

Intuition and Insightfulness in AI-assisted Proofs

MSc Thesis (*Afstudeerscriptie*)

written by

Katia Parshina

(born December 23rd, 1998)

under the supervision of **Dr. Luca Incurvati**, and submitted to the
Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense: **Members of the Thesis Committee:**

August 24th, 2023

Dr. Luca Incurvati

Dr. Maria Aloni

Prof. Dr. Arianna Betti

Dr Thomas Schindler



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

In this work, I analyze how AI was implemented by Davies et al. (2021) in a mathematical proof. The authors claim that they used AI to guide mathematical intuition. First, I analyze the existing accounts of mathematical intuition and mathematical insight and distinguish these two notions. Then, I discuss the way Davies et al. implemented AI into their work and compare it to the classical, non-AI-based computer assistants. Finally, I discuss whether the notions of mathematical intuition and mathematical insight can be used to describe the work AI performs in the discussed proofs. As a main result, I conclude that the discussed AI cannot, as suggested by Davies et al., guide mathematical intuition, and the role it plays can only be described as assisting a mathematician in obtaining mathematical insight.

Contents

Introduction	1
1 Classical computer-assisted proofs	5
1.1 What are computer-assisted proofs?	5
1.2 Approaches to the critique	8
2 Intuition and Insight In Mathematical Practice	12
2.1 Mathematical intuition and mathematical insight	12
2.2 Different approaches to intuition	15
2.3 Perceptual approach: intuition <i>of</i>	19
2.4 Insight and mathematical models	24
2.5 Insightfulness and intellectual virtues	30
2.6 Insight, knowledge, and human body	34
3 AI in Mathematical Proofs	40
3.1 The case of Davies et al.	40
3.2 AI-assisted proofs <i>vs</i> computer-assisted proofs	47
4 The Place of AI: Between Insight and Intuition	53
4.1 Agency of AI	53
4.2 Intuition and AI	59
4.3 Insight and AI	62
Conclusion	69

Introduction

In 2021, Davies et al., a group of researchers consisting of software engineers and mathematicians, published a paper “Advancing Mathematics by Guiding Human Intuition With AI”. In the paper, the authors describe how they obtained two mathematical proofs with the help of AI. The usage of AI in mathematics is not new: it was previously used to solve differential equations (Lample & Charton, 2019), speed up time-consuming calculations (Peifer et al. 2020), and find counterexamples to conjectures (Wagner 2021). However, in all the previous examples, AI functioned as a tool for deriving a proof the framework of which was already established by a human mathematician. Even if AI was used to detect patterns in data that could help a mathematician define the strategy of the proof, the decision of what conjecture to prove was always done exclusively by a mathematician. Davies et al. claim that their implementation of AI into mathematical practice is essentially different from the previous cases, as it demonstrates a novel approach to AI assistance in mathematics: it shows that AI can not only assist in obtaining a proof but also guide mathematical intuition in choosing *what* to prove (Davies et al., 71).

The debates around the use of computer assistants in mathematical proofs emerged in 1977, with the publication of the first proof that was obtained with the help of a computer program (Appel & Haken 1977). With the continued use of computer assistants in mathematics, several arguments were raised against the equal acceptance of computer-based and fully human-performed proofs in mathematical practice¹. One of the arguments, described by J. D. Barrow (1995) and W. P. Thurston (1994), focuses on the fact that proofs obtained with the help of computer assistants are not insightful, which means that they do not demonstrate any new methods, patterns, or connections that could lead to new discoveries in mathematics. Thurston insisted that computer software cannot reproduce mathematical intuition, which is human by nature (Thurston 1994, 5). The implementation of AI presented by Davies et al. can be used as a counter-argument against the described position: it can be used to argue that mathematical intuition and mathematical insight are not essentially human since one can use a computer assistant to guide and discover them. If the authors indeed show that a mathematician’s intuition can be directed and assisted by

¹The arguments against the equal acceptance of computer-based and human-based proofs are described by several authors, for example, T. Tymoczko (1979), J. D. Barrow (1995), J. R. Brown (2010), C. W. H. Lam (1990).

AI, it means that computer software can be used to discover new methods and connections, contrary to the presented arguments. If computer software can be used to discover new methods and connections, we can argue that insightfulness and intuition in mathematics are not necessarily purely human phenomena.

Davies et al.’s case can be distinguished not only from the previous implementations of various computer assistants in mathematics but also from the general implementation of AI in scientific research. AI is used for data analysis and classification in various domains, including but not limited to biology (Webb 2019), chemistry (de Almeida et al. 2019), and physics (Melnikov et al. 2018, Liu et al. 2023). However, the researchers of AI implementation in science openly state that so far, despite providing significant help to human scientists in revealing patterns that would be otherwise hard to discover, AI cannot fulfill the function of an agent in scientific discovery (Melnikov et al. 2018, 1224; Krenn et al. 2022, 761). Davies et al. do not state that in their work, AI took on the role of a scientific researcher. However, the authors state that they used AI to guide the researcher’s exploratory creativity, which could represent the first step of sharing or transferring the agency in scientific discovery from fully human to partly artificial. The article was perceived as a breakthrough in the use of AI in scientific discovery: the headlines describing it were formulated as “DeepMind AI collaborates with humans on two mathematical breakthroughs” (Sparkes 2021) and “AI is discovering patterns in pure mathematics that have never been seen before” (Nelid 2021). The paradigm in which the paper by Davies et al. has been discussed suggests that the authors have programmed AI to independently prove theorems in mathematics (Davis 2021). That is, of course, not the case, as we will show in the next chapters. However, the authors claim that they used AI to guide the intuition of a mathematician, which, as I will show in this thesis, is also a big deal.

Davies et al. do not give a definition for the term ‘mathematical intuition’ they use in their work. Moreover, they do not clarify the exact role mathematical intuition plays in the discovery and proving process in mathematical practice. Hence, their statement about the possibility of guiding mathematical intuition is not yet supported by sufficient grounds: is it really mathematical intuition that they are guiding with AI? The goal of this thesis is to determine whether it is admissible to describe the role of AI in the proofs obtained by Davies et al. as ‘guiding mathematical intuition’. If it is determined that the process can be described as AI guidance of mathematical intuition, we would obtain an argument against some critique of computer assistants: it would mean that they can help us develop new methods and strategies and, what is more, they can guide mathematicians in choosing what they are going to prove, not merely how they are going to prove it. Hence, it would be possible for computer assistants to possess some sort of mathematical creativity. If, however, we determine that the AI Davies et al. use does not guide mathematical intuition but rather provides help close to the function of a classical, non-AI computer assistant, we would not be able to ascribe it any sort of exploratory creativity.

The structure of my work is the following. In the first chapter, I briefly describe the classification of non-AI computer proofs and their critique, focusing

on the critical approach to the ‘insightfulness’ of computer-based proofs. In the second chapter, I describe the existing philosophical accounts of mathematical intuition and mathematical insight². I explain which accounts I am going to use to describe the role of AI in mathematical proofs and why. I also discuss some philosophical implications of these accounts that may allow or restrict their application in the context of AI. In the third chapter, I describe the implementation of AI in mathematics. In section 3.1, I describe how exactly Davies et al. implemented AI into their work and what results they achieved with it. In section 3.2, I compare the usage of classical computer assistants and the AI assistant mathematical practice: I focus on how the implementation of AI by Davies et al. is different from the implementation of non-AI assistants described in Chapter 1. In the last chapter, I approach the role of AI described in Chapter 3 with the notions of mathematical intuition and insight described in Chapter 2. I discuss whether it is possible to explain the work of AI in the case of Davies et al. in terms of intuition and insight and the philosophical implications of the answer to this question.

²The notions of mathematical intuition and mathematical insight are both used in philosophy of mathematics while talking about different kinds of exploratory creativity. I describe the differences between these notions in Chapter 2.

Chapter 1

Classical computer-assisted proofs

1.1 What are computer-assisted proofs?

To understand the novelty of AI-assisted proof introduced by Davies et al., we need to understand the general classification of computer-assisted proofs. That is why, in this section, I will describe the general framework of computer-based proofs as it existed before the introduction of the AI assistant by Davies et al. (e.g., before 2021) In the next section, I am going to explore the existing critique of the computer proofs, including the critique of their insightfulness – the side of criticism which we will try to apply later to the AI-assisted proofs.

There are several types of computer-assisted proofs. What vaguely defines a computer-assisted mathematical proof is that a computer was used in order to get to some (sometimes crucial) step of the proof. There are, however, different methods to use a computer in gaining mathematical results: that is one way to build a taxonomy within the computer use. I use the taxonomy outlined in Parshina (2023). There are computer-assisted proofs that:

1. Use a computer to find a counterexample. If a mathematician finds a counterexample, they can simply use it in their proof, sometimes without even mentioning that the result was obtained with the use of a computer.
2. Use a computer to prove the lack of counterexamples. There are at least two ways in which that can be done:
 - (a) Non-formalized (or not fully formalized) proof. Computer software is engaged at the step where a large set of cases needs to be checked for the existence of counterexamples. A mathematician formulated the initial proof and uses a computer assistant to do that.
 - (b) Fully formalized proof. The mathematician uses a proof assistant (for example, *Coq*, *Metamath*, *Isabelle*) or a similar software tool as

a framework for their proof. One needs to formalize the theorem for a proof assistant, which derives the proof itself, according to the encoded inference rules.

Now, I provide some examples for each case of the described taxonomy in chronological order. Proofs of type 2(a) are the so-called ‘old’¹ examples of computer-assisted proofs: the most classical example for this type is the proof of the four-colour theorem by K. Appel & W. Haken (1977). The theorem is formulated as follows:

Theorem 1 (Four-colour theorem) *Every map (whether it is located on a plane or a sphere) can be colored with no more than four colors in such a way that regions sharing a common boundary are never colored alike.*

The theorem was initially introduced by F. Guthrie in 1852 (MacKenzie 2004, 103) and was only proven in 1977 in the article “The solution of the four-colour map problem” by Appel & Haken. The authors reduced the general case to a triangulated graph (i.e. all its faces are triangles). Then, the authors proved that the minimal vertex degree m of such a graph can be either 3, 4, or 5. Cases $m = 3$ and $m = 4$ are easily proved (Whitney & Tutte, 1992). The problematic case is $m = 5$. The authors decide to prove it based on the notion of *reducible configuration* (first implemented by Birkhoff, 1913). The notion of reducible configuration is defined as follows:

Definition 1 (Reducible configuration) *A graph S is called the reducible configuration if the 4-colorability of any planar graph containing S as a subgraph is deducible from the 4-colorability of any planar graph with fewer vertices.*

From the definition, it follows that the minimal counterexample of Theorem 1 cannot contain a reducible configuration. Appel and Haken used a computer program to produce an unavoidable list U of 1834 configurations (i.e. any planar graph contains at least one of these configurations) and used a computer program to prove that every configuration in U is reducible².

Appel and Haken did most of the proof by themselves, without the use of a computer. The computer was needed for going through a large set of configurations at the final step of the proof and stating their reducibility. Therefore, Appel and Haken’s proof is of type 2(a).

The proof of the four-colour theorem initially presented by Appel and Haken contained several mistakes (Wilson 2003, 225), therefore, it was corrected and partly formalized by Robertson et al. (1996), and fully formalized in 2005 by Gonthier. Gonthier’s and Robertson et al.’s proofs are examples of type 2(b). Both of them are formalized but to different scales. Robertson et al. reduced

¹I call these proofs ‘old’ because they were introduced the earliest, therefore they were the center of the discussion on the validity of computer-assisted proofs in the period of 1970-1990.

²The authors could have proved the unavoidability of U by hand, but writing the proof down would have taken 2 months. Therefore, Appel and Haken decided to use a computer to prove it, as well as for proof of reducibility.

the list U to 633 configurations and proved its reducibility with the use of a computer program. While the authors did not formalize the whole proof (i.e. the first part of the proof was still done by hand, even though its complexity was reduced), they were able to formalize the last part of it - they present the reader with a link to a computer program where they can check the reducibility of the list by themselves.

Gonthier's proof of the four-colour theorem is fully formalized. The author worked with Robertson et al.'s version of the proof and formulated Theorem 1 in the calculus of inductive constructions, CiC (Gonthier 2005, 3-5). After that, he was able to present the theorem as an input to *Coq*, an interactive theorem prover³, and get the fully-formal proof.

It is harder to present clear cases of the proofs of type 1, but a possible example is the proof of the Boolean Pythagorean triples theorem presented by M. Heule, O. Kullmann, and V. Marek (2016). The Boolean Pythagorean triples theorem is formulated as follows:

Theorem 2 (Boolean Pythagorean triples theorem) *Set $N = \{1, 2, \dots\}$ can be divided into two parts such that no part contains Pythagorean triples.*

Definition 2 (Pythagorean triple) *A Pythagorean triple is a triple of integers (a, b, c) s.t. $a^2 + b^2 = c^2$.*

In 2016, the authors proved that it is possible for positive integers to be colored in such a way, but only up to the number 7824. They found the number by using a Boolean SAT solver to go through a set of cases and find the first counterexample. Since the authors used a computer assistant, their proof is partly of type 1. However, it is arguable that the authors could have come up with the counterexample themselves since the counterexample was produced by an exhaustive computer search.

