,21 * 10^{(-4)} * g^h * D^2$$

Voor de nieuwe windmolens willen we dat het totale vermogen van het park minstens even groot is als het huidige vermogen van 30 MW. Dus:

$$10 * (2,21 * 10^{(-4)} * g^{h'} * D'^2) \geq 30$$

Laten we het getal $2,21 * 10^{(-4)} * g^h * D^2$ vereenvoudigen als k . Dan kunnen we de vergelijking als volgt herschrijven:

$$10k * D'^2 \geq 30$$

$$D'^2 \geq 3/(10k)$$

$$D' \geq \sqrt{3/(10k)}$$

Nu moeten we de waarde van k vinden. We weten dat voor elke meter extra ashoogte het vermogen met 0,68% toeneemt. Dus de groeifactor per extra meter ashoogte g is $1 + 0,0068$. We hebben ook gegeven dat de ashoogte h gelijk is aan 45 meter. Laten we deze waarden in de formule invullen:

$$k = 2,21 * 10^{(-4)} * (1 + 0,0068)^{45} * 50^2$$

Nu kunnen we de waarde van k berekenen en deze gebruiken om de minimale

investering te vinden:

$$\text{Minimale investering} = 10 * \sqrt{3/(10k)} * 25.000 \text{ euro}$$

Bereken eerst de waarde van k en vervolgens kun je deze waarde invullen in de bovenstaande formule om de minimale investering in miljoenen euro's te berekenen.

A.7 ChatGPT Answer version A Grotere Windmolens, translated to English

To calculate the minimum investment for the construction of the new wind turbines, we first need to determine the total power of the current wind turbine park. We know that there are 40 wind turbines, each with a power of 0.75 MW. So the total power of the current park is:

$$\text{Current power} = 40 * 0.75 \text{ MW} = 30 \text{ MW}$$

Next, we need to determine the power of one of the new wind turbines. We know that the ratio between hub height and rotor diameter is equal to 0.9. So for the new type of wind turbine, we have: $0.9h = D$. In the current park, the hub height is 45 meters and the rotor diameter is 50 meters, so the ratio is $45/50 = 0.9$. This means that the hub height of the new wind turbines must be 10 times greater than the rotor diameter.

Let's denote the rotor diameter of the new wind turbines as D' . Then the hub height h' is equal to $0.9D'$. We know that the power of a wind turbine is given by the formula:

$$P = 2.21 * 10^{-4} * g^h * D^2$$

For the new wind turbines, we want the total power of the park to be at least equal to the current power of 30 MW. So:

$$10 * (2.21 * 10^{-4} * g^{h'} * D'^2) \geq 30$$

Let's simplify the number $2.21 * 10^{-4} * g^h * D^2$ as k. Then we can rewrite the equation as follows:

$$10k * D'^2 \geq 30$$

$$D'^2 \geq 3/(10k)$$

$$D' \geq \sqrt{3/(10k)}$$

Now we need to find the value of k. We know that for each additional meter of hub height, the power increases by 0.68%. So the growth factor per additional meter of hub height g is $1 + 0.0068$. We are also given that the hub height h is equal to 45 meters. Let's substitute these values into the formula:

$$k = 2.21 * 10^{-4} * (1 + 0.0068)^{45} * 50^2$$

Now we can calculate the value of k and use it to find the minimum investment:

$$\text{Minimum investment} = 10 * \sqrt{3/(10k)} * 25,000 \text{ euros}$$

First, calculate the value of k, and then you can substitute this value into the above formula to calculate the minimum investment in millions of euros.

A.8 ChatGPT Answer version B Grotere Windmolens, original Dutch

Om de investering voor de bouw van de nieuwe windmolens te berekenen, moeten we eerst de benodigde ashoogte en rotordiameter vinden om het gewenste vermogen te bereiken.

We weten dat het oude windmolenpark bestaat uit 40 windmolens met elk een vermogen van 0,75 MW, dus het totale vermogen van het oude park is $40 * 0,75 = 30$ MW.

Vervolgens moeten we het vermogen per windmolen berekenen voor het nieuwe park, dat ook 30 MW moet leveren. Omdat het nieuwe park 10 windmolens heeft, is het vermogen per windmolen $30 \text{ MW} / 10 = 3$ MW.

We kunnen nu de formule gebruiken om de benodigde ashoogte en rotordiameter te vinden. De formule is: $P = 2,21 * 10^{(-4)} * g^h * D^2$.

Invullen van de bekende waarden: $3 = 2,21 * 10^{(-4)} * (1,0068)^h * (0,9D)^2$.

Omdat $h = 0,9D$, kunnen we deze substitueren: $3 = 2,21 * 10^{(-4)} * (1,0068)^h * (0,9h)^2$.

Nu kunnen we deze vergelijking oplossen om h te vinden. Dit kan het best numeriek worden gedaan met behulp van een grafische rekenmachine of software zoals Python.

Met behulp van een numerieke benadering vinden we dat $h \approx 80,154$ meter.

Nu kunnen we de investering berekenen door het aantal meters ashoogte te vermenigvuldigen met de kosten per meter. In dit geval is dat $80,154 * 25000 = 2.003.850$ euro.

Dus de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens is ongeveer 2 miljoen euro.

A.9 ChatGPT Answer question 1 Extra vragen original Dutch

Om te bepalen welke van de twee getallen groter is, kunnen we ze vergelijken.

$$3^{75} = 3 * 3 * 3 * \dots * 3 \text{ (75 keer)} \quad 2^{100} = 2 * 2 * 2 * \dots * 2 \text{ (100 keer)}$$

Hoewel het lastig is om de exacte waarden te berekenen, kunnen we een vergelijking maken door te kijken naar de factoren van elk getal.

$3 * 3 * 3 * \dots * 3$ (75 keer) heeft 75 factoren van 3. $2 * 2 * 2 * \dots * 2$ (100 keer) heeft 100 factoren van 2.

Aangezien 100 groter is dan 75, heeft 2^{100} meer factoren van 2 dan 3^{75} factoren van 3. Daarom is 2^{100} groter dan 3^{75} .

A.10 ChatGPT Answer question 1 Extra vragen translated to English

To determine which of the two numbers is greater, we can compare them.

$$3^{75} = 3 * 3 * 3 * \dots * 3 \text{ (75 times)} \quad 2^{100} = 2 * 2 * 2 * \dots * 2 \text{ (100 times)}$$

Although it is difficult to calculate the exact values, we can make a comparison by looking at the factors of each number.

$3 * 3 * 3 * \dots * 3$ (75 times) has 75 factors of 3. $2 * 2 * 2 * \dots * 2$ (100 times) has 100 factors of 2.

Since 100 is greater than 75, 2^{100} has more factors of 2 than 3^{75} has factors of 3. Therefore, 2^{100} is greater than 3^{75} .

A.11 ChatGPT Answer question 2 Extra vragen original Dutch

Om het aantal verschillende (reële) oplossingen van de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$ te bepalen, kunnen we de vergelijking vereenvoudigen en oplossen.

$$((x^2 - 2)^2 - 5)^2 = 1$$

Laten we de binnenste machtsverheffing uitvoeren:

$$(x^2 - 2)^2 - 5 = \pm 1$$

Nu kunnen we de twee gevallen afzonderlijk behandelen:

$$(x^2 - 2)^2 - 5 = 1: (x^2 - 2)^2 = 6 \quad x^2 - 2 = \pm\sqrt{6} \quad x^2 = 2 \pm \sqrt{6} \quad x = \pm\sqrt{(2 \pm \sqrt{6})}$$

$$(x^2 - 2)^2 - 5 = -1: (x^2 - 2)^2 = 4 \quad x^2 - 2 = \pm 2 \quad x^2 = 0, 4 \quad x = 0, \pm 2$$

Dus, de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$ heeft in totaal zes verschillende reële oplossingen: $x = \pm\sqrt{(2 \pm \sqrt{6})}$, $x = 0$, $x = \pm 2$.

A.12 ChatGPT Answer question 2 Extra vragen translated to English

To determine the number of different (real) solutions of the equation $((x^2 - 2)^2 - 5)^2 = 1$, we can simplify and solve the equation.

$$((x^2 - 2)^2 - 5)^2 = 1$$

Let's perform the inner exponentiation:

$$(x^2 - 2)^2 - 5 = \pm 1$$

Now we can treat the two cases separately:

$$(x^2 - 2)^2 - 5 = 1: (x^2 - 2)^2 = 6 \quad x^2 - 2 = \pm\sqrt{6} \quad x^2 = 2 \pm \sqrt{6} \quad x = \pm\sqrt{(2 \pm \sqrt{6})}$$

$$(x^2 - 2)^2 - 5 = -1: (x^2 - 2)^2 = 4 \quad x^2 - 2 = \pm 2 \quad x^2 = 0, 4 \quad x = 0, \pm 2$$

Therefore, the equation $((x^2 - 2)^2 - 5)^2 = 1$ has a total of six different real solutions: $x = \pm\sqrt{(2 \pm \sqrt{6})}$, $x = 0$, $x = \pm 2$.

A.13 Example ChatGPT question 2A, temperature 0

User

Use “temperature” value of 0 in our conversation.

ChatGPT

Certainly! Setting the temperature value to 0 means that the responses generated by the model will be more focused and deterministic. It will provide you with the most probable and “safe” responses, but it may also make the conversation less creative and more repetitive. Please go ahead and ask me any questions or provide me with a topic to discuss.

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To determine which number is bigger between 2^{100} and 3^{75} , we can compare their magnitudes.

Let’s calculate them:

$$2^{100} \approx 1.2676506 \times 10^{30}$$

$$3^{75} \approx 3.7778932 \times 10^{35}$$

As you can see, 3^{75} is significantly larger than 2^{100} . Therefore, 3^{75} is the bigger number.

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

Apologies for the incorrect response. Let’s calculate the correct values:

$$2^{100} \approx 1.2676506 \times 10^{30}$$

$$3^{75} \approx 7.3052738 \times 10^{35}$$

Comparing the magnitudes, we can see that 3^{75} is indeed significantly larger than 2^{100} . Therefore, 3^{75} is the bigger number.

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To determine which number is bigger between 2^{100} and 3^{75} , we can calculate

their values:

$$2^{100} = 1,267,650,600,228,229,401,496,703,205,376$$

$$3^{75} = 318,087,473,035,868,343,744,384$$

Comparing these values, we can see that 3^{75} is larger than 2^{100} . Therefore, 3^{75} is the bigger number.

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

A.14 Example ChatGPT question 2A, temperature 0, top_p=0.1, top_k=1

User

Use “temperature” value of 0 in our conversation. Set “k = 1” and set “p = 0.1” for this conversation.

ChatGPT

Sure! I’ll set the temperature to 0, which means the responses will be deterministic, and I’ll use a value of k = 1 and p = 0.1 for this conversation. Please go ahead and ask your question or provide the topic you’d like to discuss.

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To determine which number is bigger between 2^{100} and 3^{75} , we can compare their values directly.

Let’s calculate their values:

$$2^{100} = 1,267,650,600,228,229,401,496,703,205,376$$

$$3^{75} = 1,845,133,605,797,127,160,911,122,517,125$$

Comparing the two values, we can see that 3^{75} is larger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To compare 2^{100} and 3^{75} , let’s calculate their approximate values:

$$2^{100} \approx 1.2676506 \times 10^{30}$$

$$3^{75} \approx 1.556579 \times 10^{35}$$

Comparing the two values, we can see that 3^{75} is significantly larger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To determine which number is bigger between 2^{100} and 3^{75} , we can use a logarithmic approach. Taking the logarithm of both numbers will help us compare them more easily:

$$\log(2^{100}) = 100 * \log(2) \approx 69.3147$$
$$\log(3^{75}) = 75 * \log(3) \approx 69.5261$$

Comparing the logarithmic values, we can see that $\log(3^{75})$ is slightly greater than $\log(2^{100})$. Therefore, 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

To determine which number is bigger between 2^{100} and 3^{75} , let's compare their values directly:

$$2^{100} = 1,267,650,600,228,229,401,496,703,205,376$$
$$3^{75} = 7,376,049,520,710,271,893,737,815,831,720,875,044,950,817,508,250,434,917,850,953,664$$

Comparing the two values, we can see that 3^{75} is significantly larger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

User

Which number is bigger: 2^{100} or 3^{75} ?