All the presented examples of computer-assisted proofs were done with the help of either an 'old-fashioned' extensive search through a set of cases or an interactive theorem prover. However, these are not the only types of computer-assisted proofs existing at the moment. In 2021, Davies et al. published an article where they explain how they proved two mathematical theorems by using an AI assistant. We are going to discuss the differences between classical computer-assisted proofs and AI-assisted proofs in Chapter 3. Now, we will see the main points of critique of the classical computer-assisted proofs, in order to understand, in the following chapters, whether AI-assisted proofs could be subject to the same criticism.

³An interactive theorem prover is different from the computer programs used by Robertson et al. or by Appel and Haken since it does not simply go over a given set of cases with the given algorithm. An interactive prover relies on logical inference rules and checks the proof of mathematical statements typed into it, or helps to find the proof.

1.2 Approaches to the critique

In this section, we will describe three main points of approach to the criticism of classical computer-based proofs. We will mainly focus on the position discussing the lack of originality and insightfulness in computer-based proofs, as this position can be argued against by finding an example of mathematical proof that, even though it was achieved with the help of computer software, presents novel approaches to the topic, – and AI-assisted proof by Davies et al. could be this example.

There exists a lot of critique aimed at computer-assisted proofs. Several philosophers and mathematicians refuse to give them the same level of trust as to human-based proofs for different reasons. Following the points of criticism underlined in Parshina (2023), we can distinguish three main approaches to the critique:

1. There is a high possibility of a mistake in computer-assisted proofs, and there is a good chance that, if the computer software produced a mistake, it will never be found.

This position is supported by J. R. Brown, C. W. H. Lam, and T. Tymoczko⁴. Brown (2010) talks about the fallibility⁵ of computer proof assistants. In proofs like Appel and Haken’s or Robertson et al., various software pieces are used: the operating system, the programming language, a compiler, special computation software, etc. We may theoretically understand how each of these pieces works, but when we combine them, we rely heavily on practical matters in understanding how these things work altogether. Brown suggests an example of applying Newton’s classical mechanics to extremely massive objects: just like this application was wrong, our hypothesis about the inner structure of the software we use may be wrong. Even if we have some theories, we cannot go and verify them with some sort of experiment, like Newton could not check if his mechanics describe the world ‘correctly’. We can run a computer program several times, but we will never obtain the full reliability of the software we use.

2. Computer-assisted proofs bring empirical truths into mathematical practice.

That objection is strongly connected to the previous one, as the fallibility of computer proofs is the reason why mathematics, for some authors, can become quasi-empirical. The reason why computer-assisted proofs make mathematics (at least partly) empirical is connected to the reliability of the software. If we are unable to trace every step of the reasoning behind the proof, we rely on the computer’s output in a different way than we rely on a highly-skilled mathematician’s results. If the software we use is unsurveyable, there is a possibility of the

⁴Here I should note that despite Tymoczko stating that there is an increased chance of producing a software error, he does support the implementation of computer-assisted proofs into mathematical practice (Tymoczko 1979, 72).

⁵Brown defines fallibility as the possibility of a proof containing a mistake.

existence of mistakes we will never be able to notice. We rely on the computing program, the compiler, and other software or hardware pieces that produce a proof - but there does not exist a way for us to check that they all indeed work as we wanted them to. Therefore, every time we use them to produce a certain output, we need to recognize its (partly) empirical nature - we use the output only because we conditionally believe that our program works as we theorize it to work.

Tymoczko equally accepts human-based and computer-based proofs into mathematics, which, according to him, leads to a big change in our understanding of mathematics (from purely rational to the partly-empirical site). Brown and Lam, however, do not support the change and insist on the problem of the software's unsurveyability.

3. Computer-assisted proofs do not bring any new 'insightful' methods into mathematics.

That is the position supported by J. D. Barrow (1995) and W. P. Thurston (1994). According to Barrow, mathematical practice does not consist of collecting all possible truths (Barrow 1995, 232-233). It is rather centered around demonstrating new methods, patterns, and links that can provide some insights and intuitions about objects, notions, and relations in mathematics. According to Barrow and Thurston, the result of the proof is not as important as the new methods that were demonstrated to the mathematical community. One may argue that Appel and Haken indeed demonstrated a new method in mathematics: after all, they proved a theorem by writing and running a computer program - a method that was never used before them in mathematics. To that, the authors supporting Barrow and Thurston's position could answer that the proof that Appel and Haken demonstrated did not provide us with any new connections between mathematical notions since all the computations were produced inside of a computer program. When a human mathematician produces a new proof, even if no important connections between mathematical objects are described in it, the (skilled enough) reader is still able to trace the line of thought that is intended by the author. In a computer-assisted proof, even if a part of the proof is explained by a human mathematician, the computer-assisted part still provides only the result, not the reasoning behind it. For some, the mathematical result would be valuable by itself. However, several discussions about the disinterest of mathematicians in Appel and Haken's proof show that it did not provide the information the mathematical community was expecting it to provide. Several mathematicians, when they discovered that the four-colour theorem was proved, experienced some sort of disappointment when it was revealed to them that the proof included a supercomputer. Mathematician P. Halmos writes, discussing the value of Appel and Haken's proof: "I want to ask: what did we learn from the proof? What do we know now that we didn't know before? I do not find that an easy question to answer. [...] We may still be far from finding a new "good" proof for the four-color theorem" (Halmos 1990, 18). Barrow provides an excerpt from a mathematician M. Davies: "[...] I expected some brilliant new

insight, a proof that had in its kernel an idea whose beauty would transform my day. But when I received the answer [...], I felt disheartened. My reaction was: “So it just goes to show, it wasn’t a good problem after all” (Barrow 1995, 323). The fact of the confirmation of the conjecture to be true does not concern the named mathematicians. What concerns them is the fact that the potential human-based proof of the theorem could have demonstrated some deep mathematical connections, but it did not.

For Thurston, the problem with computer-assisted proofs is the lack of ‘human-like’ mathematical thinking. Thurston states that one of the main features of mathematical practice is its form: it is concentrated on human language, vision, hearing, deduction, intuition, stimulus-response, etc. We do not only communicate mathematical results in human language (as in, articles, books, lectures), we think in human language when we think about mathematics (Thurston 1994, 169). For example, when we use the equation sign “=”, we think about it as of verb, and we use it as a verb. What is more, Thurston supports the ‘human-language’ approach to mathematical thinking because it leaves room for the phenomena like intuition, association, metaphor, etc. Some mathematicians, according to Thurston, make use of mathematical intuition even without a specific understanding of where this intuition comes from (Thurston 1994, 165). Thurston states that all knowledge transfer in mathematics should be human-based, and we should distinguish between a computer-assisted proof (as relying on the capabilities of a computer) and a human-based proof (as relying on human thought and information intake).

According to the authors like Barrow, Thurston, and Halmos, computer-assisted proofs cannot provide the mathematical community with any new insights that could advance our understanding of mathematics. Computers can be used to go through some important steps of the proof, and we would obtain some new information, but the resulting proofs would not demonstrate any methods or perspectives we could use to expand our toolkit for mathematical practice. And for authors like Barrow and Thurston, the non-computer proofs that demonstrate some insight into a studied entity are the core of mathematical practice, – that is what mathematics is ‘about’. In the following chapters, we are going to explore the proof method presented by Davies et al., including the step where a mathematician interacts with AI to obtain a better understanding of the studied topic. This case could be used as an example showing that proofs obtained with the use of computer software can be insightful – or, at least, that computer assistants do not always make mathematical proofs unhelpful. Moreover, if the authors indeed do what they state they do, which is, guiding mathematical intuition with AI, the example of Davies et al. can demonstrate that it is not only possible to use a computer assistant and obtain an insightful proof, but it is also possible to use a computer assistant specifically to find, or improve, the novel and insightful approaches to the studied topic. That way, we could not only answer the critique of the helpfulness of computer-based proofs but also show that it may be the computer that brings helpfulness to the proof. We will discuss Davies et al.’s case and compare it to the classical use of computer assistants in Chapter 3. Before that, we should clarify all the

concepts we are going to use during the discussion of mathematical insightfulness and intuition of Davies et al.'s AI, and that is exactly what we do in the next chapter.

Chapter 2

Intuition and Insight In Mathematical Practice

2.1 Mathematical intuition and mathematical insight

In this chapter, I am going to discuss the most well-established accounts of mathematical intuition and mathematical insight, so that later we could understand whether these notions are eligible to apply to the role of AI in mathematical proofs. Specifically, I plan to establish whether we can describe the work completed by Davies et al.'s AI as guiding mathematical intuition or guiding mathematical insight. To understand whether AI has anything to do with mathematical intuition and insight, we need to understand these notions first. Even though Davies et al. only mention mathematical intuition, I add the notion of mathematical insight to the discussion because these phenomena are closely related and sometimes even confused, or used interchangeably. Further in this chapter, I will elaborate on the connections and differences between mathematical insight and mathematical intuition. In Chapter 4, I will discuss which of these notions is more suitable to describe the work of AI in Davies et al. (2021).

All the accounts I am going to cover characterize mathematical intuition or insight from within different domains. Very few of the authors I discuss provide a precise definition of the concepts they use. Most of them compare the notions of mathematical intuition and insight to some other notions - for example, to sensory perception. However, the authors list several essential characteristics of the two notions which reflect the positions of different communities on the matter: E. Chudnoff (2018) and C. Parsons (1979) provide the philosophical perspective on the notion of mathematical intuition¹; I. D. MacDonald (1979)

¹In sections 3.2 and 3.3, I explain why I chose to base the philosophical perspective on mathematical intuition on Parsons' and Chudnoff's accounts and did not choose other authors. I will also justify the choice of all the other listed authors.

and H. B. Griffiths (1971) frame the discussion of mathematical insight and present a perspective on the topic from the community of professional mathematicians; J. A. Carter (2017) analyzes the famous cases of mathematicians publicly discussing their ‘moments of insight’ to describe the essential human-related features of this phenomenon; Nathan and Walkington (2017) research the extent to which the process of obtaining mathematical insight is connected to the human body and gestures. The chosen set of authors allows us to approach the topic of mathematical insight and intuition from different paradigms: philosophy of mathematics, mathematical practice, and cognitive and behavioral science. I do not suggest that the chosen accounts and studies are the only correct way of approaching the topic of mathematical intuition and insight. My motivation was to present the most thoroughly-described theories with a strong theoretical and/or empirical framework so that we could have a strong base for the notions we are going to apply to AI.

Parsons and Chudnoff describe the perceptual account: an account that uses human perception as an analogy for the description of the role of mathematical intuition. The authors generalize some of the characteristics of human perception to describe mathematical intuition. Specifically, Parsons talks about the human ability to perceive abstract mathematical objects based on our previous experience with some specific examples of these objects in practice. He compares it to our ability to perceive and identify empirical objects in real life based on all our previous experiences with the empirical world. There are, however, several features that make perception different from mathematical intuition. For example, mathematical intuition about an object is not connected to the possibility of the actual existence of the object (Parsons 1979, 151). The differences and similarities between sensory perception and mathematical intuition can make it either easier or harder for us to transfer the characteristics of classical accounts of mathematical intuition to the discussion of AI in mathematics: if mathematical intuition, like human perception, is necessarily closely associated with a human subject processing the internal experience similar to qualia², it makes it harder to attribute this feature to AI. If, on the other hand, mathematical intuition is an ability more independent, up to some point, from a unique inner experience than perception, it makes it easier to attribute the capacity of mathematical intuition to AI. I discuss why I chose the perceptual approach to mathematical intuition in section 2.2, and the approach itself and its implications for a mathematician’s agency and subjectivity in section 2.3.

MacDonald and Griffiths, whose approaches to mathematical insight I describe in section 3.4, were practicing mathematicians, unlike the other chosen authors. Both of the authors’ articles I discuss were published in the 1970s, which is more than 50 years ago, therefore one might suggest that there is no need to discuss the authors’ approaches since we already have some discussion of the historical perspective on mathematical intuition from Chudnoff and Parsons. However, there are no recent publications on the notions of mathematical intuition and insight that are authored by practicing mathematicians.

²Qualia is an introspectively accessible experience of a person (Chumley & Harkness, 2013).

The discussion on the matter in the mathematical community was active in the 1970-1990s. Since then, there exist two active communities that discuss mathematical intuition, insight, and their origin: philosophy and epistemology of mathematics (Chudnoff 2018, 2-3) and cognitive studies (Dehaene 2009, 234-235; Nathan and Walkington 2017, 2-4). The works written by MacDonald and Griffiths are, of course, a part of the heritage of philosophy of mathematics. However, their work also reflects the attitude of the mathematical community towards the notion of mathematical insight³. That is why their accounts are essential in this research: they allow us to look at this notion from within the mathematical community and, therefore, obtain a first-hand description of the studied phenomena.

There are some accounts that associate intuition and insightfulness in research with certain human virtues: for example, H. Battaly (2008) states the motivation for truth is a core characteristic of a successful scientist. For B. G. Charlton (2009), all scientific discoveries depend on the explicit pursuit of truth by the researcher. Such characteristics may be the reason why the concepts of intuition and insight cannot be applied to the AI system used by Davies et al. since we would be left with the problem of ascribing human virtue to an artificial agent. While there are some studies that allow attributing some level of agency to AI (Legaspi et al. 2019, Andrada et al. 2022), there are no well-established philosophical accounts that allow ascribing human virtues to AI systems in their current state of development (Gill 2020, 293-294). Therefore, it is crucial to understand how human virtues are connected to insight and intuition in mathematics before trying to apply these notions to AI. That is why we are going to discuss the connections between mathematical insight and human virtues in section 2.5, following the approach of Carter (2017).

The discussion of mathematical intuition in the paradigm of cognitive science is also an essential part of our research: Nathan and Walkington (2017), Nathan et al. (2014), and Uttal et al. (2012) are some examples of papers that present case studies of obtaining mathematical insight and intuition in mathematical practice. The authors mostly study cases with children or college students as participants while researching the transfer of mathematical knowledge. That allows us to look at mathematical intuition in a simpler and more comprehensible environment than by analyzing the cases of mathematical insightfulness in successful adult mathematicians as Carter (2017) suggests. Cognitive approaches to mathematical insight and intuition allow us to start the discussion on the characteristics of agents obtaining the insights and using the intuition: in what circumstances is it possible for an agent to experience an insight, what could stimulate it, what cognitive processes are occurring during the obtaining of the

³Griffiths provides several examples of the discussion of mathematical insight by mathematicians. MacDonald critiques some aspects of Griffiths' approach, but generally agrees with him on the general framework of mathematical insights, and analyzes some examples from mathematical practice. Even if the authors do not represent the general attitude of all mathematicians towards mathematical insight, they represent a significant percentage of the mathematical community by grounding their approach in various examples from the community.

insight? If we discuss these notions from the cognitive point of view, we most certainly would be talking about human beings as the agents in whose minds these processes are occurring. Since, in this work, I talk about the possibility of describing the work of AI by the notions of mathematical intuition or insight, the approaches that connect these notions to human cognitive states are one of the possible critical accounts against this enterprise. Hence, in section 2.6, we discuss the possible problems that cognitive science approaches may bring into the discussion of the agency and subjectivity behind mathematical insights.

I should note that while mathematical insight and mathematical intuition are related notions, they cannot be used interchangeably. There are different suggestions on how the notions of insight and intuition can be distinguished, for instance: insights only appear in consciousness, while intuition is an unconscious process (Volz & Von Cramon 2006); intuition is a continuous process and insight is an intermittent process, appearing and establishing within several seconds (Bowers et al., 1990). Zhang et al. (2016) distinguish intuition and insight as two different approaches to knowledge and information processing. The authors state that intuition is mainly involved in processing judgments on questions with ‘yes/no’ responses – in such cases, we are not interested in the particular solution of the problem we are engaged with, but rather just want to know whether there *exists* a solution. In the case of insight, we are interested in the particular solution for the problem we are working with – when we obtain insight while looking for a solution to a mathematical problem, we obtain the information on the solution itself, not just on the fact that it exists/does not exist (Zhang et al. 2016, 2). Also, the kind of obtained knowledge is different in intuition and insight: while we can obtain tacit knowledge⁴ with mathematical intuition (we know whether some fact is true, but we do not know *why*), we can only obtain insight about an explicit knowledge⁵ (we know specifically *why* some fact is true). This distinction between insight and intuition is a result of theoretical and empirical cognitive studies of mathematical practice. There are no authors who discuss the differences between insight and intuition in the domain of philosophy of mathematics. However, as we will see in sections 2.4 and 2.6, the separate discussions of insight and intuition in mathematics generally follow the distinction drawn by Zhang et al. We will come back to this distinction in section 2.6.

2.2 Different approaches to intuition

There exist several approaches to mathematical intuition. They can be distinguished by the description of the method one obtains the intuition and the

⁴While we do not know the specific solution to the problem, we may feel a subjective closeness to the solution, as Reber et al. (2007) describe it. That is what Zhang et al. call ‘tacit knowledge’.

⁵Explicit knowledge is factual knowledge that can be easily communicated. Of course, we cannot always correctly define the knowledge we obtained – there is a harmful possibility of mistaking tacit knowledge for explicit knowledge while obtaining mathematical insight (Knoblich et al. 1999).

conceptual content of the intuition. In this section, we talk about the account of intuition we have chosen to discuss and why.

Chudnoff (2018) has attempted to classify the approaches to mathematical intuition in philosophy of mathematics. He suggests distinguishing three main accounts:

1. Concrete illustration approach
2. Spontaneous impression approach
3. Perceptual approach

The concrete illustrative approach is based on the view suggested by C. F. Klein (described in Ewald 1996, 959). Klein states that when we are to imagine an abstract mathematical object, we do not, in reality, imagine a fully ‘abstract’ object lacking any certain characteristics. When we are asked to imagine a line, we always imagine a line of a certain width and length, therefore endowing our object with some fixed characteristics. We imagine a concrete illustration of an abstract object, not the abstract object directly. Therefore, the object we are intuiting is never abstract.

The spontaneous impression approach is represented by S. Dehaene (2009). According to Dehaene, to have an intuition that p is to have a spontaneous impression that p : an impression that is “fast, automatic, and inaccessible to introspection” (Dehaene 2009, 233). The type of intuition Dehaene describes is based on the high level of familiarity with symbolic representations of numbers: specifically, intuition appears when we are able to understand the connections between the symbols so fast that the whole process becomes automatic and unconscious.

The perceptual approach is represented by several authors, for example, by Parsons (1979) and Chudnoff (2018). According to the perceptual approach, mathematical intuition can be compared to sensory perception in several aspects. Mainly, the comparison is based on the assumption that mathematical intuition allows us to experience the abstract mathematical object like it is present in our immediate environment. Chudnoff calls this process a “non-inferential access” (Chudnoff 2018, 2).

While I agree with Chudnoff on distinguishing the presented accounts of mathematical intuition, I do not agree with the classification he suggests. Chudnoff lists these three accounts as different approaches within philosophy of mathematics and bases his analysis of these accounts purely on the philosophical background – he starts with the discussion of the role of I. Kant, R. Descartes, and K. Gödel in the formation of the notion of intuition. However, while the concrete illustrative and perceptual approaches are described in the paradigm of philosophy of mathematics (and the authors of these approaches are either mathematicians or philosophers), the spontaneous impression approach is suggested by Dehaene as a case study of the concepts of knowledge associated with specialized cerebral subsystems. Dehaene does not define a specific account of mathematical intuition in philosophy of mathematics; what he suggests is to use

the teaching of elementary arithmetic in school as a case study to research some specific neurocognitive functions in children. Therefore, while Klein and Parsons talk about the cases of human intuition in adults who consistently work in a professional mathematical community, Dehaene talks about the development of understanding of arithmetic in children (Dehaene 2009, 235). What is more, Dehaene limits the notion of mathematical intuition to “the sense of number” (Dehaene 2009, 233), hence only discussing mathematical intuition applied in elementary arithmetic. Since Dehaene’s account of mathematical intuition is limited to intuition in children and only studies a specific area of mathematics, I am not going to regard it as a complete account of mathematical intuition, as it is unclear how we can extend it to the use of mathematical intuition in professional mathematical practice. I agree, however, that the concrete illustrative approach and perceptual approach are the two main directions the notion of mathematical intuition can take in the paradigm of philosophy of mathematics.

Chudnoff (2018) critiques the illustrative approach because it is only able to describe *partial* aspects of mathematical intuition, “namely the use of visual imagination as a means to awareness of the abstract subject matter of mathematical propositions” (Chudnoff 2018, 6). We may go even further and argue against the concrete illustrative approach as it does not provide a basis for sustainable mathematical knowledge. If abstract mathematical objects are only accessible to our intuition in the form of illustrations of specific cases, how are we able to incorporate our knowledge of mathematical entities into mathematical practice? Even if we have an intuition about a specific object with strictly defined characteristics, we need to understand how we can generalize the features of this object to work further in the chosen area of mathematics. To find new patterns in the area of mathematics one is working with, or even to connect two initially distinct areas, a mathematician should often come to a certain level of abstraction. Otherwise, one would only be able to conclude statements within the scope of specific examples, which would not allow the achievement of any significant novel results in mathematics. Some mathematicians argue that a certain level of decontextualization should always be allowed in mathematics: it allows mathematicians to introduce new objects, present complex processes, and, generally, learn new information (Ferrari 2003, Mitchelmore & White 1995). For example, when a mathematician works in abstract group theory, they do not refer to the properties of any specific groups (Ferrari 2003, 1226), and if they did, such an approach would not lead to any important results that can be applied to the abstract group theory directly. According to the illustrative approach, mathematical intuition is responsible for producing a specific illustrative example, based on which a person could obtain a better understanding of the subject matter they are studying. The approach does not explain mathematical intuition as an experience of an abstract object (relation, structure, etc.), even though the agent of the intuition somehow obtains the representation of the object in general, not only the specific example they came up with. Hence, the illustrative approach does not include abstraction and generalization as methods that can help one obtain knowledge, awareness, or some sort of experience in mathematics. One may argue that the proponents of the

concrete illustration approach simply do not describe abstraction as a method of obtaining intuition, but they include it into the general mathematical practice. However, Klein does not explain how one can obtain knowledge or at least some awareness about the area you study from merely a specific example from this topic. To understand a mathematical entity, one needs to make some inferences, conclusions, and abstractions based on previous experience: awareness of the topic, let alone knowledge of it, cannot be produced based solely on one specific illustration (Tieszen 1989). If the knowledge is not obtained by abstraction, one should suggest another mental process as a base. Klein does not. Therefore, the concrete illustrative approach does not successfully describe the main processes behind mathematical intuition and, consequently, cannot be used to talk about all the nuances of this phenomenon⁶.

The perceptual approach to mathematical intuition is defended by several authors (Parsons 1979, Maddy 1980, Chudnoff 2018). The comparison of the two mental processes is based on the ability to access an object directly - either to approach a material object through the sensory perception or to approach an abstract object through mathematical intuition. The perceptual approach does not fall under the critique aimed at the concrete illustrative approach: the proponents of the account define the objects of mathematical intuition as abstract mathematical entities which may lack specific characteristics. Generally, the perceptual approach to mathematical intuition has stronger support than the concrete illustrative or spontaneous impression approaches: Parsons and Chudnoff provide several arguments in its favor. According to Chudnoff, mathematical intuition should be compared to sensory perception because both are examples of non-inferential knowledge (Chudnoff 2017, 5): in both cases, we obtain knowledge by direct access to the object available to us (but in different ways). Chudnoff also points out that the perceptual approach is advantageous because it covers not a specific case of mathematical intuition but its general use – unlike Klein’s account, which is limited to specific cases of illustrative intuition, and Dehaene’s approach, which only covers the area of mathematical education in children. Parsons’ arguments in favor of the perceptual approach consist of the observations and examples that are aimed to illustrate the manifestations of mathematical intuition: the author describes the experiences of mathematical intuition that we may possibly have, and suggests that the best way to describe this experience is to compare it to sensory perception. We will further discuss these observations in the next section. Since, in this research, we are discussing the ability to use the notion of mathematical intuition to describe an AI-based system, it is of the great matter for us to understand what the discussed account of mathematical intuition implies about the human agency. In the next section, we are going to analyze the main characteristics of the perceptual approach and its implications that concern the human agency of a mathematician.

⁶Moreover, there are not many authors who argue for the concrete illustrative approach. Other than Klein’s position accessible in Ewald (1996), I was not able to find authors defending this account.

2.3 Perceptual approach: intuition *of*

In this section, we are going to discuss the perceptual approach to mathematical intuition. The description of the approach presented here is mainly based on Parsons' presentation of the theory, hence, I will mainly refer to Parsons. I will also use Chudnoff (2017) to comment on the comparison of sensory perception to mathematical intuition. After the description of the approach, I am going to discuss its implications for the agency of a mathematician.

According to Parsons (1979), some of the first authors to comment on the notion of mathematical intuition were Kant, Husserl, Descartes, and Gödel. These authors did not provide a detailed account of mathematical intuition, but they either commented on mathematical intuition as a part of general human intuition or briefly touched on the concept of mathematical intuition in their works. For Kant and Gödel, mathematical intuition was not separated from general human intuition. In “Critique of Pure Reason”, Kant states that we, as humans, possess sensibility - a capacity to acquire representations. By the means of sensibility, we can have intuitions about objects. For Kant, the notion of intuition is strictly defined as the intuition of *something*, which means that intuition is always directed to an object – more specifically, to the representation of an object (Jankowiak 2014, 3). Husserl used the notion of ‘categorical intuition’ as analogous to perception, also directed to a specific object and not a proposition (Lohmar 2002, 125). Gödel specifically defined the notion of mathematical intuition as the perception of the objects of set theory (Gödel 1947, 520). Descartes considered the notion of intuition to be connected to perception as well. He, however, defined intuition as a clear and distinct perception of propositions, not objects (Slezak 2010, 401-402). Parsons reviews the listed accounts to show that intuition often was compared, in its nature, to perception. Therefore, the novelty of Parsons' approach consists in defining this comparison exclusively for mathematical intuition⁷.