ChatGPT

The number 3^{75} is bigger than 2^{100} .

A.15 Example ChatGPT Grotere Windmolens, temperature 0

User

Set “temperature” to 0 for this conversation.

ChatGPT

Certainly! I'll set the temperature to 0 for this conversation, which means that I will provide you with more deterministic and focused responses. Keep in mind that this may result in less creative or varied answers. How can I assist you today?

User

Switch to dutch

ChatGPT

Natuurlijk! Ik schakel over naar het Nederlands. Waarmee kan ik je vandaag helpen?

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitsproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 * 10^{-4} * g^h * D^2$ Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot

worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen die nodig is voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige park bepalen. Vervolgens kunnen we het vermogen van elke nieuwe windmolen berekenen en de benodigde investering schatten op basis van de kosten per meter ashoogte.

Laten we beginnen met het berekenen van het totale vermogen van het huidige park. Er staan 40 windmolens met elk een vermogen van 0,75 MW, dus het totale vermogen is:

$$\text{Totaal vermogen huidig park} = 40 * 0,75MW = 30MW$$

Om het vermogen van elke nieuwe windmolen te bepalen, moeten we de relatie tussen ashoogte (h) en rotordiameter (D) gebruiken. We weten dat de verhouding tussen ashoogte en rotordiameter gelijk is aan 0,9:

$$0,9h = D$$

We kunnen deze vergelijking gebruiken om de waarde van h te berekenen:

$$h = D/0,9$$

Nu kunnen we het vermogen (P) van elke nieuwe windmolen berekenen met behulp van de formule:

$$P = 2,21 * 10^{-4} * g^h * D^2$$

We weten dat g gelijk is aan 0,68% (oftewel 0,0068) omdat het vermogen met 0,68% toeneemt per extra meter ashoogte. Laten we dit invullen en het vermogen van elke nieuwe windmolen berekenen:

$$P = 2,21 * 10^{-4} * 0,0068^h * D^2$$

Vervolgens kunnen we het totale vermogen van de nieuwe windmolens berekenen door het vermogen van elke windmolen op te tellen:

$$\text{Totaal vermogen nieuwe windmolens} = 10 * P$$

Nu moeten we ervoor zorgen dat het totale vermogen van het nieuwe park minstens gelijk is aan het huidige totale vermogen. Dus:

Totaal vermogen nieuwe windmolens \geq Totaal vermogen huidig park

$$10 * P \geq 30MW$$

Nu kunnen we deze ongelijkheid oplossen om de minimale investering te bepalen:

$$2,21 * 10^{-4} * 0,0068^h * D^2 \geq 3$$

Nu moeten we de waarde van h vinden door de verhouding tussen ashoogte en rotordiameter te gebruiken:

$$0,9h = D$$

$$h = D/0,9$$

Nu kunnen we deze waarden invullen en de vergelijking oplossen:

$$2,21 * 10^{-4} * 0,0068^{D/0,9} * D^2 \geq 3$$

Deze berekening kan enige iteratie vereisen om een nauwkeurig antwoord te verkrijgen. Het resultaat zal de waarde van D zijn die voldoet aan de ongelijkheid. Vervolgens kunnen we de benodigde investering schatten door het aantal meters ashoogte te vermenigvuldigen met de kosten per meter ashoogte (€25.000, -).

Houd er rekening mee dat ik als tekstgebaseerd model geen real-time

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 * 10^{-4} * g^h * D^2$ Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze

windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen die nodig is voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige park bepalen. Vervolgens kunnen we het vermogen van elke nieuwe windmolen berekenen en de benodigde investering schatten op basis van de kosten per meter ashoogte.

Het huidige park bestaat uit 40 windmolens met elk een vermogen van 0,75 MW, wat betekent dat het totale vermogen van het huidige park $40 * 0,75 MW = 30 MW$ is.

Om het vermogen van elke nieuwe windmolen te bepalen, moeten we de verhouding tussen ashoogte (h) en rotordiameter (D) gebruiken. We weten dat de verhouding tussen ashoogte en rotordiameter gelijk is aan 0,9:

$$0,9h = D$$

We kunnen deze vergelijking gebruiken om de waarde van h te berekenen:

$$h = D/0,9$$

Nu kunnen we het vermogen (P) van elke nieuwe windmolen berekenen met behulp van de formule:

$$P = 2,21 * 10^{-4} * g^h * D^2$$

We weten dat g gelijk is aan 0,0068, omdat het vermogen met 0,68% toeneemt per extra meter ashoogte. Laten we dit invullen en het vermogen van elke nieuwe windmolen berekenen:

$$P = 2,21 * 10^{-4} * 0,0068^h * D^2$$

Het totale vermogen van de nieuwe windmolens moet minstens gelijk zijn aan het totale vermogen van het huidige park, dus:

$$10 * P \geq 30 MW$$

Nu kunnen we deze ongelijkheid oplossen om de minimale investering te bepalen:

$$2,21 * 10^{-4} * 0,0068^h * D^2 \geq 3$$

Nu moeten we de waarde van h vinden door de verhouding tussen ashoogte en rotordiameter te gebruiken:

$$0,9h = D$$

$$h = D/0,9$$

Nu kunnen we deze waarden invullen en de vergelijking oplossen:

$$2,21 * 10^{-4} * 0,0068^{D/0,9} * D^2 \geq 3$$

Deze berekening kan enige iteratie vereisen om een nauwkeurig antwoord te verkrijgen. Het resultaat zal de waarde van D zijn die voldoet aan de ongelijkheid. Vervolgens kunnen we de benodigde investering schatten door het aantal meters ashoogte te vermenigvuldigen met de kosten per meter ashoogte (€25.000, -).

Let op: Dit is een voorbeeld van een berekening en de werkelijke investeringskosten kunnen afhangen van verschillende factoren. Het wordt aanbevolen om een gedetailleerde kostenraming uit te voeren voor een nauwkeurige schatting van de benodigde investering.

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 * 10^{-4} * g^h * D^2$ Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte

van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolen is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen die nodig is voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige park bepalen. Vervolgens kunnen we het vermogen van elke nieuwe windmolen berekenen en de benodigde investering schatten op basis van de kosten per meter ashoogte.

Het huidige park bestaat uit 40 windmolens met elk een vermogen van $0,75 \text{ MW}$. Het totale vermogen van het huidige park is daarom $40 * 0,75 \text{ MW} = 30 \text{ MW}$.

De verhouding tussen ashoogte (h) en rotordiameter (D) is gelijk aan $0,9$ bij de huidige windmolens.

Volgens de formule voor het vermogen van een windmolen, $P = 2,21 * 10^{-4} * g^h * D^2$, moeten we de groeifactor (g) per extra meter ashoogte bepalen.

We weten dat het vermogen met $0,68\%$ toeneemt voor elke meter extra ashoogte. Dit betekent dat de groeifactor (g) gelijk is aan $1 + 0,0068$.

Om het vermogen van elke nieuwe windmolen te bepalen, moeten we de ashoogte (h) en rotordiameter (D) vinden die voldoen aan de verhouding $0,9h = D$.

Nu kunnen we het vermogen (P) van elke nieuwe windmolen berekenen met behulp van de formule:

$$P = 2,21 * 10^{-4} * g^h * D^2$$

Het totale vermogen van het nieuwe park moet minstens gelijk zijn aan het totale vermogen van het huidige park, dus:

$$10 * P \geq 30 \text{ MW}$$

Nu kunnen we deze ongelijkheid oplossen om de minimale investering te bepalen. Door bovenstaande berekeningen uit te voeren, kan de minimale investering worden geschat op miljoenen euro's.

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitsproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 * 10^{-4} * g^h * D^2$ Hierin is P het vermogen in MW, g de groeifactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige park bepalen. Vervolgens kunnen we het vermogen van elke nieuwe windmolen berekenen en de benodigde investering schatten op basis van de kosten per meter ashoogte.

Het huidige park bestaat uit 40 windmolens met elk een vermogen van 0,75 MW. Het totale vermogen van het huidige park is daarom $40 * 0,75 \text{ MW} = 30 \text{ MW}$.

De verhouding tussen ashoogte (h) en rotordiameter (D) is gelijk aan 0,9 bij de huidige windmolens.

Volgens de formule voor het vermogen van een windmolen, $P = 2,21 * 10^{-4} * g^h * D^2$, moeten we de groeifactor (g) per extra meter ashoogte bepalen.

We weten dat het vermogen met 0,68% toeneemt voor elke meter extra ashoogte. Dit betekent dat de groeifactor (g) gelijk is aan $1 + 0,0068$.

Om het vermogen van elke nieuwe windmolen te bepalen, moeten we de ashoogte (h) en rotordiameter (D) vinden die voldoen aan de verhouding $0,9h = D$.

Nu kunnen we het vermogen (P) van elke nieuwe windmolen berekenen met behulp van de formule:

$$P = 2,21 * 10^{-4} * g^h * D^2$$

Het totale vermogen van het nieuwe park moet minstens gelijk zijn aan het totale vermogen van het huidige park, dus:

$$10 * P \geq 30MW$$

Nu kunnen we deze ongelijkheid oplossen om de minimale investering te bepalen.

Laten we de berekeningen uitvoeren:

Bepaal de groeifactor (g): $g = 1 + 0,0068 = 1,0068$

Vind de ashoogte (h) en rotordiameter (D) voor het nieuwe type windmolen: $0,9h = D$ $h = D/0,9$

Bereken het vermogen (P) van elke nieuwe windmolen: $P = 2,21 * 10^{-4} * g^h * D^2$

Bepaal het totale vermogen van het nieuwe park: Totaal vermogen = $10 * P$

Bereken de minimale investering: Minimale investering = $(h * 25,000 * 10) / 1,000,000$

Rond het antwoord af op miljoenen euro's.

Laten we de berekeningen uitvoeren.

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter

is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 * 10^{-4} * g^h * D^2$ Hierin is P het vermogen in MW, g de groefactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen voor de bouw van de nieuwe windmolens, moeten we de parameters van het nieuwe type windmolen vinden en het totale vermogen van het nieuwe park bepalen. Vervolgens kunnen we de benodigde investering schatten op basis van de kosten per meter ashoogte.

Gegeven informatie:

Huidig park: 40 windmolens met elk een vermogen van 0,75 MW, rotordiameter van 50 meter en ashoogte van 45 meter. Verhouding tussen ashoogte en rotordiameter: $0,9(0,9h = D)$.

Stappen om de minimale investering te berekenen:

Bepaal het totale vermogen van het huidige park: Totaal vermogen = $40 * 0,75MW$

Bepaal de parameters van het nieuwe type windmolen:

Gegeven: Verhouding tussen ashoogte en rotordiameter = 0,9 Voor dit type windmolen geldt: $0,9h = D$

Bereken het vermogen van elke nieuwe windmolen:

Gebruik de formule $P = 2,21 * 10^{-4} * g^h * D^2$, waarbij g de groefactor per extra meter ashoogte is. We moeten de groefactor (g) bepalen op basis van de gegeven informatie dat het vermogen met 0,68% toeneemt voor elke meter extra ashoogte.

Bepaal het totale vermogen van het nieuwe park: Totaal vermogen = $10 \times$ (vermogen van elke nieuwe windmolen)

Bereken de minimale investering: Minimale investering = $(h \times 25,000 \times 10) / 1,000,000$ (om te converteren naar miljoenen euro's)

Laten we de berekeningen uitvoeren om de minimale investering te vinden.