Parsons develops Gödel's account of intuition in set theory into the general account of mathematical intuition. First, Parsons distinguishes two possible focuses of mathematical intuition: intuition *that* and intuition *of*. Intuition *that* is what Parsons calls “the traditional understanding of intuition” (Parsons 1979, 146): it denotes an intuition of certain statements, or propositions, to be true. Intuition *that* is not based on habit, convention, or previous practice and experience, – it is considered to support propositions based on the self-evidence of the statements. Descartes' account of intuition is an example of intuition *that*. For him, the objects of intuition are self-evident elementary truths, obtained not by experience but by systematic reasoning (Beck 1954, 50). Intuition *of*, on the other side, is not an intuition about a specific proposition being true; it is rather a clear representation of a mathematical object. Parsons uses Gödel's description of mathematical intuition to define intuition *of*. For Gödel, mathematical intuition is a direct relation between a person and a mathematical

⁷While Kant indeed writes about mathematical intuition, he mainly focuses on intuition in geometry and does not develop it into the general account of mathematical intuition (Parsons 1979, 145-147).

object. He writes: “[...] we do have something like a perception of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true” (Benacerraf et al. 1983, 271). For Gödel, intuition *of* [the objects of set theory] leads to intuition *that* [certain statements in set theory are true]. If we have an intuition that certain axioms are true, we must have an intuition of the objects that are used to define those axioms. Parsons argues that the modality in which mathematical objects are given to us is very similar to the manner in which we perceive physical objects. He also suggests that the perceptual account of intuition *of* is the most suited to describe intuition in mathematics, because, as Parsons notices, a considerable amount of arithmetic is given to us in a more direct way than propositional reasoning (Parsons 1979, 152). Hence, Parsons develops the perceptual account of intuition *of*.

Parsons bases his explanation of how intuition *of* works on the following example. Suppose, we introduce a stroke “|” as a basic symbol of the syntax of a given language. The well-formed expressions of the language are arbitrary strings |, ||, |||, ... and the presented sequence of strings is isomorphic to natural numbers. Two strings are of the same type if their strokes can be placed in a one-to-one correspondence (Parsons 1979, 153). Now, according to Parsons, we have an understanding of types which makes them *less abstract* to a listener and is close to our experience of words and sentences in natural language. Now, we have an intuition about what a type is, even if we do not fully understand all the nuances of type theory. We understand the concept of a type and can apply it. We have some background knowledge to start to think about this concept and to intuitively understand what is a type and what is not: mathematical intuition, according to Parsons, is always based on our previous experience in mathematics. Our basic mathematical intuition is token-oriented, or object-oriented – it can only be *about* an abstract object. Intuition *of* can give rise to intuition *that*: the former would postulate the existence of certain objects, and the latter would postulate the existence of some relations between these objects. For instance, if we have an intuition of || and the subsequent intuition of |||, we have an intuition *that* ||| is the successor of ||.

Parsons notes that by acquiring the understanding that it is possible to extend every string of strokes, we do not acquire an understanding that it is *physically possible* to do so. Neither mathematical possibility is connected to the capabilities of a human mind and body. Here, mathematical intuition can be distinguished from sensory perception. As Parsons formulates it, “What is *intuited* depends on the concept brought to the situation by the subject” (Parsons 1979, 162), even if the concepts we use have developed spontaneously. According to Parsons’ Kantian view, we cannot have an intuition of an object falling under a concept we do not grasp yet. While with ordinary perception, the perceivable object is not restricted by any concepts the intuiting subject should understand. For instance, I can perceive heat without realizing that a specific movement of molecules is involved in this experience.

Chudnoff (2018) is another author who argues for the perceptual approach to mathematical intuition. Chudnoff defines sensory perception as a “way of gaining information from one’s immediate environment” (Chudnoff 2018, 7).

According to Chudnoff, mathematical intuition has the same phenomenological connotation as sensory perception, which means that these two processes provide a similar internal experience to a subject. Both mathematical intuition and sensory perception provide a direct experience of an object. When we directly perceive a physical object, for example, visually, we are not provided with any kind of additional support for its existence: we are not guessing, not receiving a testimony, not using clairvoyance, etc. We simply perceive the object. When we experience mathematical intuition, according to Chudnoff, we also perceive the abstract object directly, without additional proof or inference about the reasons for its existence. Chudnoff calls this experience the ‘presentational feel of a phenomenon’ because we are directly presented with the object.

Chudnoff’s theory of mathematical intuition is closely related to Parsons’ theory. Both accounts connect mathematical intuition to sensory perception by describing both processes as direct experiences of objects, physical or mathematical. We have seen that, for Parsons, the main difference between the two notions is that mathematical intuition has nothing to do with the possibility of the existence of experienced objects in the real world, unlike sensory perception. Parsons does not explain this difference in more detail: he does not clarify the ontological status of the objects of mathematical intuition any further, as well as he does not explain how the possibility of the existence of physical objects is connected to human perception. Chudnoff partly closes that gap by explaining what enables us to find the abstract objects we experience with mathematical intuition reliable in mathematical practice: the previous experience of a mathematician. The more previous contact with a chosen area of mathematics a person had, the more reliable their mathematical intuition is. We are able to use our mathematical intuition to think about abstract objects because of our previous experiences with still abstract, but more detailed and specified mathematical objects.

I must add that the concept of mathematical intuition introduced by Parsons and Chudnoff may not correspond to the notion of mathematical intuition used by mathematicians in their day-to-day life. As I have stated earlier, I was not able to find a thoroughly defined account of mathematical intuition authored by a professional mathematician. MacDonald and Griffiths, the mathematicians whose work we are going to discuss in the next section, have only built an account of mathematical insight, while sometimes briefly mentioning the intuition. Some mathematicians, like Gödel, only describe some particular cases of mathematical intuition, like the intuition of the axioms of set theory, and do not present an account that would help us understand this concept in full (Gödel 1947, 518). Some other mathematicians choose to not define the concept of intuition specifically for mathematics, and use the general notion of intuition: for example, D. Hilbert approached the notion of intuition applied to finitary mathematics through the Kantian conception of pure intuition (Kitcher 1976, 110; Parsons 1979, 147). Hence, we cannot explore a detailed and established account of mathematical intuition formulated by a mathematician – not a philosopher

of mathematics, – as it does not exist⁸. What mathematicians and teachers of mathematics usually call ‘mathematical intuition’ does not correspond to the philosophical accounts of mathematical intuition we discussed in the present and previous sections. For example, authors who research mathematical education often use the notion of intuition to describe the process of problem-solving, therefore connecting the notion of intuition to the process of gaining mathematical knowledge, not to experiencing an abstract mathematical entity (Arwanto et al. 2018, Baiduri et al. 2022). The authors of self-educational literature who claim that the reader can improve their mathematical skills also describe problem-solving and similar analytical skills as mathematical intuition⁹. S. Feferman examines several examples of the use of the words ‘mathematical intuition’ in mathematical practice and concludes that, on the everyday level, the notion is used to describe the sudden realization of the solution to a studied problem: “[...] *Erlebnis* of a flash of insight or illumination on the road to the solution of a problem” (Feferman 2000, 1). As we can see, the notion of mathematical intuition in the mathematicians’ day-to-day life denotes the skills or experience that allow a mathematician to naturally find the solution to a given problem. This notion does not correspond to the perceptual account of mathematical intuition described earlier in the section: the latter describes the experience of an abstract mathematical entity that may not lead to a specific solution or proof, while the former, as described, is an ability or phenomenon that allows an agent to find the said solutions, proofs, or understand a particular mathematical structure. As we will show in the next chapter, the so-called day-to-day notion of mathematical intuition is closer to the notion of mathematical insight, as it is discussed in philosophy of mathematics. Now, we come back to the discussion of the perceptual account of intuition.

Parsons states that the mathematical possibility of the abstract objects we contemplate using mathematical intuition is not connected to the capabilities of our mind and body (Parsons 1979, 158). What Parsons means by that is that it does not matter whether we can or cannot construct the intuited entity with our mind – the possibility of the existence of such an entity is based solely on theoretical mathematical definitions. It does not matter if the mathematician fully understands how to construct a string and what structure the resulting string will have. They just have an intuition about a possible further string. The mathematician’s intuition is centered on abstract objects but does not define the possibility of the existence of these objects in the actual world: the

⁸I made this assumption based on my research of the topic. I extensively searched through online libraries and archives but was not able to find an account of mathematical intuition formulated by a professional mathematician, e.g. a person whose job is directly related to the research of pure mathematics. This task may include more than online libraries search since the potential author’s occupation should be recognized. This task, I believe, is too big of a project to make it a part of this thesis. However, I hope to investigate it further in the future.

⁹Example: Azad K. (2013) *Math Better Explained: Learn to Unlock Your Math Intuition*. Amazon Press. For other examples, I suggest to google ‘mathematical intuition’ and read numerous online discussions from users who aim to improve their mathematical intuition. Often, if not always, the discussions consist of tips on how to improve one’s problem-solving skills, as well as in the book mentioned earlier.

object a mathematician experiences may not match all the nuances of the actual object in mathematics. However, Parsons writes that an object of mathematical intuition is formed based on what is brought to the attention by the subject. Parsons' theory connects mathematical intuition to the previous experience of the intuiting mathematician. Therefore, Parsons' account strongly connects mathematical intuition to the previous experience of the agent. It follows from this assumption that the subject is eligible to have an intuition about an object based on the experience they had before: for example, whether it is possible to add one more stroke to a string without breaking the rules of syntax. Parsons' theory of mathematical intuition connects fully theoretical abstract objects with the subjective experience of the intuiting agent. However, how strong is this connection?

According to Parsons, we are able to have intuitions of mathematical objects because, among other things, we had previous experience in the related field of mathematical practice. By 'previous experience' Parsons understands the history of interacting with specific examples and problems in mathematics that are connected to the abstract object we are intuiting *of*. Interaction, in that case, is a direct work with examples - describing their characteristics, substituting different variables in formulas, modeling a specific mathematical structure, etc. But what if I did not have any experience with specific objects, but was given detailed instructions on how to decide if there exists a certain object? Consider the following example: I never had any experience with strings and strokes, but was given an instruction "If there is a string, you can construct another string by adding it to itself"¹⁰. I still do not have any experience with specific examples of strings and strokes, but if I was asked "Is there a string that follows the string |?" I would answer 'yes' because I know I can add a string to itself. As in the cases of mathematical intuition presented by Parsons, I would not have a specific understanding of the structure of the object I suggest there is. However, I do not base it on previous experience with specific mathematical cases. One may argue that in that case, it is not mathematical intuition, but a simple case of following the rule. To decide whether the presented example would be a case of mathematical intuition, we would need to analyze a specific definition of mathematical intuition given by Parsons. However, Parsons does not give a specific definition of mathematical intuition but rather lists some characteristics of the phenomenon: first, it is an intuition *of* an abstract object. In the presented case, it is - we are given a rule that helps us to identify if there is an object with a certain characteristic, and, when asked if there is such an object, we are able to answer. Second, we are able to think about the abstract object in what Parsons calls the 'non-factual way' - we do not have an intuition that this object is connected to some other structure in a particular way, we are only able to have an intuition about whether such an object exists in mathematics. In the presented example, I can say that I have an intuition about such an object existing - the rule told me so, hence, I intuit, this is right. The third

¹⁰Even though strings and strokes are not mathematical objects, I use them as an example based on Parsons' explanation of intuition.

characteristic of mathematical intuition presented by Parsons is connected to the object of intuition - mathematicians are able to perceive it, in a restricted sense. In this case, I believe, my example would fail, - if I only had a rule stating how to decide whether a specific abstract object exists, I would not be able to experience the token directly – only when presented with an example. We can see that an essential characteristic of Parsons’ account of mathematical intuition is its connection to the experience of a mathematical object which builds on our previous interactions with similar cases in mathematical practice. The experience of a mathematical object implies the existence of an agent with certain mental states that allow them to access the intuited entity. These mental states combine the previous experience of a mathematician into an intuition about a mathematical entity. To conclude, it is not possible to imitate mathematical intuition by merely following the pre-defined rules: in such cases, the agent would not be able to experience the object based on previous interactions with similar entities, and the intuition, presumably, would not appear. This is a characteristic of the perceptual account of intuition we will come back to in section 4.2, to show how it affects the possibility of (partially) reproducing mathematical intuition with AI or any other kind of computer software.

Parsons and Chudnoff connect mathematical intuition to specific mental states of the agent. Parsons suggests that intuition is not an ‘act’ of mind, but rather the state of it, in which an abstract object is placed ‘before our mind’ (Parsons 1979, 150). So, mathematical intuition is a state of the immediate presence of the object, and it is directly connected to a specific state of mind, which is, experiencing the said object. Chudnoff relates mathematical intuition to demonstrative thoughts and defines demonstrative thoughts as mental activity. According to Chudnoff, mathematical intuition is a process where the subject becomes aware of the object through demonstrative thoughts (Chudnoff 2014, 27). Hence, Chudnoff also connects mathematical intuition to the mental state of awareness of the object. Both authors representing the perceptual account of mathematical intuition describe it through mental states: the state of direct experience of the object for Parsons and the state of awareness about the object for Chudnoff. Essentially, these two descriptions indicate the same state of mind – the ‘perception’ of an abstract mathematical object, as if it was presented to them in their immediate surroundings.

2.4 Insight and mathematical models

Now, we discuss the notion of mathematical insight and its connection to mathematical intuition. One way to distinguish these two notions in philosophy of mathematics was suggested by MacDonald (1978), who uses the notion of a mathematical model initially presented in Griffiths (1971) to describe the core feature of mathematical insight. In this section, we consider several examples of mathematical models Griffiths provides and try to describe the main characteristics of a model, according to the author. Then, we compare it to the account of mathematical insight presented by MacDonald, which is also based

on mathematical models, but differently defined. We conclude by discussing the role of a model in mathematical insight and its differences from the notion of mathematical intuition as described by Parsons.

Griffiths argues for the connection of mathematical insight to mathematical models. The author does not give a precise definition of a model but gives several examples of how models can be used in mathematical reasoning, producing mathematical insight. Here are some of the Griffiths' examples of mathematical insight:

- (1) "If a student takes a course in transfinite arithmetic, and cannot even begin to see how the theory might help him to explain why $3 + 5 = 8$, then his course has given him no real insight".
- (3) "[...] A Ph.D. student was well versed in the spectral theorem for normal operators and Hilbert Space, but could not simplify it to diagonalize a Hermitian form. That student lacked insight".
- (5) "A course on homology groups which concludes by merely stating that "cohomology groups are just the same except that you reverse all arrows" is unlikely to yield insight; a course on groups, which includes topological proofs of certain theorems, might well yield more insight than if a purely algebraic approach were given".
- (12) "It is intuitively obvious that the two equations $f(x, y) = 0 = g(x, y)$, of the second degree, have no more than four common roots".
- (16) "A continuous function on a real interval is differentiable at each point, with at most a countable number of exceptional points".

All the examples present mathematical insight in different forms. Griffiths writes that all the examples he presents have a common feature – they are all connected to the realization that a studied theory has a model in an accessible part of mathematics, for example, geometry. Under the "accessible part of mathematics," Griffiths understands the areas that allow the possibility of representation of a theory through generalizations or examples. In example (1), the student should have obtained an insight about elementary arithmetic because of the generalized knowledge they obtained during the course on transfinite geometry. This generalized knowledge can be considered as a model for the understanding of elementary examples in mathematics, like ' $3 + 5 = 8$ '. In the case of (3), the student was not able to come up with examples to check the general theory, which identifies their inability to understand the structure of the theory. If one is not able to derive some representative examples from the general theory, one does not have an insight about the theory. In case (5), the purely algebraic approach to the course does not give any visual approaches to the content of the course, therefore not providing a possibility to obtain an insightful understanding of the topic. When, however, we are given several visual examples (models) and we have seen the theorems' topological proofs, we might be able to capture the structural content of the theory. All the examples

so far have the common feature Griffiths suggests: a connection to some visual representation in mathematics, either it is a specific illustrated example or an abstract model of the mathematical theory. Griffiths even mentions that computer programs that check all the steps of formal proofs are not able to show insight since they do not have the ability to process illustrative representations of the mathematical models (Griffiths 1971, 156). The author connects insight into a mathematical theory with finding an abstract model in physics, geometry, or some other accessible and/or practical part of mathematics. MacDonald (1978) calls this statement ‘Griffiths’ suggestion’.

Griffiths does not clearly distinguish mathematical insight and mathematical intuition. In his work, he starts to talk about mathematical insight and states that he will provide examples of mathematical insight. However, when discussing example (12), he uses the phrase “intuitively obvious”, again, without a strict definition. Griffiths recognizes statement (12) to be correct, but only on the condition that “the model is known to be a good one” (Griffiths 1971, 156). Griffiths does not give a definition for a ‘good’ model. MacDonald suggests his understanding of what it means for a model to be good. Suppose, MacDonald writes, one becomes familiar with basic arithmetic and learns some foundational properties of integers. The learned list of properties later becomes the list of axioms for the said person, containing commutative law, distributive law, etc. Suppose, the said person uses the set of axioms to their knowledge. However, how can they be sure that that list contains all the existing properties of the integers? MacDonald states that our confidence in the use of the axioms comes from the lack of evidence that the axiom list is wrong. We trust it as we do not encounter cases that would contradict the laws. However, we would never have a *proof* that the integers we are working with (that we are familiar with, use as examples, etc.) are defined by a strict list of axioms. MacDonald suggests defining integers in the formal theory according to the axioms but using ‘familiar’ integers as models. Therefore, we would have a formal definition of an integer that satisfies the list of axioms we have, and a less formal concept of an integer that we can use to describe our intuitive understanding of the object. That ‘less formal’ concept of an integer is what Griffiths calls a good mathematical model - it represents the key features we need to understand the mathematical theory we are working with. A model is good when it is *representative*. According to Griffiths, a mathematical statement is intuitively obvious when there exists a good model representing it. However, as was already stated before, Griffiths does not give a detailed account of mathematical intuition and its connection to mathematical insight.

Griffiths uses the characteristic of “intuitively obvious” to describe mathematical statements that appear to be true, or mental pictures that somehow represent a good mathematical model. MacDonald (1978) tries to give a strict definition to the notion of mathematical insight, and therefore clarifies the relations between insightfulness and mathematical models. The author defines insight as a *trivially true suggestion*, making the notion more primitive compared to what Griffiths tried to describe (MacDonald 1978, 417). The author’s definition of insight is connected to visual input - we obtain it by understanding

a theory represented by some model (either in physics, geometry, or some visually accessible part of mathematics). MacDonald also notes that some models in Griffiths' examples are mathematical, and some are not. Mathematical models are less problematic to define since they can be regarded as a representative selection of examples. Cases (1) and (3) are mathematical, and (4) is not a model, according to MacDonald's definition, since it does not contain a visual example. (16) contains a model that can either be regarded as mathematical or physical. Physical models are less definable for MacDonald: they cannot serve as representative examples. Therefore, while operating with physical models, one needs to choose some specific properties one is interested in and reject the rest, therefore working not with a strict example but with an abstraction.

For MacDonald, mathematical insight has a strict connection to models, given that we define insight as a trivially true statement. Mathematical models that play a role of a representative selection of examples can be used to obtain mathematical insight. However, some models in physics and geometry cannot be defined as representative examples, and to use them, we need to work with an abstract model, leaving some essential characteristics of a formal theory behind. Therefore, according to MacDonald, models can serve as a base for mathematical insights only when they are *mathematical* models that are constructed in geometry, topology, or other part of mathematics that allows the construction of visual examples.

As I already noted, despite a detailed discussion of specific examples of mathematical insight, Griffiths does not formulate a complete account of it. The author gives several examples of what he considers to be mathematical insights, but he does not strictly distinguish it from intuition. Hence, while we are able to describe the role of models in mathematical insights, we cannot accept or reject the role of models, as Griffiths defines them, in mathematical intuition. MacDonald applies the notion of a model to only mathematical insight, restricting Griffiths' account. In this section, we are only going to discuss the notion of mathematical insight and its connection to mathematical models. Therefore, while comparing mathematical insight and mathematical intuition, I will focus my attention on MacDonald's account and some details of Griffiths' account that do not contradict it.

Now, as we have described both the notion of mathematical intuition and the notion of mathematical insight, we can compare the two concepts as they are discussed in the field of philosophy of mathematics. We have seen that Parsons' account of mathematical intuition includes the existence of a subject who has an experience of an abstract mathematical object. The subject does not necessarily have an understanding of all the object's characteristics, they do not necessarily understand how to prove that this object can be defined in mathematics, they may not fully understand all the nuances of the definition of the object: they merely have an intuition that there is such an object. Mathematical insight, on the other hand, as described by MacDonald and Griffiths, includes a significant level of understanding of the internal structure of an object – the representation of its constitution and properties provided to us by a mathematical model. Mathematical intuition, as it is described by Parsons

and Chudnoff, is an experience, close to perceptual, of a mathematical entity, and mathematical insight, as it is described by MacDonald and Griffiths, is an understanding of why and how this entity exists in mathematics: how it is constructed, what conjectures can be used to prove a statement about it, and how the object is connected to other mathematical entities. This understanding may not imply the grasp of all the nuances, especially if it is based on an abstract model, but it allows the mathematician to catch the core of the statement or object they are studying. To have a mathematical intuition, we do not need a model that would represent the structure of the theory we are studying: our attention is concentrated on a specific mathematical entity. In the case of mathematical insight, we need to understand the connection of a specific object to the general theoretical framework we are working with.

We can see that the experience of mathematical insight is closely connected to the knowledge of specific information. When one obtains mathematical insight, they obtain something close to the solution to a problem they have been studying. I say ‘something close to a solution’ because mathematical insights may come in different forms, as we have seen from the examples Griffiths lists. For example, as described in case (5), I can obtain insight about what cohomology groups are based on the information on homology groups I studied in a related course. In that case, I would not obtain a solution plan for some specific problem but rather understand how one piece of theory follows from another. I can also obtain an insight that a certain equation has only a restricted number of roots, as described in the example (12). In that case, I would obtain knowledge that is qualified as something closer to the solution plan than the previous example. We can see that, unlike mathematical intuition, mathematical insight allows one to obtain some sort of explicit, structural knowledge related to the problem one is studying - it may be the connection between the two theories, solution of an equation, proof of a conjecture, etc. Mathematical intuition, on the other hand, does not help one to achieve explicit knowledge directly: it allows an agent to assess what direction is worth pursuing, but the solution strategy needs to be chosen unrelatedly. One may object that mathematical intuition *that* described by Parsons or intuition about the axioms of set theory described by Gödel provides explicit knowledge: we get to know that some statements are true. However, when one intuits a mathematical proposition, they do not get the knowledge of how this proposition can be verified. For Gödel, the axioms that one can intuit do not come with a strategy to justify them - they “force themselves upon us”. Even if we omit axioms as something not requiring formal proof to justify it, Parsons’ intuition *that* also does not come with a proof strategy for the intuited proposition. The intuited proposition is not a formally justified knowledge, as it is composed merely based on intuitions about abstract objects one previously had. It is possible to approach the result of intuition *that* as some sort of knowledge - after all, we obtain a piece of information about mathematical objects, - but it would be knowledge that is not yet supported by a formal proof. Intuiting a proposition does not imply intuiting the way it can be proved. That is how mathematical insight is different from mathematical intuition: the former comes with the knowledge of how to formally verify

a proposition, and the latter does not. Here and further, I will refer to the verifiable knowledge obtained with insight as specific, or explicit knowledge.

One may notice that the discussion of mathematical insight in philosophy of mathematics is somewhat resembling the discussion of mathematical intuition in ordinary life that we have introduced in the previous section. I would agree, to some extent. The so-called ‘day-to-day’ notion of mathematical intuition is a skill allowing one to easily and quickly find solutions to the required problems, hence, the notion is connected to a specific kind of explicit knowledge - the knowledge of a solution to a problem. Mathematical insight, as we have just established, may lead to several kinds of explicit knowledge - proof strategy, connections of two or more theories, the formulation of a conjecture. We may consider the ordinary concept of mathematical intuition to be a specific case of mathematical insight¹¹. However, as the day-to-day mathematical intuition is not formed into any specific account and the description of this concept varies from author to author (Feferman 2000, 2-3), we cannot place this notion into a specific place in our theoretical framework. We established that the ‘ordinary’ concept of intuition is closer to the notion of mathematical insight, as it is described by MacDonald and Griffiths, than to the notion of mathematical intuition, as it is described by Parsons and Chudnoff. That is all we can conclude so far about the so-called ordinary understanding of mathematical intuition.

We have seen that the definition of mathematical intuition, as given by Parsons, is strongly connected to a mathematician as a human agent capable of having intentional mental states. MacDonald and Griffiths, however, base their description of mathematical insight on the ability to construct a representative model of a theory. Now, we should characterize the connection between human agency and the ability to connect a theory to a model. Under the term ‘model’ we should understand MacDonald’s definition of a mathematical model, which is a representative selection of examples. We follow MacDonald’s reasoning and do not consider physical models representative examples as they are only helpful in studying specific features of a theory, not the general structure behind it. Therefore, we will only talk about mathematical models as representative examples.

A representative example in mathematics is an example of a mathematical structure (object, pattern, etc.) that is helpful to understand the general structure of the studied theory. There exist several types of representations in mathematics (Mainali 2021, 1-2), but MacDonald and Griffiths talk specifically about *visual representation* when they talk about models. A representative model is not always a visible drawing or a diagram: for instance, in example (1) by Griffiths, the student has a representative model of transfinite arithmetic *in their mind*¹² that helps them understand the connection between the theory

¹¹The are several ways practicing mathematicians, educators, and authors choose to describe the notion of day-to-day intuition. It may be described as a skill, experience, feeling, intellectual capacity, etc. In this thesis, we do not research the ordinary notion of intuition, hence, we are not able to establish which form of this notion is the most reliable.

¹²In this case, I use the term ‘mind’ very vaguely with an intention to say that the student obtained an understanding of the general structure of transfinite algebra based on the examples

they studied and a specific example. However, representative models can as well exist as a selection of visually accessible examples, for example, as graphic illustrations for a course in algebra. If a model, i.e. a representative structure of a theory, can exist as an illustration or a study case for a specific area of mathematics, then a model can exist independently of a human mathematician¹³. Hence, the existence of a model, as it is defined by MacDonald, is not connected directly to the subjective experience of a human mathematician.

However, even while the model itself is not connected to human agency, the production of a mathematical insight is always described by MacDonald and Griffiths as done *exclusively* by a human mathematician. Griffiths even directly comments on the exclusive human ability to produce insight: “A computer cannot show insight in this sense when it checks the steps of a formal proof for correctness [...]” (Griffiths 1971, 156). To obtain mathematical insight is to extract a representative model from the studied area of mathematics and to be able to use this model to explain how structures and objects are connected to this area of mathematics. While a representative model can be obtained with the use of a computer program (Hege & Polthier, 1997), the explanation of how this model is connected to specific examples in related areas of mathematics is problematic to obtain with a computer, as it involves such mental processes as generalization and abstraction. In Chapter 4, we will discuss whether AI systems used in mathematical proofs can engage with mathematical models.