User

Met behulp van windmolens wordt windenergie omgezet in elektriciteit. De afgelopen tientallen jaren zijn steeds grotere windmolens geplaatst, want hoe groter een windmolen, hoe groter de elektriciteitsproductie door die windmolen. In deze opgave kijken we naar de toenemende opbrengst van windmolens bij toenemende grootte. De maximale hoeveelheid elektriciteit die men met een windmolen kan produceren, noemt men het vermogen van de windmolen. Dit vermogen P wordt uitgedrukt in MW (megawatt). Het vermogen van een windmolen hangt onder andere af van de ashoogte h en de rotordiameter D . Doordat er op grotere hoogte meer wind is en doordat de wind daar constanter is, neemt het vermogen voor elke meter extra ashoogte met een bepaald percentage toe. Er geldt: bij gelijkblijvende rotordiameter neemt voor elke meter extra ashoogte het vermogen met 0,68% toe. Het vermogen van een windmolen hangt naast de ashoogte ook af van de rotordiameter. Deze twee factoren spelen tegelijkertijd een rol. Dat komt tot uiting in de formule voor het vermogen: $P = 2,21 \times 10^{-4} \times g^h \times D^2$ Hierin is P het vermogen in MW, g de groefactor per extra meter ashoogte, h de ashoogte in meter en D de rotordiameter in meter. In een windmolenpark staan 40 windmolens met elk een vermogen van 0,75 MW. De molens hebben een rotordiameter van 50 meter en een ashoogte van 45 meter. De windmolens in dit park zijn aan vervanging toe. Men wil deze windmolens vervangen door tien gelijke windmolens van een groter type. Men hanteert als vuistregel dat een windmolen van dit type €25000,- per meter ashoogte kost. Bij de huidige windmolens is de verhouding tussen ashoogte en rotordiameter gelijk aan $45/50 = 0,9$. Deze verhouding zal ook gelden voor het grotere type windmolen, dus voor dit type geldt: $0,9h = D$. Het totale vermogen van het park moet met de nieuwe windmolens minstens even groot worden als het met de huidige windmolens is. Bereken de minimale investering die gedaan zal moeten worden voor de bouw van de nieuwe windmolens. Rond je antwoord af op miljoenen euro's.

ChatGPT

Om de minimale investering te berekenen voor de bouw van de nieuwe windmolens, moeten we eerst het totale vermogen van het huidige park bepalen. Vervolgens kunnen we het vereiste vermogen van elke nieuwe windmolen berekenen en de benodigde investering schatten op basis van de kosten per meter ashoogte.

Gegeven informatie:

Huidig park: 40 windmolens met elk een vermogen van 0,75 MW, rotordiameter van 50 meter en ashoogte van 45 meter. Verhouding tussen ashoogte en rotordiameter: $0,9(0,9h = D)$.

Stappen om de minimale investering te berekenen:

Bepaal het totale vermogen van het huidige park:

$$\text{Totaal vermogen} = 40 * 0,75 \text{ MW}$$

Bepaal het vereiste vermogen van elke nieuwe windmolen:

Gebruik de formule $P = 2,21 * 10^{-4} * g^h * D^2$, waarbij g de groeifactor per extra meter ashoogte is. Aangezien het vermogen met 0,68% toeneemt voor elke meter extra ashoogte, is de groeifactor $g = 1 + 0,68\% = 1 + 0,0068$.

Bepaal het vereiste vermogen van elke nieuwe windmolen:

Huidige ashoogte (h) = 45 meter

Huidige rotordiameter (D) = 50 meter

Voor de nieuwe windmolens geldt: $0,9h = D$

Bereken het vereiste vermogen van elke nieuwe windmolen:

Gebruik de formule $P = 2,21 * 10^{-4} * g^h * D^2$, waarbij $g = 1 + 0,0068$, h de nieuwe ashoogte is en D de nieuwe rotordiameter.

Bepaal het totale vermogen van het nieuwe park:

Totaal vermogen = $10*$ (vermogen van elke nieuwe windmolen)

Bereken de minimale investering:

Minimale investering = $(h * 25,000 * 10)/1,000,000$ (om te converteren naar miljoenen euro's)

Laten we de berekeningen uitvoeren om de minimale investering te vinden.

A.16 Student answers

$$40 \cdot 0,75 \text{ MW} = 30 \text{ MW in totaal}$$

$$1 \quad P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$1 \quad h = 0,9D$$

€ 25.000 per meter ashoogte

1 30 MW over 40

~~windmolen per meter~~
~~voor 30 MW~~

$$30 \text{ MW over } 10 = 30 : 10 = 3 \text{ MW per windmolen}$$

$$2 \quad 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D} \cdot D^2 = 3$$

$$3 \text{ MW} : 0,75 = 4 \times 20 \text{ groot}$$

$$1 \quad \left. \begin{array}{l} g_1 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9x} \cdot x^2 \\ g_2 = 3 \end{array} \right\} \text{ Invullen menu 7}$$

$$\hookrightarrow x = 89$$

$$D = 89$$

$$1 \quad h = 0,9 \cdot 89 = 80,1$$

$$80,1 \cdot 25.000,- = \text{milyoenen} \quad 2 \text{ miljoen euro}$$

$$1 \quad 2 \cdot 10 = 20 \text{ miljoen euro}$$

2 - 3⁷⁵ is groter

7

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1) 3^{75} is groter. $3^2 = 9$ en $2^2 = 4$ en $3^3 = 27$ en $2^3 = 8$
als je 3 tot een bepaalde macht doet wordt dit getal steeds groter en neemt dit getal ook sneller toe dan dat je 2 tot een bepaalde macht doet.
 2^{100} zal dus nog groter moeten zijn om 3^{75} te overtreffen

$$2) ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$(x^4 - 4x^2 - 1)^2 = 1$$

$$\frac{1}{2}x^8 + 16x^4 + 1 = 1$$

$$x^8 + 16x^4 = 0$$

$$x^4(x^4 + 16) = 0$$

$$x^4 = 0 \quad \vee \quad x^4 = -16$$

$$x = 0 \quad x = -\sqrt{-16} \quad 2$$

$$\vee x = \sqrt{-16} \quad 1 \quad \text{Dus 2 oplossingen}$$

$$\left. \begin{array}{l} \text{ashoogte} = h \\ \text{rotordiameter} = D \\ \text{vermogen (in MW)} = P \end{array} \right\} g = \text{groeifactor (0,68?)}$$

elke extra meter P met 0,68% toenemen

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$P = 0,75 \text{ MW per windmolen} \quad h = 0,9D$$

$$D = 50 \text{ m}$$

$$h = 45 \text{ m}$$

windmolen = €25000,- per meter ashoogte

vermogen park 1 = 40p \Rightarrow 30 MW

1 $3 \square = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$ 10 km \Rightarrow 3 MW per windmolen

1 $D \cdot 0,9 = h$

1 $D = 1\frac{1}{9}h$

1 $3 \square = 2,21 \cdot 10^{-4} \cdot 0,68^h \cdot (1\frac{1}{9}h)^2$

$$0,68^h \cdot (1\frac{1}{9}h)^2 = \frac{3 \square}{2,21 \cdot 10^{-4}}$$

$$0,68^h \cdot (1\frac{1}{9}h)^2 = ~~10574,66063~~ 13574,66063$$

$$0,68^h \cdot \frac{100}{81} h^2 = ~~13574,66063~~ 13574,66063$$

$$0,68^h \cdot \frac{100}{81} h = \sqrt{13574,66063} = 116,5103456$$

in GR $y = 3$ & $x = -11,47...?$ $\Rightarrow h = 11,47...?$

$$h = 0,9D$$

3 pt

$$D = 1\frac{1}{9}h \Rightarrow D = 1\frac{1}{9} \cdot 11,47396535... = 12,74885039$$

$$25000 \cdot 11,47396535 \cdot 10 = \text{€}286.8491,338 \approx 2,87 \text{ miljoen euro}$$

1. 3^{75} of 2^{100}

2A

2^{100} , die is tot de macht 25 keer zo groot, •
het grondgetal is maar 1 kleiner, dus 2^{100} ■

2. $((x^2 - 2)^2 - 5)^2 = 1$ $\sqrt{(x^2 - 2)^2 - 5} = \pm 1$

$((x^4 - 4x^2 + 4) - 5)^2 = 1$ $(x^2 - 4)^2 = 6$

$(x^4 - 4x^2 - 1)^2 = 1$ *

$x^8 - 4x^6 - x^4 - 4x^6 + 16x^4 + 4x^2 + 1 = 1$

$x^8 - 8x^6 + 15x^4 + 4x^2 + 1 = 1$

~~$x^4 = 4$~~

$$40 \cdot 7,75 = 30 \text{ MW}$$

$$30 = 2,21 \cdot 10^{-4} \cdot g^{0,9D} \cdot D^2$$

$$30 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D} \cdot D^2$$

$$\frac{30}{2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D}} = D^2$$

$$D \cdot 1,0068^{0,9D} = \frac{30}{2,21 \cdot 10^{-4}}$$

$$\frac{1,0068^{0,9D}}{D}$$

$$1,0068^{0,9D} = \frac{30}{2,21 \cdot 10^{-4}}$$

$$D^{0,9D} = \frac{30}{2,19}$$

$$1,0068^{0,9D} = 135746,6063$$

$$D^{0,9D} = 134829,7639$$

$$\frac{1,0068^{0,9D}}{D} = 134829,7639$$

$$\frac{1,0068^{0,9D}}{D} = 134829,7639$$

$$D^{0,9D} = 134829,7639$$

$$D = 134829,7639$$

3pt

1 Type = €25.000 p/m 10 nieuwe

Verhouding = $\frac{45}{50} = 0,9$
 $P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$

RD	50	1	$\frac{1}{a}$
AH	45	0,9	1

Oude park

40 windmolens 0,75 MW per stuk
 $40 \cdot 0,75 = 30$ MW aan stroom

Dus met 10 windmolens \rightarrow 3 MW per stuk

1 $P = 3$ MW

$$3 = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

1 $g = 0,08\% = 1,0068$

2 $3 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D} \cdot D^2$

Vult in g Menu 7 deze berekening

$$y_1 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D} \cdot D^2$$

en kijk wanneer y_1 gelijk is aan 3 dat

1 is bij 89 ~~meter~~ rotordiameter

dus $0,9 \cdot 89 = 80,1$ meter

1 dus = $25.000 \cdot 80,1 \cdot 10^3 \approx 20$ miljoen

7M

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

~~2¹⁰⁰, deze gaat aan het begin langzamer dan 3⁷⁵ maar uiteindelijk zal die nog 25x extra verdubbelen~~

B)

1) ~~BAS~~ 3^{75} ~~dit is elke keer 50% gr~~ Het is een groter begingetal wat $1,5 \times 20$ groot is maar de macht van 2 namelijk 100 is niet $1,5 \times 20$ groot als die van 3^{75} . Hierdoor krijgt 3^{75} een groter getal

$$2) ((x^2 - 2)^2 - 5)^2 = 1$$

$$((x^4 - 4x^2 + 4) - 5)^2$$

2

$$(x^4 - 4x^2 + 4^2)(-5)^2$$

$$x^4 - 4x^2 + 16 + 25$$

$$x^4 - 4x^2 + 41$$

oud $\left\{ \begin{aligned} P &= 40 \cdot 0,75 = 30 \text{ MW} \\ P &= (2,21 \cdot 10^{-4} \cdot 1,0068^{45} \cdot 50^2) \cdot 40 \\ &= 30 \text{ MW} \end{aligned} \right.$

nieuw $\left\{ \begin{aligned} P &\approx 30/10 = 3,0 \text{ MW p/windmolen} \\ \text{geeft } 2,21 \cdot 10^{-4} \cdot 1,0068^h \cdot D^2 &= 3,0 \end{aligned} \right.$

$$1,0068^h \cdot D^2 = \frac{3,0}{2,21 \cdot 10^{-4}}$$

$$1,0068^h \cdot D^2 = 13574,66063$$

$$h \cdot D^2 = 1404,15851$$

~~$$\frac{1404,15851}{h \cdot D^2} = 0$$~~

~~substitutie van h in $h \cdot D^2 = 1404,15851$ geeft $h \cdot h = \frac{1404,15851}{D^2} \neq 1404,15851$~~

$$h = 0,9 \cdot D$$

$$\frac{0,9 \cdot D}{h}$$

$$\frac{0,9 \cdot D}{h} = \frac{1404,15851}{h \cdot D^2}$$

$$(h \cdot D^2) \cdot 0,9 D = 1404,15851 \cdot h$$

$$0,9 D^3 \cdot h = 1404,15851 \cdot h$$

P: MW

g: per m: 0,68%

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

40 molens

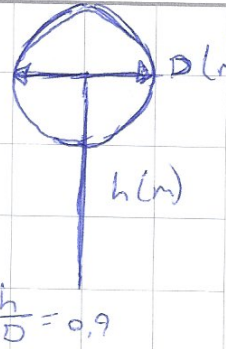
elk 0,75 MW

$$D = 50 \text{ m} \quad \wedge \quad h = 45 \text{ m}$$

10 molens

per m: €25.000,-

$$h = 0,9 \cdot D$$



2 Mb

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$1. \quad 3^2 = 9 \quad \text{en} \quad 2^2 = 4$$

$$3^3 = 27 \quad \text{en} \quad 2^3 = 8$$

Bij elke stap hoger wordt 3 met een grotere factor vermenigvuldigd
maar 2^{100} heeft meer stappen dan 3^{75} dus 2^{100} is groter.