2.5 Insightfulness and intellectual virtues

In this section, we discuss the notion of insightfulness and its connection to intellectual virtues that are normally attributed to human beings. First, we are going to discuss the account of mathematical insightfulness that connects this notion to intellectual virtues. Then, we are going to discuss how strong this account is and what are its philosophical implications. We will mostly focus on the account of mathematical insightfulness presented by J. A. Carter. While there are other authors who try to define insightfulness through intellectual virtues (which we will mention here), Carter’s approach is focused on the domain of mathematics specifically. Hence, it is the most applicable to our research topic.

Carter (2017) analyzes the connection of mathematical insight to the notion of intellectual virtue. According to Carter, there exist two approaches to insight. First, in the general sense, an act of insight is an episode in which one grasps X (object, property, or relation) and its relatedness to something F (object, property, or relation). Carter calls this act “grasping under general mode” (Carter 2017, 540). Second, insight can also be defined as a result of the

they study.

¹³Here, I will not go into the debates about ontological commitments in mathematics as this topic is too big to discuss additionally. I talk about the independent existence of a model merely in the sense that the illustration is *drawn* in a book and it does not only exist as an object of the human mind.

cognitive act of grasping the relatedness of the objects. However, the notion Carter mainly prefers to discuss in the paper is insightfulness: not a process or the result of insight, but a characteristic of an individual that is expressed in the ability to achieve mathematical insights.

According to Carter, insightfulness is not simply a property of having many insights. The author points out four relational modes that constitute an individual's ability to produce insights: depth, non-triviality, originality, and frequency. Depth describes the ability to connect an abstract entity X with an abstract entity F unexpectedly, for example, to connect two theories in physics in a surprising way. Carter illustrates this characteristic with James Maxwell's discovery of electromagnetic force (meaning that before the discovery electricity and magnetism were regarded as two separate forces). The non-triviality mode describes the property of being meaningful and somehow spontaneous, revealing the connections that are not so easily noticed by non-professionals. The originality mode simply states that if a lot of individuals have the same insight as you do, your insightfulness is less valuable. Frequency is defined in a formula 'the more insights you have, the better'.

There are several approaches that define insightfulness as an intellectual virtue. For instance, virtue responsibilism proponents (Montmarquet 1993, Battaly 2008) consider insightfulness an intellectual virtue because they believe insightful individuals to have a motivation for truth. Carter argues that motivation for truth is not yet enough to call insightfulness an intellectual virtue: an individual might have a certain motivation that they believe to be truthful, but still not achieve the needed level of reliability in issuing true beliefs. What is more, Carter gives an example: suppose, there are two agents, and both of them have absolutely zero motivation for truth. While the first one fails in obtaining any results or achievements in mathematics, the second one succeeds in it: they produce several deep mathematical results, prove some theorems, etc. According to Carter, this situation is not impossible: nothing really stops a person not motivated by truth, but, say, motivated by the strict use of rules, from obtaining mathematical insights (Carter 2017, 544). Therefore, Carter concludes, aiming for truth is not a basis for mathematical insight. Another option is virtue reliabilism (Sosa 1991, 2011; Greco 1999, 2012), according to which insightfulness can be measured by the ability to produce more truthful insights and less false ones. Carter rejects this approach as merely statistical - there are several individuals, he states, who are certainly considered to be insightful people, but who still produce a lot of false insights (Carter sets Elon Musk as an example).

Instead, Carter suggests his own definition of insightfulness. The author bases it on the connection between the structure of insights and our epistemic goals. There are two main episodes in the structure of an insight that Carter highlights: the first he calls *eureka moments* - a sudden moment of realizing the connection between certain theories/properties/etc; and a stage that precedes an eureka moment, *incubation* - a break of attention from the task which stimulates the solution process. It is the basic structure of insight - first, an incubation stage where we take our attention from the task we are working on; then, an

eureka moment - a sudden realization of the solution structure for the problem we are working with. In this structure, however, it is unclear how we distinguish correct and incorrect insights. For Carter, correct insights are related to the factual understanding of a problem, which is a process of gaining knowledge that is related to the real, non-subjective world. Incorrect insights, on the other hand, are not related to the factual world, even though for the individuals obtaining them, they can seem to be true. However, if a mathematical insight is a process (or event) obtained by an individual agent, how we can distinguish subjective and non-subjective truths within it? Carter tries to answer this question by including intellectual virtue in the definition of a correct insight.

Carter solves the problem of distinguishing correct and incorrect insights by pointing out the connection between the cultivation stage of insight (eureka and incubation moments) and the management stage of insight (verification, epistemic conclusions). The success of the management stage is based on the insightfulness of the mathematician. For Carter, the insightfulness of a mathematician is directly connected to their mathematical abilities and it can be measured by a given mathematician being more successful than the mean. When a mathematician is better at skills of preparation and incubation (which are the features that usually can be historically traced for every given mathematician) which preceded his breakthrough, they obtain an indirect control over their insights. Hence, even though eureka moments are not directly controlled by an individual (who can receive them spontaneously), they can be attributed to the level of their previous performance. The more successful the previous mathematical performance of an individual is, the more reliable their insights are. Carter also directly connects the success of the previous performance to the skills of a mathematician: if, in the past, they could successfully process their insights into mathematical results by moving from emotional eureka episodes to correct epistemic judgments, they have proven themselves to be able to distinguish between epistemically valuable and invaluable insights. That ability – to distinguish valuable insights from false ones, – is an intellectual virtue that an insightful mathematician always has. “[...] The virtuously insightful individual is appropriately cautious in moving from emotional experience to endorsement,” - writes Carter (2017, 552). Therefore, for Carter, insightfulness is an intellectual virtue that is measured by the individual’s ability to perform at different stages of achieving and processing an insight, and that ability is defined by the skillset of a mathematician.

We have seen that Carter, while connecting the ability to produce mathematical insight to human virtues, does not directly connect it to an individual motivation for truth. The author connects it to the ability of a mathematician to process eureka moments into epistemically reliable judgments. For Carter, this ability is defined by the skillset of a mathematician, and the skillset can be measured by the mathematician’s previous achievements. One can notice that Carter’s definition seems to be circular: a mathematician is insightful if they have previously shown that they are insightful. It is unclear how one can enter this definition without a significant, already existing and recognized background in mathematical practice. If the definition of insightfulness worked how Carter

described it, which is, purely based on the already demonstrated skills, there would be no possibility for new mathematicians to enter the field and achieve insightfulness in mathematics. One may suggest that we should understand Carter's position on mathematical skills as measured in degrees: for example, a mathematician can build up their experience and start by producing a little number of mathematical insights (for example, by obtaining mathematical results that are not that essential in their area of expertise) and gradually develop their work so that at some point they can be called insightful in the mathematical community. This interpretation is more feasible than a strict separation of insightful and non-insightful mathematicians once and forever, but it can be met with obvious critique: how do we understand that one's experience is enough to be called insightful? For example, suppose, a mathematician obtained 3 different results in mathematics, all of them - not very significant, but they have some value. We can say that these results contain some non-essential insights. Now, suppose, the same mathematician proves a significant theorem by obtaining a presumably essential mathematical insight. We try to understand whether the insight the mathematician obtained can be trusted and we look at their previous work. There, we see three non-essential proofs. Is this background enough for us to call the mathematician insightful? If it is, how have we measured it? What if they had two results? One? If such a background is not enough to call the mathematician insightful, will it ever be? To raise the level of insightfulness of a mathematician, there should be a significant mathematical insight in their experience. For there to be a significant mathematical insight in their experience, we need to raise their accumulated level of insightfulness. We enter a vicious circle by trying to interpret Carter's theory into real-life mathematical practice.

Despite an unclear explanation of the connection between mathematical insightfulness and previously documented mathematical ability, Carter's account establishes a strong connection between the previous mathematical experience of an individual and their ability to produce mathematical insights. We have already seen this connection in MacDonald's and Griffiths' definitions of mathematical insight: the authors state that insight is always connected to a model, which can be defined as a representative example of the theory one is working with. For MacDonald, a good selection of representative examples is a basis for insights in mathematics. The selection of examples can serve as a basis for mathematical insight only when it is processed by an individual – a mathematician – in their previous experience. During the processing of the examples, an individual gains an understanding of the main structural features of the studied theory, which allows them to obtain an insight – the understanding of the basic structure of the theory and how it is related to the studied examples. We have discussed that the models, as MacDonald and Griffiths define them, can be, in theory, generated by some other source than a human agent. However, when it comes to the production of insight, a human mathematician is an essential variable. A human agent uses selective models as content for generalization and abstraction to obtain insight about a specific topic in mathematics. To use the models as content for mental processing, mathematician, obviously, needs to

encounter the models in their practice. Therefore, for Griffiths and MacDonald, as well as for Carter, a mathematician's experience also plays a significant role in mathematical insight.

There is, however, a difference between how MacDonald and Griffiths discuss a mathematician's experience and how Carter discusses it. For Carter, the previous experience of a mathematician is based on their previous performance in mathematical practice. That 'performance' includes not only the ability to understand various mathematical structures, patterns, definitions and use them in practice: for Carter, a level of mathematician's experience is mainly based on the mathematician's previous achievements: if you have not yet obtained any significant results in mathematics, you have an essentially low possibility of obtaining mathematical insights. For MacDonald and Carter, however, it does not matter whether you are an independent mathematician with a publication history or you are a student who just tries to understand a new topic in class. If you have a good understanding of a model of the theory you study, or you are able to comprehend a selection of representative examples which illustrate the main features of the studied theory, you are eligible to obtain a mathematical insight. What is more, if you understand a mathematical model of a theory, you are already halfway in to obtaining an insight – what is left is to understand the connections this model has with other models in mathematics. Therefore, according to Carter, mathematical insight is based on previous recognition in the mathematical community, while for MacDonald and Griffiths, insight can be obtained by any potential individual who understands the studied subject. Carter's account of mathematical insight is much less stable than MacDonald's and Griffiths'. What is more, Carter discusses only a specific subset of mathematical insight – mathematical insight in successful, widely-recognized mathematicians. Hence, Carter's account does not provide strong support for the definition of mathematical insight based on the mathematician's previous recognition, even though Carter strengthens the connection between mathematicians' experience and their ability to produce insights.

Nevertheless, Carter's account is valuable because of its particular focus on the property of insightfulness. MacDonald and Griffiths have built a stable account of mathematical insight, but they barely discussed the property of being insightful. We have established that for a mathematician to be insightful is to be cautious in transferring their previous experience into new insights, and be successful in doing so. Hence, mathematical insightfulness is based on one's intellectual virtue, specifically, cautiousness, and their previous experience in mathematics. In Chapter 4, we will try to apply the notion of insightfulness to AI in mathematical proofs.

2.6 Insight, knowledge, and human body

In this section, I want to talk about the connections between mathematical insight and the human body of a mathematician. This discussion unfolds in a paradigm different from the pure philosophy of mathematics: the field we will

be discussing is closer to cognitive and behavioral science. The discussion will be concentrated on the experimental studies conducted by M. J. Nathan and C. Walkington, as their research is specifically concentrated on mathematical reasoning and mathematical insight. While there are several studies in cognitive science that discuss mathematical cognition, Nathan and Walkington developed a concluded theory of mathematical cognition that thoroughly describes mathematical insight and the role of the human body in it. Since, in this paper, it is crucial for us to understand the nature of the connection between the human body and mathematical insight, we will discuss the authors' theory in this section.

Nathan and Walkington (2017) introduce the theory of grounded and embodied mathematical cognition (GEMC) which explains how the human body is connected to mathematical reasoning. The authors suggest that human-based mathematical communication shapes cognition states that allow us to obtain mathematical insight. Nathan and Walkington state that, combined with human speech, gesture use in mathematics provides a basis for the emergence of mathematical insight.

Non-specified for mathematics, grounded cognition is a theory stating that human intellectual behavior can be analyzed based on different external factors: states of the body, actions, and gestures (Barsalou, 2008, p. 619). When we concentrate our attention on the role of the body in human intellectual behavior, we are talking about the theory of embodied cognition (Nathan & Walkington 2017, 3). With GEMC, the embodied cognition theory applied to mathematical practice, Nathan and Walkington research the effects of direct actions used in mathematical reasoning. The authors specifically use the word "gestures" to distinguish directed actions from spontaneous body movements (Nathan & Walkington 2017, 3-4). There exist several experiments that show the connection between the improvement of the mathematical ability in students and the use of dynamic 'depictive gestures' in teaching and learning - actions in which people represent an object or an idea with their own body (Donovan et al., 2014). In some other studies, the researchers improved primary school children's mathematical understanding by manipulating the students' hands to achieve a certain representation of a mathematical object (Abrahamson and Trninic 2015; Abrahamson 2015). In the study by C. Petrick and T. Martin (2012), high-school students were enabled to engage in dynamic geometric constructions which improved the learning process significantly compared to the group that was only able to visually interact with geometrical objects. All these and multiple other studies¹⁴ show that mathematical reasoning and mathematical learning are significantly influenced by body gestures. Nathan and Walkington suggest that the ability to produce mathematical insight is strongly connected to human body movements.

Within GEMC, Nathan and Walkington distinguish mathematical intuition and mathematical insight. Based on the studies by Zander, Öllinger, & Volz

¹⁴For a more detailed description of the experiments supporting the GEMC theory, see Nathan & Walkington (2017), 2-6.