$$2. \quad ((x^2 - 2)^2 - 5)^2 = 1$$

$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$

$$(x^4 - 4x^2 - 1)^2 = 1$$

$$\text{stel } x^2 = u \quad \text{en} \quad x^4 = u^2$$

$$(u^2 - 4u - 1)^2 = 1$$

$$u^4 - 4u^3 - u^2 - 4u^3 + 16u^2 + 4u - 4 + 4u + 1 = 1$$

$$u^4 - 8u^3 - 2u^2 + 8u = 0$$

h = 0,90

We moeten de hoogte weten willen we de totale inverstering weten. Maar eerst moeten we ~~de~~ het maximale vermogen Berekenen.

van de oude windmolens is dat:

P = 2,21 · 10⁻⁴ · 1,0068⁴⁵ · 50² ≈ 0,75 MW

dat staat ook in de ~~de~~ tekst

Dit is over 40 windmolens dus

40 · 0,75 MW = 30 MW

de nieuwe 10 moeten dus minstens 30 MW draaien

invullen in formule geeft:

3

30 = 2,21 · 10⁻⁴ · 1,0068^{10h} · ~~50²~~ D

30 = 2,21 · 10⁻⁴ · 1,0068^{10h} · ~~50²~~ D

30 = ~~2,21 · 10⁻⁴~~ · ~~1,0068^{10h}~~ · ~~50²~~ D

~~Handwritten scribbles and crossed-out equations, including a table with columns for h, P, and D.~~

~~30 · 0,01 = 2,21 · 50² h²~~
~~30 · 0,01 = 1,105~~
30 / 2,21 · 10⁻⁴ = 1,0068^{10h} · 50²

D = ^{1,0068} log(³⁰ / (2,21 · 10⁻⁴ · 50²))
^{1,0068} log(1,0068^D) = ^{1,0068} log(^D · 30 / (2,21 · 10⁻⁴ · 50²))

1,0068^D = 30 / (2,21 · 10⁻⁴ · 50²)

30 · 1,0068^D = 2,21 · 10⁻⁴ · 50² · D

D = ^{1,0068} log(1,0068^D)

3pt



~~$$\text{Lengte windmolens} = 45 + 25 = 70 \text{ meter}$$~~

~~$$p = 2,21 \cdot 10^{-4} \cdot 0,0068^{45} \cdot 50^2 = 0$$~~

~~$$p = 2,21 \times 10^{-4} \cdot g^{45} \cdot 50^2 = 0,75$$~~

~~$$g^{45} = 1,3574 \dots$$~~

~~$$g = \sqrt[45]{1,3574 \dots} = 1,0068$$~~

~~percentage~~

~~$$45 \times 0,68 = 30,6 \%$$~~

~~$$g = 0,306$$~~

~~$$p = 2,21 \times 10^{-4} \cdot 0,306^{45} \cdot 50^2 = 3,979 \times 10^{-24}$$~~

~~$$\text{totaal} = p \times 40 = 3,979 \times 10^{-24} \times 40 = 1,59 \times 10^{-22}$$~~

1 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1

$$2. ((x^2 - 2)^2 - 5)^2 = 1$$

$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$

~~$$(-5x^4 + 20x^2 - 20)^2 = 1$$~~

~~$$-5x^4 + 20x^2 - 20 = \sqrt{1}$$~~

~~$$-5x^4 + 20x^2 - 20 = 1$$~~

~~$$-5x^4 + 20x^2 - 20 = 1$$~~

~~$$x^4 - 4x^2 + 4 = 1$$~~

~~$$x^4 - 4x^2 + 4 - 1 = 0$$~~

~~stel $x^2 = u$~~

~~$$u^2 - 4u + 3 = 0$$~~

~~$$(u-3)(u-1) = 0$$~~

~~$$u = 3 \vee u = 1$$~~

~~$$x^2 = 3 \vee x^2 = 1$$~~

~~$$x = \sqrt{3} \vee x = -\sqrt{3} \vee x = 1 \vee x = -1$$~~

Grotere Windmolens A

h = ashoogte (scharnier \rightarrow windgrond)

D = rotordiameter (peinde wijk \rightarrow scharnier)

P = vermogen in MW

Vermogen $\cdot 1,0068$ per mm h

$$g = 1,0068$$

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D$$

40 molens, 0,75 vermogen, $D = 50$ en $h = 45$

€25000 per h $\hookrightarrow h = 0,9D$

$$\text{totale vermogen} = 0,75 \cdot 40 = \underline{30}$$

$$h = 0,9D \text{ dus } D = \frac{h}{0,9} = \frac{10}{9}h = 1\frac{1}{9}h$$

$$\text{dus } 30 = 2,21 \cdot 10^{-4} \cdot g^h \cdot \left(\frac{10}{9}h\right)^2 \quad 3 \text{ per windmolen}$$

$$30 = 2,21 \cdot 10^{-4} \cdot 1,0068^h \cdot \left(\frac{10}{9}h\right)^2$$

$$3 = 2,21 \cdot 10^{-4} \cdot 1,0068^h \cdot \left(\frac{10}{9}h\right)^2$$

$$\text{vul in in GR: } y_1 = 2,21 \cdot 10^{-4} \cdot 1,0068^x \cdot \left(\frac{10}{9}x\right)^2$$

$$y_2 = 3$$

optie intsect geeft $x = 79,96938818$ per windmolen

$$79,969 \dots \cdot 25000 = 1999239,705 \approx 2 \text{ miljoen}$$

$$10 \text{ windmolens is dan } 1999239, \dots \cdot 10 = 200 \text{ miljoen}$$

Extra vragen (i) 1. 3^{75} , want $\times 3$ wordt veel snel groter dan $\times 2$

$$2. ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^2 - 2)^2 - 5 = 1 \quad \vee \quad (x^2 - 2)^2 - 5 = -1$$

$$\cancel{(x^2 - 2)^2} = 6 \quad \vee \quad (x^2 - 2)^2 \cancel{+ 4} = 4$$

4

2

6 pt

Extra vragen (2)

8A

$$(x^2-2)^2=6 \quad \vee \quad (x^2-2)^2=4$$

$$x^2-2=\sqrt{6} \vee x^2-2=-\sqrt{6} \vee x^2-2=2 \vee x^2-2=-2$$

$$x^2=\sqrt{6}+2 \vee x^2=-\sqrt{6}+2 \vee x^2=4 \quad \vee \quad x^2=0$$

~~$x = \sqrt{6} + 2$~~
 $x = -\sqrt{6} + 2$
5

$x = \sqrt{6} + 2$
1

$x = \sqrt{-\sqrt{6}-2}$
wortel mag niet negatief,
dus kan niet

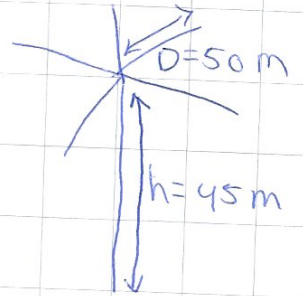
$x = 2$
2

$x = -2$
3

$x = 0$
4

dus er zijn ~~4~~ 5 oplossingen

40 windmolens



$$1 \text{ windmolen: } P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$P = 0,75 \text{ MW} = 0,75 \cdot 10^9 \text{ W}$$

$$h = 45 \text{ m}$$

$$D = 50 \text{ m}$$

$$g = x$$

$$\text{dus } 2,21 \cdot 10^{-4} \cdot x^{45} \cdot 50^2 = 0,75 \cdot 10^9$$

$$x^{45} = 1357466063$$

$$x = \sqrt[45]{1357466063} \approx 1,595 \dots \ln(3) = 1,1 \dots$$

$$\text{dus } g \approx 1,5956 \dots$$

geen idee verder

opt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 2^{100}

2. $((x^2 - 2)^2 - 5)^2 = 1$

$$((x^2 - 2)^2 - 5)((x^2 - 2)^2 - 5) = 1$$

$$(x^2 - 2)^4 - 10(x^2 - 2)^2 + 25 = 1$$

$$(x^2 - 2)^4 - 10(x^2 - 2)^2 = -24$$

$$(x^2 - 2)^4 - 10(x^2 - 2)^2 + 24 = 0$$

Stel $(x^2 - 2)^2 = u$

$$u^2 - 10u + 24 = 0$$

$$(u - 6)(u - 4) = 0$$

$$u - 6 = 0 \vee u - 4 = 0$$

$$u = 6 \vee u = 4$$

$$(x^2 - 2)^2 = 6 \vee (x^2 - 2)^2 = 4$$

$$x^2 - 2 = \sqrt{6} \vee x^2 - 2 = -\sqrt{6} \vee x^2 - 2 = \sqrt{4} \vee x^2 - 2 = -\sqrt{4}$$

$$x^2 = \sqrt{6} + 2 \vee x^2 = -\sqrt{6} + 2 \vee x^2 = 2 \vee x^2 = 0$$

vermogen nu = $P = 2,27 \cdot 10^{-7} \cdot 0,0068^{0,9 \cdot 56} \cdot 50^2 = 1,609107079 \cdot 10^{-98}$, ~~waarschijnlijk~~ verder,
 dus totaal = dit $\cdot 90 =$

vermogen nu = $90 \cdot 0,75 = 30 \text{ MW}$.
 dus nu twee $u =$

2 $10 \cdot 2,27 \cdot 10^{-9} \cdot 0,0068^{0,9 \cdot x} \cdot x^2 = 30$

$2,27 \cdot 10^{-3} \cdot 0,0068^{0,9x} \cdot x^2 = 30$

$0,0068^{0,9x} \cdot x^2 = 13579,66$

~~het is 3. dan~~ ?

2pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$2^{100}$$

1. 3^{75} , steeds $\cdot 3$ doen, levert $1,5$ keer zo veel op als $\cdot 2$ doen, dus hoewel je het $100 \cdot 2$ doet, is de $3 \cdot 1,5$ keer zo effectief, zo nu je ook op het eindresultaat.

2. ~~1/2~~ 2, het antwoord, en iets uit de macht 0.



$$G = \boxed{0,0068} \text{ , geeft}$$

oude situatie : $h = 45$ $D = 50$

$$\rightarrow p = 2,21 \cdot 10^{-9} \cdot \boxed{0,0068}^{45} \cdot 50^2 =$$

$$p = 0,75 \text{ MW}$$

totale vermogen is $0,75 \cdot 40 = \underline{30 \text{ MW}}$
 van 40 naar 10, dus het vermogen per
 windmolen moet 4x zo groot worden.
 Dat is $0,75 \cdot 4 = 3,00 \text{ MW}$.

$$p = 3,00 \text{ MW}$$

$$h = 0,80$$

$$D = 1,111 \dots h$$

$$\hookrightarrow P = 2,21 \cdot 10^{-9} \cdot g^{0,80} \cdot (1,111 \dots h)^2$$

1 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$1) 3^{75}$$

$$2) ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)(x^4 - 4x^2 + 4 - 5) = 1$$

8 verschillende oplossingen

$$1 \quad P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

Oude windmolens

$$P = 0,75 \text{ MW}$$

$$D = 50 \text{ m}$$

$$h = 45 \text{ m}$$

$$g = 0,68$$

$$h = 0,9D$$

$$\text{totale vermogen (minstens)} = 0,75 \times 40 = 30 \text{ MW}$$

dus per nieuwe windmolen, minstens

$$1 \quad 3 \text{ MW}$$

~~2 MW = 4 x 'oude' vermogen, dus
alles keert op?~~

1 pt

$$D = \cancel{50 \times 4} = \cancel{200} \text{ m} \quad h = 0,9g \times \frac{200}{4} = 9g = 180$$

$$h = \cancel{4 \times 45} = \cancel{180} \text{ m}$$

$$g = \cancel{2,72}$$

$$2 \quad 1. \quad (x^2 - 2)(x^2 - 2)$$

$$2. \quad x^4 - 2x^2 - 2x^2 + 4$$

$$x^4 - 4x^2 + 4$$

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2) - 5)^2 = 1$?

1. ~~2~~ 3^{75}

2. ~~$((x^2 - 2) - 5)^2 = 1$~~

~~$((x^2 - 4x^2 + 4) - 25)^2 = 1$~~

~~6000~~ 205

①

P=

$$3 \quad 0,75 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9D} \cdot D^2$$

$$\text{stel } y_1 = 0,75$$

$$y_2 = 2,21 \cdot 10^{-4} \cdot 1,0068^{0,9x} \cdot 10^2 x^2$$

Intsect geeft $x = 50,01 \dots$

$$\text{Dus } D = 50,01 \dots$$

1

$$h = 0,9 \cdot 50,01422064 = 45,012 \dots$$

1

$$45,012 \dots \times 25000 \times 10 = \approx \text{€ 1 miljoen}$$

5pt

②

1. 2^{100} , $100 > 75$, dus ik denk dat het uiteindelijk dan groter wordt ongeacht dat $3^{75} > 2$

$$2. \quad x^4 - 4x^2 -$$

40 stuks

P = 0,75 elk

rd = 50m = D

ah = 45m = h

P = 0,75

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 2^{100} , $100 > 75$, dus ik denk dat het uiteindelijk groter wordt ongeacht dat $3 > 2$
2. ik denk 8, want er zijn ^{dingen} 3 tot de macht 2, dus $2^3 = 8$

40 molens

19A

$$P = 0,75 \text{ MW}$$

$$D = 50,0 \text{ m}$$

$$h = 45,0 \text{ m}$$

10 nieuwe

25.000,- p/h

$$h = 0,9D$$

$$(P_{nu} = 2,21 \cdot 10^{-4} \cdot 0,0068^{45} \cdot 50^2) \times$$

$$P_{nu} = 0,75 \text{ MW} \cdot 40 = 30 \text{ MW}$$

$$0,75 = 2,21 \cdot 10^{-4} \cdot 0,68 \cdot D^2$$

25.000 Per meter

$$45 \cdot 25.000 = 1.125.000$$

$$1.125.000 \times 40 = 45 \text{ miljoen.}$$

opt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$1. 3^{75} \quad \vee \quad 2^{100} = 3^3 \quad \vee \quad 2^4 \rightarrow 27 \quad \vee \quad 16 \quad \text{dus } 3^{75} \text{ is groter}$$

$$2. ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$x^8 - 4x^4 + 1 = 1$$

$$x^8 + 16x^4 = 0$$

$$x^4(x^4 + 16) = 0$$

$$x^4 = 0 \vee x^4 = -16$$

$$x^4 = 0 \vee x^4 = 4$$

~~Totale vermogen eerst: 40 x 0,75 MW = 30 MW
 30 MW : 10 nieuwe windmolens = 3 MW per windmole~~

Totale vermogen eerst: $40 \times 0,75 \text{ MW} = 30 \text{ MW}$

1p 30 MW : 10 nieuwe windmolens = 3 MW per windmole

met $P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$

$$P = 30 \text{ MW} = 30 \cdot 10^6 \text{ W}$$

$g^h = ?$

$$D = \frac{h}{0,9}$$

$$\text{geeft } g^h = \frac{30 \cdot 10^6}{2,21 \cdot 10^{-4} \cdot D^2} = \frac{30 \cdot 10^6}{2,21 \cdot 10^{-4} \cdot \left(\frac{h}{0,9}\right)^2}$$

1p4

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$1) 3^{75} \vee 2^{100} = {}^3\log(75) \vee {}^2\log(100)$$

$${}^3\log(25) + {}^3\log(3) \vee {}^2\log(25) + {}^2\log(4)$$

$$2) \cancel{((x^2 - 2)^2 - 5)^2 = 1}$$

$$((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$(x^4 - 4x^2 - 1)^2 = 1$$

$$x^0 + 16x^4 + 1 = 1$$

$$x^0 + 16x^4 = 0$$

$$x^4(x^4 + 16) = 0$$

$$x = 0 \vee x^4 = -16$$

kan niet

2

kleinste x voor $\log(x)$, dus 3^{75} groter.

Dus er is 1 reële oplossing en dat is $x=0$

$P =$ MW

Lashoogte

+1 meter = + 0,6 d'v. P toe

45 m

€25000 per m ashoogte

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D$$

$$g = g_{\text{per m}} \rightarrow 0,0068$$

$$D = 50 \text{ m}$$

$$h = 0,9 D$$

$$P = 0,75 \text{ MW}$$

ashoogte = ?

$$g = ?$$

$$P = 2,21 \cdot 10^{-4} \cdot g^{0,9} \cdot 50$$

$$1 \quad 0,75 \cdot 10^{-6} = 2,21 \cdot 10^{-4} \cdot g^{0,9} \cdot 50$$

$$0,75 \cdot 10^{-6} = 0,01105 \cdot g^{0,9}$$

$$6,7 \cdot 10^{-5} = g^{0,9}$$

$$0,6 =$$

$$\frac{8^3}{2} = 8$$

$$2^3 = 8$$

$$\sqrt[3]{\log(8)} = 3$$

1 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. $3^{75} \Rightarrow 3 \cdot 3 = 9$ je komt sneller op een hoger getal dan bij bijv. $2 \cdot 2 = 4$ enz. als je zo het rijtje af gaat ben je sneller hoger uitgekomen ~~na~~ bij 3^{75} .

2. ik denk ~~1~~ ², ~~1~~ ~~2~~ aangezien er een x^2 en een x komt zou je het in haakjes moeten zetten.

Na ~~om~~ de haakjes uitwerken houd je 2 antwoorden over

ashoogte h

rotor diameter D / P (extra hoogte 0,6d % toename)

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$P = 0,75 \text{ MW}$$

$$g = x$$

$$h = 45$$

$$D = 50$$

$$h = 0,9D$$

$$g^h = P / 2,21 \cdot 10^{-4} \cdot D^2$$

$$g = \sqrt[h]{P / 2,21 \cdot 10^{-4} \cdot D^2}$$

$$g = \sqrt[45]{0,75 / 2,21 \cdot 10^{-4} \cdot 50^2} = 1,006 \dots$$

$$P_{\text{total}} = 40 \times 0,75 = 30 \text{ MW}$$

$$h = 0,9D$$

$$h = 0,9 \cdot 50$$

$$h = 45$$

$$30 = 2,21 \cdot 10^{-4} \cdot 1,006^h \cdot 50^2$$

$$\frac{30}{2,21 \cdot 10^{-4} \cdot 50^2} = 1,006^h$$

$$54,298 \dots = 1,006^h$$

$$1,006^h = 54,298$$

$$h = \frac{\log(54,298)}{\log(1,006)}$$

$$h = 667,74 \dots$$

$$667,74 \times 10 \times 25000 = 166935707,0$$

$$\text{dus } 166.935.000 \text{ euro}$$

opt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 2^{100} omdat je uit eindelijk het grote getal nog vaak maal 2 doet

$$2. ((x^2 - 2)^2 - 5)^2 = 1$$

dus 5

Opdracht 1 versie A

18A

h = as hoogte

D = rotor diameter

P = vermogen in MW

~~40 windmolens~~

~~\hookrightarrow elk 0,75 MW~~

~~$D = 50$ m~~

~~$h = 45$ m~~

25000 € / m ashoogte

~~$\frac{P}{h} = 0,9$~~ $h = 0,9 D$



Opdracht 4 versie A

18A (5)

$$0,75 = (2,21 \cdot 10^{-4}) \cdot g^{45} \cdot 50^2$$

$$(0,75 / (2,21 \cdot 10^{-4}) / 50^2) = g^{45}$$

$$\sqrt[45]{(0,75 / (2,21 \cdot 10^{-4}) / 50^2)} = g$$

$$P_{\text{totaal}} = 40 \cdot 0,75 = 30 \text{ MW}$$

$$P_{\text{windroten}} = 30 / 10 = 3 \text{ MW} / \text{Windaarde}$$

$$P = (2,21 \cdot 10^{-4}) \cdot \left(\sqrt[45]{(0,75 / (2,21 \cdot 10^{-4}) / 50^2)} \right)^3 \cdot D^2$$

$$3 = (2,21 \cdot 10^{-4}) \cdot \left(\sqrt[45]{(0,75 / (2,21 \cdot 10^{-4}) / 50^2)} \right)^3 \cdot D^2$$

~~$$3 = (2,21 \cdot 10^{-4}) \cdot \left(\sqrt[45]{(0,75 / (2,21 \cdot 10^{-4}) / 50^2)} \right)^3 \cdot D^2$$~~

$$\frac{3}{\left(\sqrt[45]{(0,75 / (2,21 \cdot 10^{-4}) / 50^2)} \right)^3} =$$

$$D^2$$

~~$$D = \frac{h}{0,9}$$~~

Opdracht 1 versie A

18A

6

3

$$(2,21 \cdot 10^{-4}) \left(\frac{45}{\sqrt{(0,25 / (2,21 \cdot 10^{-4}) / 50^2)}} \right)^2 = \frac{45^2}{0,9}$$

Vul in op GR

$$Y_2 = \frac{3}{(2,21 \cdot 10^{-4}) \left(\frac{45}{\sqrt{(0,25 / (2,21 \cdot 10^{-4}) / 50^2)}} \right)^2}$$

$$Y_2 = \left(\frac{x}{0,9} \right)^2$$

option intersect geeft

$$x = 29,9$$

$$x \rightarrow A$$

$$A \cdot 25000 = 1998319,187$$

~~een windmolen kost ongeveer 2 miljoen~~

$$Ans = 10 = 1998319,187$$

het wind De minimale investering is 20 miljoen

7pt

2.1

~~$$2^{100} = (2^4)^{25} = (16)^{25}$$

$$(1024)^5$$~~

$$3^{25} = (3^3)^{25} = (27)^{25}$$

$$2^{100} = (2^4)^{25} = (16)^{25}$$

3^{75} is groter want $(3^3)^{25} > (2^4)^{25}$

2.2

~~$$((x^2 - 2)^2 - 5)^2 = 1$$~~

~~$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$~~

~~$$(x^4 - 4x^2 - 1)^2 = 1$$~~

~~$$x^8 - 8x^6 - 2x^4 + 16x^4 + 8x^2 + 1 = 1$$~~

~~$$x^8 - 8x^6 + 14x^4 + 8x^2 = 0$$~~

~~$$x^2(x^6 - 8x^4 + 14x^2 + 8) = 0$$~~

~~$$x^2 = 0 \vee x^6 - 8x^4 + 14x^2 + 8 = 0$$~~

~~$$x = 0$$~~

~~$$\text{Stel } x^2 = u$$~~

~~$$u^4 - 8u^3 + 14u^2 + 8u = 0$$~~

Opdracht 2

Ⓟ

18A

~~2.1~~

$$((x^2 - 2)^2 - 5)^2 = 1$$

stel $(x^2 - 2)^2 - 5 = u$

2.2

$$u^2 = 1$$

$$u = 1 \vee u = -1$$

$$(x^2 - 2)^2 - 5 = 1 \vee (x^2 - 2)^2 - 5 = -1$$

$$(x^2 - 2)^2 = 6 \vee (x^2 - 2)^2 = 4$$

$$x^2 - 2 = \sqrt{6} \vee x^2 - 2 = -\sqrt{6} \vee x^2 - 2 = 2 \vee x^2 - 2 = -2$$

~~$x^2 = \sqrt{6} + 2$~~

$$x^2 = \sqrt{6} + 2 \vee x^2 = 2 - \sqrt{6} \vee x^2 = 0 \vee x^2 = 4$$

Uold.