(2016), the authors define mathematical insight as a partial understanding of the underlying ideas of a given mathematical system. Intuition is different from insight in the following: the former is based on unconscious knowledge and does not provide a specific description of the decision-making process, while the latter is based on both conscious and unconscious knowledge and provides a traceable solution to a problem. Under ‘traceable solution’ the authors understand the ability of an individual to formally prove the result they obtained with a mathematical insight. For example, if I obtained an insight that two mathematical structures are inherently connected, I should be able to provide proof of why is it so. The knowledge I obtain with insight should be explicit. In the case of mathematical intuition, one usually cannot provide the specifics of the decision-making process which led them to the intuition. The distinction between mathematical intuition and mathematical insight in cognitive science is closely related to the distinction between intuition and insight in philosophy of mathematics. In both areas, intuition is not connected to the full and nuanced understanding of a certain mathematical structure, while insight is a deep and thorough understanding of a given mathematical structure, or of the relations between two mathematical structures. Mathematical insight can serve as a basis for mathematical proof as it provides an explicit understanding of a given theory (area, problem. etc.), while mathematical intuition can only be used to research a certain topic further, in search of reinforcement for a formal proof. However, the authors who presented a completed account of mathematical intuition in philosophy of mathematics do not describe it as ‘unconscious’ knowledge. Neither Chudnoff nor Parsons uses the division of conscious and unconscious to describe the knowledge obtained with the use of mathematical intuition. MacDonald, Griffiths, and Carter state that insight can only be obtained by an individual in the case of a deep structural understanding of the area of mathematics they research: MacDonald and Griffiths connect this understanding with the existence of a model describing a studied theory. Nathan and Walkington go further than MacDonald and Griffiths in defining the knowledge an agent obtains with mathematical insight – they say that it should not only be precise but also traceable. That characteristic allows us to make sure that the knowledge one obtained with an insight is explicit: it can be formally organized and included into the theory we are working with. The traceability of the decision-making process specifies the nature of insight-produced knowledge, making it more reliable. That characteristic described by the authors researching GEMC allows us to make a strict distinction between the result of mathematical intuition and mathematical insight.

Based on several experiments (Nathan & Walkington, 2017, 64-65), Nathan and Walkington provide a description of the processes taking place during mathematical reasoning, according to GEMC. Students who study a mathematical topic use gestures to illustrate the used mathematical notions and patterns - the gestural illustration can either be initiated by students themselves or by their mentors who guide the students’ thought process. There are various kinds of gestures the students can use: a dynamic illustration of the sides of a triangle

to prove the triangle conjecture¹⁵, hand movements that illustrate proportional reasoning, dynamic gestures that illustrate a change in an angle's degree, etc. The actions of students directed by mentors stimulate the activation of the motor system of the human body which induces spatiotemporal processing (Nathan & Walkington, 2017, 6). The spatiotemporal processing alone can lead to an insight, but it would be an insight without proof, or formal explanation, as it is not supported by the appropriate language system. However, if the mentoring system includes the appropriate language as well (familiarity with the basic concepts and relations in the discussed mathematical theory), a language system is activated, which leads to speech processing, which results in allowing a student to support their insight with a formal explanation. Combined, speech and spatiotemporal processing allow dynamic gestures and transformational speech to work as the mediators of mathematical thinking, which can lead to a mathematical insight supported with proof. Nathan and Walkington claim that the combined coordination of human speech and actions is essential to perform simulated actions on abstract mathematical objects: human language and actions serve as mediators for mathematical thinking, allowing students to grasp new mathematical knowledge, including mathematical insights. To support the authors' theory, there exist several studies that show that mentors' interventions directing students' arm and eye movements have led to better performance of the students in mathematics (Nathan et al. 2014; Novack et al. 2014; Thomas & Lleras, 2007).

According to Nathan and Walkington's account, human gestures play a significant role in producing mathematical insights. However, we should note two important points here. First, the authors do not state that bodily gestures are the only possible resource that can induce mathematical insight. They do not reject any other stimuli that can help in insight-producing. Therefore, their account does not contradict MacDonald's theory of mathematical insight. It is possible for a mathematician to understand a mathematical model, use gestures to describe how it works, and, subsequently, obtain a mathematical insight about a connection of this model to some other mathematical structure. For MacDonald and Griffiths, understanding of a mathematical model is an essential and indispensable step in obtaining insight. For Nathan, Walkington, and other researchers working with GEMC, guided gestures are only *one* of all the ways of obtaining insights - it may be one of the most successful ones, but it is not an irreplaceable step of the process. The second important point we should keep in mind while talking about the GEMC account is that it mainly describes insights obtained by students, not by practicing mathematicians. The authors concentrate their research not on mathematical practice in general, but on mathematical education, mainly high-school level. It is expected that in mathematical education, the teacher's explanatory techniques play an important role in the level of mathematical understanding their students have. However, in more general mathematical practice, explanatory techniques do not

¹⁵Triangle conjecture is the following theorem: the sum of the lengths of any two sides of a triangle must be greater than the length of the remaining side.

play such a major role: it is more important how a mathematician thinks and describes the theory they are working with, as it is given that they already understand the model and structure of the chosen mathematical area. I do not reject the assumption that gestural illustration can provide a strong basis for obtaining insights into non-educational mathematical practice. However, to make a strong connection between gestural illustration and the production of mathematical insights in general mathematical practice, we would need to support it with research that does not exclusively include educational practice as case studies. Therefore, while Nathan and Walkington’s research strongly connects human gestural explanation to mathematical insights, this connection is limited to the area of mathematical education.

However, Nathan and Walkington, as well as other authors researching the notion of mathematical insight from the perspective of cognitive science, add some important points to our understanding of this notion. MacDonald and Griffiths connect the notion of insight to some specific knowledge that the mathematician obtains from it, but they do not define the type of knowledge specifically. Griffiths provides many examples of mathematical insight and the information an agent obtains with it can be very different: a connection between two theories, a solution for a mathematical problem, a precise formulation or proof of a conjecture, etc. Nathan, Walkington, and Zander give more specificity to the type of information one obtains with insight: according to their definition, when an agent has a mathematical insight, they obtain a solution for a problem they studied. And, if it is indeed mathematical insight, the solution is always traceable: a mathematician should be able to formally verify the information they claim to obtain. Even though Nathan and Walkington use this characteristic of insight to describe the educational practice specifically, I believe that this description can be expanded to the type of explicit knowledge that is generally obtained in mathematical insight. It gives us a criterion according to which we can distinguish mathematical insight from mathematical intuition: if an agent can provide a verifiable traceable solution to their insight, we accept it as such.

We have discussed various approaches to mathematical intuition and mathematical insight in this chapter. We have established the perceptual approach to mathematical intuition represented by Parsons and Chudnoff as the most well-established in philosophy of mathematics: it provides a clear explanation of the phenomenon and is applicable to all areas of mathematical practice. We discussed three approaches to mathematical insight: in connection to mathematical models (MacDonald, Griffiths), in connection to intellectual virtues (Carter), and in connection to human gestures (Nathan and Walkington). We established that MacDonald and Griffiths’ approach to mathematical insight allows for models, or representative examples, to be produced by a computer. That is a promising path for us to describe the role of AI in Davies et al.’s proof. We found some flaws in Carter’s discussion of mathematical insight, particularly, the opacity of the connection between the ability to produce insights and a particularly recognizable background of a mathematician. However, Carter’s account is valuable to us since he, unlike MacDonald and Griffiths, particularly describes the property of mathematical insightfulness. The authors studying

GEMC (Nathan & Walkington, Zander et al.), even though their domain of research is restricted, focus on the explicitness of knowledge one obtains with mathematical insight – that is a feature that allows us clearly distinguish intuition and insight in mathematics, and we will discuss it (as well as all the above) further in Chapter 4. Since MacDonald and Griffiths’s account of mathematical insight is the most well-established, we will base our discussion of mathematical insight in AI on their approach. We will discuss the insightfulness of AI based on Carter’s account, and we will use GEMC to formulate the main differences between the notion of insight and intuition. Now, to describe the role of AI in mathematical proofs, we need to discuss how exactly Davies et al. used it, and how is it different from other computer assistants. That is what we do in the next chapter.

Chapter 3

AI in Mathematical Proofs

3.1 The case of Davies et al.

We have seen various approaches to mathematical intuition and insight. To understand how they can be applied to describe the role of AI, we, obviously, need to know how AI was implemented by Davies et al. into their proofs. That is exactly what we are going to discuss in this section. First, we describe the framework the authors developed: we describe the role of AI in this framework and how it is combined with the role of a human mathematician. Then, we discuss the authors' evaluation of the role AI plays in their proof and try to define the notion of AI they use in their work. Based on the description of the role AI plays in Davies et al.'s proof that we provide in this chapter, we discuss the possibility of application of the notions of intuition and insightfulness to AI-based systems in chapter 4.

In 2021, Davies et al. published a paper “Advancing mathematics by guiding human intuition with AI”, where they proposed a new sort of computer-based mathematical proof: AI-assisted proof. Part of the paper's authors are mathematicians: András Juhász, Marc Lackenby, and Geordie Williamson, and the other part of authors consist of researchers from DeepMind, artificial intelligence research laboratory¹. Together, the authors developed a method of using machine learning (ML) algorithms to provide assistance to a human mathematician in theorem proving. They used ML to discover patterns and relations between mathematical objects that can provide a mathematician with observations that can be potentially helpful in proving a related theorem. Davies et al. call this technique “guiding human intuition with AI” (Davies et al. 2021, 70-71) and suggest that their model for human-AI collaboration can achieve significant results by engaging with the respective strengths of machine learning and human mathematicians.

¹The researchers from DeepMind are: Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, Demis Hassabis, Pushmeet Kohli.

In the introduction to their article, the authors talk about the importance of a mathematician’s intuition. They cite a mathematician T. Tao: “It is only with a combination of both rigorous formalism and good intuition that one can tackle complex mathematical problems” (Davies et al. 2021, 70). The authors do not specify if they use a pre-existing account of mathematical intuition or define mathematical intuition by themselves. The authors only specify some features of mathematical intuition when they describe the aspects of theorem-proving that can be guided by ML algorithms, for example, the ability to intuit the existence of relations between two mathematical objects (Davies et al. 2021, 71). Davies et al. state that the AI-based framework they present helps to guide a mathematician’s intuition in two different aspects: first, by closely investigating the patterns hypothetically suggested by a mathematician as an input; second, by helping them to understand these patterns. Based on the activity that the authors discuss as the object of guidance for AI, we may suggest that they consider the process of researching different mathematical patterns and objects with the goal of discovering new connections between them as one of the tasks that mathematical intuition is (or should be) occupied with. However, the exact nature of a mathematician’s intuition is not defined by the authors.

Since Davies et al. define the presumed connection between two or more mathematical entities to be the object of mathematical intuition, we can assume that the approach the authors adopt is related to the perceptual account of intuition described by Parsons. In section 3.3, we established that Parsons distinguishes two types of mathematical intuition: the first-stage intuition about the possibility of the existence of mathematical objects (intuition *of*), and the second-stage intuition about statements concerning the relations between these objects (intuition *that*). The second stage can only be built upon the first one: we can have intuition *that* only about the relations between objects about which we have intuition *of*. Despite the absence of a specific definition, Davies et al. describe an approach to intuition in mathematics that is close to the intuition *that* in Parsons’ perceptual account. For Davies et al., the goal of an agent making use of mathematical intuition is to reveal new relations between pre-defined mathematical objects. In this case, these relations, or patterns, are the objects of mathematical intuition: an intuition *that* this variable is connected to this variable by some means that are normally not strictly defined or even known in advance by a mathematician. This description fits the definition of intuition *that* presented by Parsons, even if Davies et al. did not choose any specific account of mathematical intuition. The authors emphasize the concept of intuition in their article to state that in their project, a mathematician’s intuition is guided by AI. Now, we look more closely at the process of obtaining mathematical proof with the use of AI, and see what role intuition plays in it, according to the authors.

The authors suggest the following framework describing the process of AI guidance of a mathematician’s intuition (**Fig.1**).

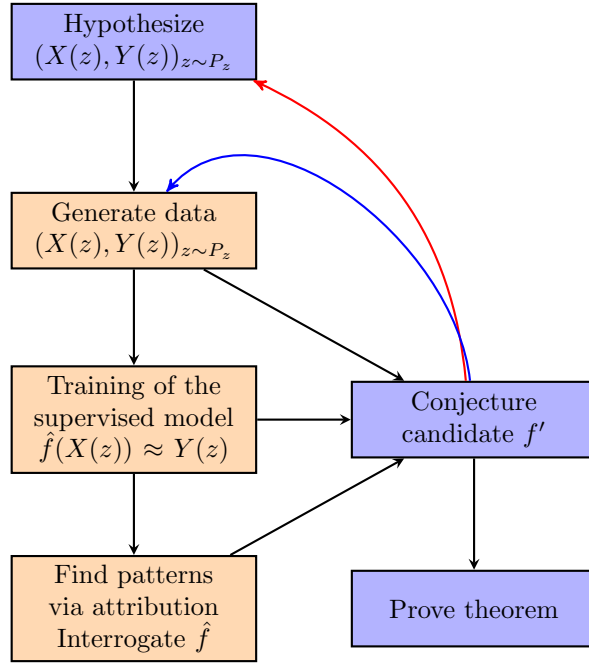


Figure 3.1: **Flowchart of the Davies et al.'s framework.**

Purple boxes represent the steps that a mathematician undertakes by themselves, yellow boxes represent the steps at which the mathematician uses a computer program or machine learning algorithms. Black arrows represent the main core of the algorithm, the blue arrow - altering the distribution sample, the red arrow - reformulating the hypothesis.