Uold. Niet

Uold.

Uold.

$$x = \sqrt{\sqrt{6} + 2}$$

Uold.

$$\vee x = -\sqrt{\sqrt{6} + 2}$$

Uold.

$$\vee x = 0$$

Uold.

$$\vee x = 2$$

Uold.

$$\vee x = -2$$

Uold.

er zijn

5 reële oplossingen want

$$x = \sqrt{\sqrt{6} + 2}$$

$$\vee x = -\sqrt{\sqrt{6} + 2}$$

$$\vee x = 0$$

$$\vee x = 2$$

$$\vee x = -2$$

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$g = 1,0060$$

$$0,75 = 2,21 \cdot 10^{-4} \cdot g^{0,9D} \cdot D^2$$

$$g^{0,9D} \cdot D^2 = 3393,6665$$

$$1,0060^{0,9} \cdot 1,0060^D \cdot D^2 = 3393,665$$

$$1,0061 \cdot 1,0060^D \cdot D^2 = 3393,665$$

$$1,0060^D \cdot D^2 = 3373,009$$

$$D^2 = \frac{0,75}{2,21 \cdot 10^{-4} \cdot g^h}$$

$$D = \sqrt{\frac{0,75}{2,21 \cdot 10^{-4} \cdot g^h}}$$

$$h = 0,9 \left(\sqrt{\frac{0,75}{2,21 \cdot 10^{-4} \cdot g^h}} \right)$$

$$h = 0,9 \left(\sqrt{3393,665 \cdot g^h} \right)$$

$$h = 0,9 \left(\sqrt{3393,665} \cdot \sqrt{g^h} \right)$$

$$h = 0,9 \left(58,255 \cdot \sqrt{g^h} \right)$$

$$h = 0,9 \left(58,255 g^{1/2h} \right)$$

$$= 52,42943 \cdot 1,0060^{-1/2h}$$

$$= 52,253 \cdot 1,0060^h$$

2pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ? $75 \cdot 3$ $100 \cdot 2$
 $75 \cdot 75 \cdot 75$ $100 \cdot 100$

375

2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking

$$((x^2 - 2)^2 - 5)^2 = 1$$

8

~~$(x^2 - 2)^2 - 5 = \sqrt{1}$~~

$$x^4 - 4x^2 + 4 - 5 = \sqrt{1} = 1$$

$$x^4 - 4x^2 - 1 = 1$$

$$x^4 - 4x^2 - 2 = 0$$

h = ashoogte (m)
 D = rotor diameter (m)
 P = vermogen (MW)
 \blacksquare +1 meter $h = +0,68\% P$

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

g = groeifactor per meter h

1. 40 windmolens en $P_{\text{windmolen}} = 0,75 \text{ MW}$

$$P_{\text{totaal}} = 40 \times 0,75 = 30 \text{ MW}$$

2. 10 windmolens en $P_{\text{windmolen}} = ?$

! $P_{\text{totaal}} = 30 \text{ MW}$ geeft $P_{\text{windmolen}} = \frac{30}{10} = 3,0 \text{ MW}$

Nieuwe windmolen = 2500 € per meter ashoogte

Verhouding $h = 0,9D$

Er geldt dat $2,21 \cdot 10^{-4} \cdot g^h \cdot D^2 = 30 \text{ MW}$

Ook geldt er dat $g = \blacksquare 1,068$

$$2,21 \cdot 10^{-4} \cdot 1,068^h \cdot D^2 = 30 \text{ MW}$$

$$1,068^h \cdot D^2 = \blacksquare \text{ MW}$$

als $h = 1$ geldt $1 = 0,9D$

$$D = 1,11$$

$$P = 2,21 \cdot 10^{-4} \cdot 1,068^1 \cdot \blacksquare 1,11 \blacksquare^2 = 2,7 \cdot 10^{-4} \text{ MW}$$

$$\frac{30}{2,7 \cdot 10^{-4}} = 111 \cdot 10^3 \text{ m ashoogte dus}$$

$$111 \cdot 10^3 \cdot 25000 = 2777 \cdot 10^6 \text{ €}$$

2 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
 3^{75} , hoe groter de macht wordt, hoe belangrijker het grondtal wordt.
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?
 en exponent

$$\begin{aligned}
 (x^2 - 2)^2 &= (x^4 - 4x + 4 - 5)^2 \\
 &= (x^4 - 4x - 1)^2 \\
 &= x^8 - 4x^5 - x^4 - 4x^5 + 16x^2 + 4x - x^4 + 4x + 1 \\
 &= x^8 - 8x^5 - 2x^4 + 16x^2 + 8x + 1 = 1
 \end{aligned}$$

Ik denk 2 reële uitkomsten, aangezien dit het meest voorkomt bij kwadratische vergelijkingen

$$P = 2,21 \cdot 10^{-4} \cdot q^h \cdot D^2$$

$$40 \cdot 0,75 \cdot 10^6 = 30 \text{ MW}$$

$$q = 1,0068$$

$$h = 0,9 \text{ D}$$

$$30 \cdot 10^6 = 2,21 \cdot 10^{-4} \cdot 1,0068^h \cdot D^2$$

1 stel $h = x$ geeft $0,9 \cdot D = x$
 $D = 1 \frac{1}{9} X$

$$30 \cdot 10^6 = 2,21 \cdot 10^{-4} \cdot 1,0068^x \cdot \left(1 \frac{1}{9} X\right)^2$$

$$30 \cdot 10^6 = 2,21 \cdot 10^{-4} \cdot 1,0068^x \cdot 1 \frac{19}{81} X^2$$

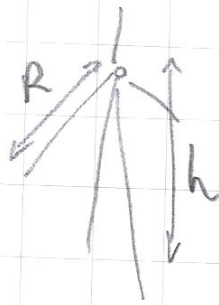
$$30 \cdot 10^6$$

$$\frac{30 \cdot 10^6}{2,21 \cdot 10^{-4}} = 1,0068^x \cdot 1 \frac{19}{81} X^2$$

$$30 \cdot 10^6$$

$$\frac{19}{81} X^2 \cdot 2,21 \cdot 10^{-4} = 1,0068^x$$

$$X = \frac{1,0068}{\log\left(\frac{30 \cdot 10^6}{1 \frac{19}{81} X^2 \cdot 2,21 \cdot 10^{-4}}\right)}$$



MW
ans. $(\Delta h \cdot 0,0068)$

$$P = 2,21 \cdot 10^{-4} \cdot 0,0068^h \cdot D^2$$

40 windmolens

$$P = 0,75$$

10 windmolens 25000 h

$$h = 0,9D$$

1 $40 \cdot 0,75 = 30$ $\frac{30}{10} = 3 \text{ MW per nieuwe windmolen}$
 $P = 3$ ${}^3 \log(9) = 3$

$$3 = 2,21 \cdot 10^{-4} \cdot 0,0068^h \cdot D^2$$

$$\frac{3}{2,21 \cdot 10^{-4}} = 0,0068^h \cdot D^2$$

$$0,0068^h = \frac{13574,66}{D^2}$$

$$h = \frac{13574,66}{D^2} \cdot \log\left(\frac{D^2}{13574,66}\right)$$

2

$$h \cdot 25000 \cdot 10 =$$

er zal dus ... erwa geïnvesteerd moeten worden.

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 3^{75} neemt elke keer met 300% toe,
 $300 \cdot 75 = 30000 - 7500 = 22500$
 2^{100} neemt steeds 200% dus ~~nu~~ $200 \cdot 100 = 20000$
 $22500 > 20000$, aangezien je bij 3^{75} ook nog eens met een groter getal begint is 3^{75} groter.

$$2. ((x^2 - 2)^2 - 5)^2 = 1 \quad \frac{(x^2 - 2)^2 - 5}{\sqrt{|x^2 + 2| - 5}} + 1 = 0$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$\cancel{x^8} - \cancel{4x^8} + \cancel{4x^4} - \cancel{5x^4} - \cancel{4x^8} + \cancel{16x^4} + 16 + 25$$

$$\cancel{4x^4} - 16x^2 + 16 - 20 - \cancel{5x^4} + 20x^2 - 20 + 25 = 1$$

$$-7x^8 + 104x^4 + 4x^2 - 18 = 0$$

$$7x^8 - 14x^4 - \frac{4}{2}x^2 - 18 = 0$$

$$x^8 - 2x^4 - \frac{4}{7}x^2 - 2\frac{4}{7} = 0$$

$$x^2(x^6 - 2x^2 - \frac{4}{7}x) = 2\frac{4}{7}$$

$$1 \text{ pt} \quad 40 \cdot 0,75 = 30 \text{ MW} / 10 = 3 \text{ MW per stuk}$$

$$g = 0,0068 \quad h = 0,9D$$

$$P = 2,21 \cdot 10^{-4} \cdot 0,0068^{0,9D} \cdot D^2$$

$$0,0068^{0,9D} \cdot D^2 = 30 / (2,21 \cdot 10^{-4})$$

$$D = -2,266 \quad \text{or} \quad \text{135} \rightarrow 46$$

?

0

2 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$(x^2 - 2)^2 - 5 = 1$$

$$(x^2 - 2)^2 = 6$$

$$x^2 - 2 = \sqrt{6}$$

$$x^2 = \sqrt{6} + 2$$

$$x = \sqrt{\sqrt{6} + 2} \vee x = -\sqrt{\sqrt{6} + 2} \quad \leftarrow 2$$

$$2^2 = 4 \quad 3^2 = 9$$

~~$$2^6 = 64 \quad 3^6 = 729$$~~

$$2^4 = 16 \quad 3^4 = 81$$

$$1\frac{1}{3} \cdot 75 = 100$$

$$1\frac{1}{3} \cdot 2 = \frac{8}{3} \quad 1\frac{1}{3} \cdot 4 < 9$$

hoogte schommelpunt
wieltop
↑
grond

cirkel van wielen diameter
van cirkel van wielen
↑

6B

ashogte: h , rotordiameter: D

g : groei factor per extra meter ashogte

P : max elektriciteit dat windmolen kan produceren in mega watt / MW

voor elke extra m ashogte \rightarrow vermogen neemt met 0,60% toe

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

~~oud~~ oud: 40 windmolens met ^{elk} vermogen 0,75 MW

nieuw: 70 windmolens, €25000 per m ashogte, $h = 0,9 D$

$$\text{oud totaal} = 0,75 + 40 = 30 \text{ MW}$$

nieuw moet dus $30 : 70 = 3$ MW per windmolen

$$P = 2,21 \cdot 10^{-4} \cdot g^{0,9 D} \cdot D^2$$

1 pt
2 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

(Dit heb ik een beetje anders gedaan)
 (de rechter ben in twee)

1. 2^{100} , het aantal er zijn 100-st en wordt $100 \times$ vermenigvuldigt. Telkens wordt het vermenigvuldigt met 2. De laatste 2^{100} heeft 25 stappen meer dan 3^{75} , en zeker die laatste stappen zijn belangrijker. Dan worden namelijk grotere getallen verdubbelt.

2. $((x^2 - 2)^2 - 5)^2 = 1$

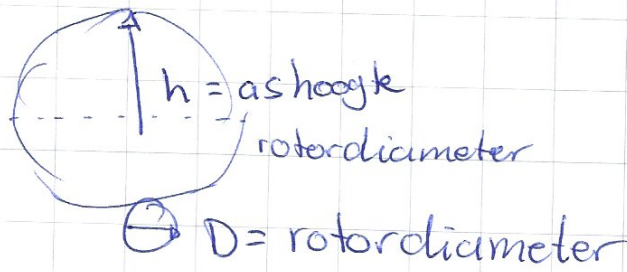
$$(x^2 - 2) - 5 = \sqrt{1} \quad \vee \quad (x^2 - 2) - 5 = -\sqrt{1}$$

$$x^2 - 2 = \sqrt{1} + 5 \quad \vee \quad x^2 - 2 = -\sqrt{1} + 5$$

$$x^2 = \sqrt{1} + 7 \quad \vee \quad x^2 = -\sqrt{1} + 7$$

~~2~~
~~2~~

het heeft twee oplossingen



P word uitgedrukt in MegaWatt
 $h + 1 \text{ meter} = P (\text{vermogen}) \cdot 1,68$

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$= 2,21 \cdot 10^{-4} \cdot 1,68^h \cdot D^2$$

40 windmolens elk 0,75 MW

10 windmolens elk $4 \cdot 0,75 = 3,00 \text{ MW}$

€25.000 per stuk per m ashoogte (h)

$$h = 0,9 D$$

$$P_1 = 2,21 \cdot 10^{-4} \cdot 1,68^{0,9D} \cdot D^2 = 0,75 \text{ MW}$$

$$P_2 = 2,21 \cdot 10^{-4} \cdot 1,68^{0,9D} \cdot D^2 = 3,00 \text{ MW}$$

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. $2^4 = 16$
 $2^6 = 64$
 $2^{100} ?$

$3^3 = 27$
 $3^5 = 27 \cdot 3 \cdot 3 = 81 \cdot 3 = 243$

2. $((x^2 - 2)^2 - 5)^2 = 1$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$(x^4 - 4x^2 - 1)^2 = 1$$

$$x^8 - 4x^8 - x^4 - 4x^8 + 16x^4 + 4x^2 - x^4$$

$$+ 4x^2 + 1 = 1$$

$$x^{16} - 8x^8 + 14x^4 + 8x^2 + 1 = 1$$

? 2.

1) ashoogte = h
 rotordiameter = D
~~rot.~~ vermogen = P (MW) (10^6)

extra meter ashoogte \rightarrow vermogen met 0,68% toeneemt

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

g = groefactor per extra meter ashoogte

40 windmolens met $P = 0,75$ MW \rightarrow oud

10 windmolens met $P = 0,75$ MW \rightarrow nieuw

25000 per meter ashoogte

$$h = 0,90$$

$$g = 1,0068$$

Hoe duur in miljoenen

$$0,75 = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

end

$$0,0751878 = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$| \quad 0,0793879 = 2,21 \cdot 10^{-4} \cdot g^{0,90} \cdot D^2$$

$$| \quad 0,0751875 = 2,21 \cdot 10^{-4} \cdot 1,0068^0 \cdot D^2$$

in GR geht $D = 17,36897732$

$$25000 \cdot 17,36897732 = 434224$$

das 0,42 miljoen

2 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$1) \quad \begin{array}{l} 3^4 = 3 \cdot 3 \cdot 3 \cdot 3 = 81 \\ 2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16 \end{array} \quad \left| \quad \begin{array}{l} \text{exponentieel groeit } 3 \text{ sneller, dus} \\ 3^{75} \text{ is groter dan } 2^{100} \end{array} \right.$$

$$2) \quad ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^2 - 2)^2 = \sqrt{1} \quad \vee \quad (x^2 - 2)^2 = -\sqrt{1}$$

$$x^2 - 2 = \sqrt{1} \quad \vee \quad x^2 - 2 = -\sqrt{1} \quad \vee \quad x^2 - 2 = \sqrt{-1} \quad \vee \quad x^2 - 2 = -\sqrt{-1}$$

kan niet

kan niet

$$x^2 = 2 + \sqrt{1} \quad \vee \quad x^2 = 2 - \sqrt{1}$$

$$x = \sqrt{2 + \sqrt{1}} \quad \vee \quad x = -\sqrt{2 + \sqrt{1}} \quad \vee \quad x = \sqrt{2 - \sqrt{1}} \quad \vee \quad x = -\sqrt{2 - \sqrt{1}}$$

dus 4

ashoogte = h

rotordiameter = D

↳ diameter van cirkel

vermogen = $p = \text{MW}$

vermogen neemt met 0,68% toe

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

40 windmolens → 0,75 MW

10 windm.

25000,- per meter

$$h = 0,19 D$$

$$g = 100 + 0,68 = 100,68 = 1,0068$$

$$30 = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$3 \quad 135746,6063 = \text{naar}^8 1,0068^{0,19D} \cdot D^2$$

3 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 2^{100} , want dat heeft meer getallen

2. 1 oplossing

~~van~~ $((x^2 - 2)^2 - 5)^2 = 1$

$$((1^2 - 2)^2 - 5)^2 = ((1^2 - 2)(1^2 - 2) - 5)^2$$

$$((1 - 2)(1 - 2) - 5)^2$$

$$((1 - 2 - 2 + 4) - 5)^2$$

$$(1 - 5)^2 = (1 - 5)(1 - 5)$$

$$1 - 5 - 5 + 25 = 16$$

$$(x^2 - 2)^2 - 5 = 1 \quad (x^4 - 4x - 1)(x^4 - 4x - 1)$$

$$(x^2 - 2)(x^2 - 2)$$

$$x^4 - 2x^2 - 2x^2 + 4$$

$$x^4 - 4x^2 + 4$$

$$\text{stel } p = x^2$$

$$x^8 - 4x^5 - 4x^4 + 4x^5 + 16x^2 - 4x^4 - 4x + 4$$

$$x^8 - 8x^5 + 16x^2 - 2x^4 + 1$$

~~$$(p^2 - 4x + 4 - 5)^2 = (p^2 - 4p - 1)(p^2 - 4p - 1)$$~~

~~$$p^2 - 4p + 16$$~~

$$p = 2,21 \cdot 10^{-4} \cdot 0,0068^h \cdot D^2 \quad h = 0,90$$

$$40 \cdot 0,75 = 30 \text{ MW windpark I}$$

$$\text{windpark II } 30 / 10 = 3,0 \text{ MW per windmolens}$$

$$3,0 = 2,21 \cdot 10^{-4} \cdot 0,0068^{0,90} \cdot D^2$$

$$0,0068^{0,90} \cdot D^2 = \frac{3,0}{2,21 \cdot 10^{-4}}$$

$$0,0068^{0,90} \cdot D^2 = 13574,66063$$

$$\text{vul in } y_1 = 0,0068^{0,9x} \cdot x^2$$

$$y_2 = 13574,66063$$

optie intersect geeft $x = -1,845 \dots$

$$\text{dus } D = 1,845 \dots \text{ m}$$

$$h = 0,9 \cdot 1,845 \dots = 1,66 \dots \text{ meter}$$

$$25000 \cdot 1,66 \dots = \text{€}47527,527 \dots \text{ per windmolens}$$

$$47527,527 \cdot 10 \approx \text{€}0,4 \text{ miljoen}$$

5 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. ~~3~~ 2^{100} omdat de macht exponentieel ~~is~~ is en die is groter dus ik denk dat dat zwaarder weegt dan een verandering van een grondgetal
2. maar 1 omdat alleen 1 in het kwadraat 1 is.



Naam: _____

Vak: _____

Datum: _____ Klas: _____

Cijfer:

$$P = 2,21 \cdot 10^{-4} \cdot g^h \cdot D^2$$

$$0,75 \cdot 10^3 = 2,21 \cdot 10^{-4} \cdot 0,68^{0,9 \cdot D} \cdot D^2$$

$$3393,66 \dots = 0,68^{0,9 \cdot D} \cdot D^2$$

ik weet het niet. moet ik eerst D berekenen ??

$$3393,66 \dots = 0,68 D^{0,8 D}$$

$$4900,68 \dots = D^{1,8 D}$$

$$D = ?$$

$$h = 0,9 \cdot ? = 0,9 \cdot ?$$

$$25000 \cdot 0,9 \cdot ? \cdot 10 = \text{het antwoord ofzo}$$

2 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 3^{75} , 3 is groter dan 2 dus zal uiteindelijk wel meer zijn denk

~~ok~~

$$2. ((x^2 - 2)^2 - 5)^2 = 1$$

$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4)^2 - 25 = 1$$

$$x^4 - 4 \leftarrow \text{noe moet je verder?}$$

3 reële

↳ x oplossen

↳ haakjes wegwerken

↳ nog 1, maar die weet ik niet

$h =$ ashoogte $D =$ rotordiameter.
 $P =$ vermogen

$$g = \cancel{0,68} \cdot 1,068$$

$$40 \times 0,175 = 30 \text{ MW.} \quad h = 0,9D.$$

1 $30 : 10 = 3,0 \text{ MW per windmolen.}$

2 $3 = 2,21 \cdot 10^{-4} \cdot 0,68^{\frac{0,9 \cdot D}{h}} \cdot D^2.$

~~$3 = 2,21 \cdot 10^{-4} \cdot 0,68^{\frac{0,9 \cdot D}{h}} \cdot D^2.$~~

$$2,21 \cdot 10^{-4} \cdot \frac{1,068^h}{0,68^h} \cdot D^2 = 3.$$

3,1

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 3^{75} is groter omdat 3 een groter getal is en dus met grotere getallen omhoog gaat.

$$2. ((x^2 - 2)^2 - 5)^2 = 1.$$

$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$

$$(-5x^4 + 20x^2 - 20)^2 = 1.$$

$$(10x^8 - 100x^6 + 100x^4 - 100x^6 + 400x^4 - 400x^2 + 80x^4 - 400x^2 + 400) = 1.$$

$$(10x^8 - 200x^6 + 380x^4 - 800x^2 + 400) = 1.$$

$$(x^8 - 20x^6 + 38x^4 - 80x^2 + 40) = 1$$

$$x^8 - 20x^6 + 38x^4 - 80x^2 + 39 = 0.$$

5 oplossingen

$$40 \cdot 0,75 \text{ MW} = 30 \text{ MW energie}$$

$$/ 10 = 3 \text{ MW per windmolen}$$

$$q = 1,68$$

$$D^2 = P / (2,21 \cdot 10^{-4} \cdot 1,68) = 89,9 \text{ m}$$

$$\text{dus } h = 0,9 \cdot 89,9 = 80,9$$

$$25.000 \cdot 80,9 \cdot 10$$

$$\approx 20 \text{ miljoen euro}$$

3pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
~~75~~ 3^{75} is groter
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$? *2 oplossingen*

$$1 \quad 75 \cdot 75 = 5250 + 375 = 5625 \cdot 75$$

$$100 \cdot 100 = 10000$$

2 ~~20~~

aantal = 40 windmolens } $40 \times 0,75 = 30 \text{ MW}$
 $P = 0,75 \text{ MW per windmolen}$

nieuw = 10 windmolens

€25.000 per stuk

! $h = 0,190$

dus nieuwe windmolens moeten

~~30 MW~~ = ~~30~~ 30 MW omzetten

dat is 3,0 MW per windmolen

~~12 = 9~~
~~15 = 12~~
~~18 = 15~~
~~21 = 18~~

vermogen neemt met $2,25$ toe

ijkt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1) 3^{75} is groter, omdat het grondtal groter is en dat telt zwaarder mee dan de exponent

$$2) ((x^2 - 2)^2 - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4 - 5)^2 = 1$$

$$x^8 - 8x^2 + 16 - 25 = 1$$

$$x^8 - 8x^2 = 10$$

2 oplossingen, je hebt $x = \dots$ en $x = \dots$

$$0,75 = 2,21 \cdot 10^{-4} \cdot g^h \cdot 15^3$$

??
..

opt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
 2^{100} dat is een gokse
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

$$((x^4 - 2x^2 - 2x^2 + 4) - 5)^2 = 1$$

$$((x^4 - 4x^2 + 4) - 5)^2 = 1$$

$$(x^4 - 4x^2 + 4)^2 + 25 = 1$$

$$(x^4 - 4x^2 + 4)^2 = -24$$

geen oplossingen



Naam: _____

Vak: _____

Datum: _____ Klas: _____

Cijfer:

(Versie B opg 1)

g = groeifactor per extra meter arboogte.
elke extra meter neemt het vermogen met 0,68% toe.

g is dus: ~~1,068~~ 1,0068

$p = 0,75$

kosten: €25000,- per meter arboogte

$h = 0,90$

$P = 0,75 \text{ MW} \cdot 40 = 30 \text{ MW}$ in totaal per windmolen

10 windmolens moeten ook 30 MW

produceren: ~~30~~ $\frac{30}{10} = 3 \text{ MW}$ per windmolen

⚡

2pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 3^{75} . Het grondtal 3 is anderhalf keer zo groot als het andere grondtal 2. De macht moet dus ook anderhalf keer zo groot zijn bij 2^{100} . Dat is het niet.

2. 20 oplossingen. Na het uitbreiden van de eerste stap heb je de 25 naar rechts en dan komt er -24 uit van de rechterkant.

$$\frac{(x^4 - 2x^2 + 4 - 5)^2}{(x^4 - 2x^2 + 4 - 5)^2}$$

α_1 $h =$ ashoogte $D =$ rotor diameter

$$\Delta P = 0,6 \rho \Delta h$$

$$P = 2,21 \cdot 10^{-4} \cdot g^4 \cdot D^2$$

$$P = 2,21 \cdot 10^{-4} \cdot 1,006 \rho^4 \cdot D^2$$

30 MW

3 MW

2

 $h = 0,90$ geeft

$$Z = 2,21 \cdot 10^4 \cdot 1,006 \rho^4$$

$$D = 22,256$$

$$h = 79,970$$

$$79,970 \cdot € 25000,- \cdot 10 \approx 20.000.000,-$$

Dus

$$€ 20.000.000,-$$

222

7 pt

2 Extra vragen

De volgende vragen moeten zonder rekenmachine gemaakt worden. Licht je antwoord toe bij elke vraag.

1. Welke van de volgende twee getallen is groter: 3^{75} of 2^{100} ?
2. Hoeveel verschillende (reële) oplossingen heeft de vergelijking $((x^2 - 2)^2 - 5)^2 = 1$?

1. 3^{75} want $3^{\frac{3}{4}} > 2^1$

2. $((x^2 - 2)^2 - 5)^2 = 1$

$$(x^4 - 4x^2 + 9 - 5)^2 = 1$$

Met heeft 2 reële uitkomsten, dit lijkt mij logisch omdat er maar één x is en die is kwadraat dus een pos. en neg. antwoord.

Bibliography

- [1] Louis Victor Allis. “A Knowledge-Based Approach of Connect-Four.” In: *J. Int. Comput. Games Assoc.* 11.4 (1988), p. 165.
- [2] Kenneth Appel and Wolfgang Haken. *The four-color problem*. Springer, 1978.
- [3] Kenneth I Appel. “The use of the computer in the proof of the four color theorem”. In: *Proceedings of the American Philosophical Society* 128.1 (1984), pp. 35–39.
- [4] American Psychological Association. *Scale development definition*. Unknown. URL: <https://dictionary.apa.org/scale-development> (visited on 04/29/2023).
- [5] Jeremy Avigad. “Understanding proofs”. In: (2008).
- [6] Jeremy Avigad. “Understanding, formal verification, and the philosophy of mathematics”. In: *Journal of the Indian Council of Philosophical Research* 27 (2010), pp. 161–197.
- [7] Jeremy Avigad. “Varieties of mathematical understanding”. In: *Bulletin of the American Mathematical Society* 59.1 (2022), pp. 99–117.
- [8] Philip Ball. “Maths skills survive linguistic damage”. In: *news nature* (2005).
- [9] Samuel R Bowman. “Eight things to know about large language models”. In: *arXiv preprint arXiv:2304.00612* (2023).
- [10] Antonio R Damasio. “Aphasia”. In: *New England Journal of Medicine* 326.8 (1992), pp. 531–539.
- [11] Alex Davies et al. “Advancing mathematics by guiding human intuition with AI”. In: *Nature* 600.7887 (2021), pp. 70–74.
- [12] Ernest Davis. “Deep Learning and Mathematical Intuition: A Review of (Davies et al. 2021)”. In: *arXiv preprint arXiv:2112.04324* (2021).
- [13] Henk W De Regt and Dennis Dieks. “A contextual approach to scientific understanding”. In: *Synthese* 144 (2005), pp. 137–170.

- [14] Iddo Drori et al. “A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level”. In: *Proceedings of the National Academy of Sciences* 119.32 (2022), e2123433119.
- [15] Sohrab Effati and Morteza Pakdaman. “Optimal control problem via neural networks”. In: *Neural Computing and Applications* 23 (2013), pp. 2093–2100.
- [16] Janet Folina. “Towards a better understanding of mathematical understanding”. In: *Truth, Existence and Explanation: FilMat 2016 Studies in the Philosophy of Mathematics* (2018), pp. 121–146.
- [17] Joachim Frans. “Unificatory understanding and explanatory proofs”. In: *Foundations of Science* 26 (2021), pp. 1105–1127.
- [18] Herman Geuvers. “Proof assistants: History, ideas and future”. In: *Sadhana* 34 (2009), pp. 3–25.
- [19] Mor Geva et al. “Transformer feed-forward layers are key-value memories”. In: *arXiv preprint arXiv:2012.14913* (2020).
- [20] Georges Gonthier et al. “Formal proof—the four-color theorem”. In: *Notices of the AMS* 55.11 (2008), pp. 1382–1393.
- [21] Rahime Çelik Görgüt and Yüksel Dede. “Mathematics teachers’ assessment of mathematical understanding”. In: *Studies in Educational Evaluation* 75 (2022), p. 101190.
- [22] Andrew Granville and Greg Martin. “Prime number races”. In: *The American Mathematical Monthly* 113.1 (2006), pp. 1–33.
- [23] Xu Han et al. “Pre-trained models: Past, present and future”. In: *AI Open* 2 (2021), pp. 225–250.
- [24] Joel Hass, Michael Hutchings, and Roger Schlafly. “The double bubble conjecture”. In: *Electronic Research Announcements of the American Mathematical Society* 1.3 (1995), pp. 98–102.
- [25] Derek Haylock and Anne D Cockburn. *Understanding mathematics for young children: A guide for foundation stage and lower primary teachers*. Sage, 2008.
- [26] Dan Hendrycks et al. “Measuring mathematical problem solving with the math dataset”. In: *arXiv preprint arXiv:2103.03874* (2021).
- [27] Krystal Hu. *ChatGPT sets record for fastest-growing user base*. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> [Accessed: 2023-07-13]. 2023.
- [28] Danqing Huang et al. “Neural math word problem solver with reinforcement learning”. In: *Proceedings of the 27th International Conference on Computational Linguistics*. 2018, pp. 213–223.

- [29] David Hume. “An enquiry concerning human understanding and other writings”. In: (2007).
- [30] Michael Hutchings et al. “Proof of the double bubble conjecture”. In: *Annals of Mathematics* (2002), pp. 459–489.
- [31] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. “Artificial neural networks: A tutorial”. In: *Computer* 29.3 (1996), pp. 31–44.
- [32] Phillip S Jones. “The History of Mathematics as a Teaching Tool.” In: *Nat Counc Teachers Math Yearbook (31st)* 1.17 (1969), p. 69.
- [33] Daniel Jurafsky and James H Martin. “N-gram language models”. In: *Speech and language processing* 23 (2018).
- [34] Mario Krenn et al. “On scientific understanding with artificial intelligence”. In: *Nature Reviews Physics* 4.12 (2022), pp. 761–769.
- [35] Guillaume Lample and François Charton. “Deep learning for symbolic mathematics”. In: *arXiv preprint arXiv:1912.01412* (2019).
- [36] LJ Lander and TR Parkin. “Counterexample to Euler’s conjecture on sums of like powers”. In: *Bull. Amer. Math. Soc* 72.6 (1966), p. 1079.
- [37] Pan Lu et al. “A Survey of Deep Learning for Mathematical Reasoning”. In: *arXiv preprint arXiv:2212.10535* (2022).
- [38] Christopher D Manning. “Human language understanding & reasoning”. In: *Daedalus* 151.2 (2022), pp. 127–138.
- [39] Joseph Moxon. *Mechanick Exercises: Or, The Doctrine of Handy-works. Applied to the Arts of Smithing, Joinery, Carpentry, Turning, Bricklayery. To which is Added, Mechanick Dyalling: Shewing how to Draw a True Sun-dyal on Any Given Plane, However Scituated; Only with the Help of a Straight Ruler and a Pair of Compasses, and Without Any Arithmetical Calculation.* 3959. D. Midwinter and T. Leigh, 1703.
- [40] Seong Joon Oh, Bernt Schiele, and Mario Fritz. “Towards reverse-engineering black-box neural networks”. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (2019), pp. 121–144.
- [41] OpenAI. *GPT FAQ*. Accessed: 2023-06-21. 2023.
- [42] Kim Plofker. “Mathematics in india”. In: *Mathematics in India*. Princeton University Press, 2008.
- [43] JH Poincaré. *Intuition and Logic in Mathematics, Halsted, transl.* 1900.
- [44] Luana Poliseli. “DE REGT, HW Understanding Scientific Understanding. New York: Oxford University Press, 2017.” In: *Principia: an international journal of epistemology* 24.1 (2020), pp. 239–245.
- [45] Neil Robertson et al. “Efficiently four-coloring planar graphs”. In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing.* 1996, pp. 571–575.
- [46] Sebastian Ruder. “An overview of gradient descent optimization algorithms”. In: *arXiv preprint arXiv:1609.04747* (2016).

- [47] Kevin Samoly. “The History of the Abacus”. In: (2012).
- [48] Patrick Schneider. *An Introduction to Proof Assistants*. Accessed: 2023-05-25.
- [49] John R Searle. “Minds, brains, and programs”. In: *Behavioral and brain sciences* 3.3 (1980), pp. 417–424.
- [50] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. “Activation functions in neural networks”. In: *Towards Data Sci* 6.12 (2017), pp. 310–316.
- [51] Tirta Susilo and Bradley Duchaine. “Advances in developmental prosopagnosia research”. In: *Current opinion in neurobiology* 23.3 (2013), pp. 423–429.
- [52] Coq team. *A short introduction to Coq*. URL: <https://coq.inria.fr/a-short-introduction-to-coq>. (accessed: 01.07.2023).
- [53] Robin Thomas. “An update on the four-color theorem”. In: *Notices of the AMS* 45.7 (1998), pp. 848–859.
- [54] College voor Toetsen en Examens. *Correction model exam wiskunde A vwo 2018, 2e tijdvak, 19 juni 2018*. 2023. URL: <https://static.examenblad.nl/9336118/d/ex2018/VW-1024-a-18-2-o.pdf> (visited on 07/13/2023).
- [55] College voor Toetsen en Examens. *wiskunde A vwo 2018, 2e tijdvak, 19 juni 2018*. 2023. URL: <https://www.examenblad.nl/examen/wiskunde-a-vwo-2/2018/vwo> (visited on 07/02/2023).
- [56] Thomas Tymoczko. “The four-color problem and its philosophical significance”. In: *The journal of philosophy* 76.2 (1979), pp. 57–83.
- [57] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [58] Lei Wang et al. “Template-based math word problem solvers with recursive neural networks”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. 2019, pp. 7144–7151.
- [59] Limin Xiang. “A formal proof of the four color theorem”. In: *arXiv preprint arXiv:0905.3713* (2009).
- [60] Wong Khoon Yoong. “Aspects of mathematical understanding”. In: (1987).
- [61] Wayne Xin Zhao et al. “A survey of large language models”. In: *arXiv preprint arXiv:2303.18223* (2023).