As one can see from the chart, there are six stages to the process of guiding mathematical intuition with AI to obtain a proof. The lilac boxes are steps that a human mathematician performs by themselves, without the help of AI; the orange boxes are the stages that AI system performs under a mathematician's supervision. The process presented in the framework is interactive and iterative, which means that it cannot be reduced to a fixed series of steps: it is designed to be repeated several times in different directions until the mathematician can approach the final part, formulating and proving a theorem.

The framework in Figure 1 describes the following process. First, a mathematician formulates a hypothesis: suppose, it is “there exists some relationship between $X(z)$ and $Y(z)$ ”, where $X(z)$ and $Y(z)$ are some mathematical objects. Then, the authors generate a dataset P_z of pairs $X(z)$ and $Y(z)$ and use supervised ML algorithms to train a function $\hat{f}$ that uses $X(z)$ as an input and predicts $Y(z)$ based on it. By using ML in such a way, we can analyze a dataset that is too big to be approached by a human mathematician: if $\hat{f}$ turns out to be more accurate than we would expect it to be, it may be an indication of an existence of a nonlinear function connecting $X(z)$ and $Y(z)$. “More accurate

than we expect it to be” means, for the authors, that if the accuracy of $\hat{f}$ is above the chance compared to further examples from P_z (on which the function was not trained), then a relation between $X(z)$ and $Y(z)$ is considered to be found. If $\hat{f}$ shows patterns that may indicate such connections, we go to the next step: applying attribution techniques that may help to explain the patterns. Generally, what Davies et al. call “attribution techniques” is any kind of system that can help with the explanation of the patterns found in the previous stage². The specific attribution technique that Davies et al. use in their work is gradient saliency - a technique that calculates the derivative of the $\hat{f}$ ’s outputs (with respect to the inputs). The application of the gradient saliency technique implies, in most cases, repetition: it is an iterative process that we may need to repeat several times before we settle on a conjecture f' that characterizes the relations between $X(z)$ and $Y(z)$. However, even after we settle on a specific conjecture, we can continue to work within the cyclic framework: as you can see in Figure 1, we can continue using ML techniques to generate a dataset based on our potential conjecture’s inputs and outputs and continue testing the conjecture. In case we are dissatisfied with the conjecture f' we can continue re-formulating the initial hypothesis about the relations between $X(z)$ and $Y(z)$, generating new datasets and training a supervised model on them. When, at some point, a successful³ conjecture would be found, a mathematician can proceed to formulate and conclude the formal proof of the conjecture.

Davies et al. call this framework “a test bed for intuition” (Davies et al. 2021, 71): the designed system is supposed to help mathematicians quickly check whether the intuition they have about a connection between two mathematical entities is worth to be developing, and understand the path of this development. Hence, the help from ML algorithms comes at an early stage of the supposed proof, or, arguably, precedes the stage of the proof. In the next paragraph, we are going to see how exactly ML is used by Davies et al. to do what they call ‘guiding human intuition’.

When Davies et al. say that they used AI to guide a mathematician’s intuition, they mean two main steps of the framework that they constructed: the step where they train a supervised model to find relations between the introduced mathematical entities and the step where we use the integrated gradient techniques to explain the found relations. The supervised learning stage follows the proposition of a hypothesis. For the purposes of an easier demonstration, the authors suppose that there are no known functions that map $X(z)$ to $Y(z)$ (Davies et al. 2021, 75). The introduced framework, as the authors write, may still be valuable if there are some fixed relations already existing between $X(z)$ to $Y(z)$, but in their work, they only formulate how to work with two mathematical entities with yet not established relations. So, we generate a dataset of

²Attribution techniques applied in ML help to identify the patterns and relations we need to explore further to reveal the reason behind the found connections between $X(z)$ and $Y(z)$.

³In this case, a “successful” candidate for a conjecture is a conjecture that correctly describes the patterns relating $X(z)$ and $Y(z)$ that were discovered on a previous step of the algorithm.

pairs $X(z)$, $Y(z)$, sampling Z from a distribution P_z ⁴. To see if our hypothesis holds, we use supervised machine learning to train it to receive $X(z)$ as an input and produce a mapped $Y(z)$ as an output. Davies et al. use the approach presented by Bronstein et al. (2021) called “geometric deep learning”. Geometric deep learning is an approach to geometry based on using the domain of deep learning to study the connections between various geometries under symmetry transformations⁵. Bronstein et al. explore neural networks that are used for the analysis of unstructured sets, graphs, and manifolds, and show that these networks can be used to unify unstructured mathematical entities and present them with respect to the symmetry in their domain. Davies et al. chose the geometric deep learning as a basis for their supervised learning stage because the approach allows them to work with several different geometrical formats of $X(z)$ and $Y(z)$ (with different approaches to invariances and symmetries), which meant that the format of input data $X(z)$ could be easily processed by the network. As was already said, if the accuracy of $\hat{f}$ prevails over chance on the samples of P_z that were not used as an input, we consider there to be a relation between $X(z)$ and $Y(z)$.

The next step of the framework is to explain the found patterns, i.e., to find what exact structures are responsible for the predictions made by $\hat{f}$. Davies et al. used the integrated gradients method introduced by Sundararajan et al. (2017). Sundararajan et al. suggest using the gradient method in machine learning because, as the authors show, this technique satisfies the two most important characteristics of attribution techniques: sensitivity and implementation invariance. Experts who work with machine learning algorithms are required to constantly inspect the model coefficients and their products to obtain more correct predictions. Model coefficients are the variables that are responsible for the direction in which predictor and response variables are connected. The integrated gradients method, as Sundararajan et al. present it, is an analog to model coefficients for a deep network: it allows us to control the relations between the input and the output of the deep networks. Sensitivity is a characteristic according to which, if the function used in machine learning does not depend on a certain variable, the attribution to that variable is always 0. In that way, the integrated gradient technique focuses only on the features relevant to the input and desired output. Implementation invariance is a characteristic that makes sure that if two networks are functionally equivalent (i.e. their outputs are equal for all their inputs), the attributions of these networks are identical. By applying the implementation invariance axiom, we make sure that if two networks produce identical output based on identical input, we assign credit to the same algorithms. Sundararajan et al. state that the integrated gradients method allows us to incorporate the sensitivity axiom and the imple-

⁴The authors note that the results obtained within the proposed framework, i.e., a formulated conjecture and possibly its proof, will only hold for the chosen sample of distribution P_z , not the whole Z .

⁵In geometry, symmetries are the properties that are unchanged under some class of transformations. This approach to invariants was introduced by F. Klein in 1872 (Tobies 2019, 6-7.).

mentation invariance axiom into the attribution algorithms, therefore allowing us to: 1) conduct a precise analysis of the connections between the input and output of the machine learning algorithms, and 2) change the correct features of the input with the goal to obtain the desired output. Davies et al. state that they decided to choose the integrated gradient technique because it gives the biggest importance to the elements of $X(z)$ and helps to understand how even small changes in $X(z)$ affect $Y(z)$. The information extracted with the gradient technique, which connects the changes in $X(z)$ to changes in $Y(z)$, is analyzed by a human mathematician who may choose some candidates for f' based on the observed patterns (or continue to change the hypothesis or the sampling distribution).

It is important to note that ‘artificial intelligence’ (AI) is a versatile term and can be defined and understood differently (Russell & Norvig 2010, 1-2). Davies et al. use the term ‘AI’ to refer to a collection of different techniques that can help with effective and fast data analysis, in this case, in the area of mathematics. The helpful techniques behind the notion of AI in mathematics include finding counterexamples (Wagner 2021), surveying complex mathematical structures (Lample 2019), accelerating calculations (He, 2021), etc. The authors name themselves the first people who used AI to assist in the discovery and formulating of theorems. Specifically, what the authors call ‘AI’ in their research are two specific methods of ML that they used: supervised learning and attribution techniques. Therefore, the authors use the notion of artificial intelligence to refer to the two specific machine learning algorithms and the framework in which they are implemented. However, ML algorithms and AI are not interchangeable notions. While machine learning techniques can be used by AI, they are simply a method of analyzing and extracting data (Jordan & Mitchell, 2015); AI, however, is a broader notion – there are different ways to define it, but, generally, it is a machine that can somehow simulate human or rational intelligence (Russell & Norvig 2010, Zhang & Lu, 2021). What Davies et al. call ‘AI’ is the yellow part of the flowchart depicted in Figure 1. There, AI is the complex system that generates a dataset of $X(z)$ and $Y(z)$ as defined in the input, allows to train a supervised model defining $\hat{f}$ that connects the required variables, and finds patterns and connections that could provide an explanation for $\hat{f}$. A human mathematician should be in close communication with the system on almost every step of the algorithm: they should provide and formulate the hypothesis, supervise the model training process, look through possible restrictions for $\hat{f}$, and, often, redefine the initial hypothesis and change the conjecture. In the end, it is the mathematician who proves the found conjecture.

Davies et al. do not give a precise definition of the notion of AI they use in their paper. Since they use the term ‘AI’ to talk about the stages in their flowchart that involve ML algorithms that analyze large datasets and make predictions based on the conducted analysis, we can establish that the authors do not talk about the human-modeling approach to AI: they do not use ML to imitate or reproduce human behavior or thought process. The usage of the

term ‘AI’ in Davies et al.’s paper is closer to the rationalist approach. The rationalist approach defines AI as an agent or system that is capable to think or act rationally (Russell & Norvig 2010, 2). The rationality of the system can be achieved through successful knowledge representation and processing. In the case of Davies et al., the rationality of the AI system can be defined by its ability to provide human mathematicians with data that helps them to uncover certain patterns that connect mathematical objects. What Davies et al. call ‘AI’ is a system that uses ML as a tool to obtain data about mathematical structures that are suggested by a human mathematician to exist, and then explore this data to present the found relevant regularities to a mathematician. The goal of a mathematician here is to find and prove a connection between two variables, the goal for AI is to present information in such a way that the connection can be clearly identified. The authors used the framework presented in Figure 1 to obtain two mathematical results: proving the existence of a relationship between algebraic and geometric invariants in knot theory and solving the combinatorial invariance conjecture for symmetric groups.

The framework suggested by Davies et al. is not a new approach to the form or the content of a mathematical proof, but rather a novel method suggesting where can a mathematician start when they only have a vague intuition about the existence of some mathematical relation. The authors suggest feeding this intuition to the AI system that would help us undercover patterns and regularities that would serve as a ground for the proof of the intuited connection. In this framework, a mathematician’s intuition is a starting point. The authors start with the premise that we already have an intuition about the existence of a relation between two entities: we do not yet fully understand the nature of this relation, but we suggest that it is there. We use AI with the goal to understand how to better pursue this relation: it indicates the patterns and entities we need to research more precisely to establish the specifics of the intuited relation. We start with the intuition of a human mathematician, and we use this intuition to present some input to the AI, which will help us develop this intuition further. Davies et al. describe this process as a mathematician’s intuition being *guided* by AI. However, one may argue that the process presented by the authors is not an example of guidance. First of all, one of the advantages of the presented method is that it allows us to process large amounts of information that we would be unable to process without a computer. Even if a human mathematician can briefly look through various examples of $X(z)$ and $Y(z)$, it would be hard for them to find specific patterns that could play a significant role in establishing the exact function connecting $X(z)$ and $Y(z)$. ML helps to establish the existence of these patterns easier and faster. However, can we talk about these steps as ‘guiding mathematical intuition’? One can simply say that ML allows us to reduce a large set of examples to a small set of relevant examples, and, accordingly, reduce the studied domain. Reducing a dataset we research can hardly be called ‘guiding’ the intuition, as it has nothing to do directly with the intuition; it only limits the scope of our research. The second problem with stating that AI guides a mathematician’s intuition is that the whole process consists of constant interaction between a mathematician and an AI system.

We start with a hypothesis, use it to generate the dataset, train the model to predict $Y(z)$, apply attribution techniques to understand the connection better, select a candidate for the role of f' , and constantly decide whether we want to repeat the process again, and, if we do, what settings to change in it. It is clear that the usage of AI makes our task of formulating the conjecture candidate easier as we find the needed connections faster, but why call it ‘guidance’? Why can’t we say that it is a *tool* that assists us in some stages of the work? Or, if its role is more than a tool, why not describe this process as a joint work of AI and a mathematician? To discuss both of these questions, we need to establish the kind of agency the AI system possesses in Davies et al.’s framework. To do that, we need to start by comparing the role software plays in classical computer-assisted proofs to the role AI plays in AI-assisted proofs, which we do in the next section. Then, in the next chapter, we discuss the agency of AI in Davies et al.’s proof and whether it is possible for it to guide a mathematician’s intuition.

3.2 AI-assisted proofs *vs* computer-assisted proofs

In this chapter, I want to compare two different types of assistance used by mathematicians in their work: classical computer-assisted proofs (discussed earlier in Chapter 1) and AI-assisted proofs presented by Davies et al. and discussed in the previous chapter.

The most basic feature that distinguishes the two types of proof is the kind of software used in them. In all classical computer-based proofs, like Robertson et al.’s (1996) or Gonthier’s (2005) proofs of the four colour theorem (both based on the initial proof by Appel & Haken, 1977), the ‘computer-based’ part of the proof is produced by a program which receives the needed conjectures and inference rules as an input and provides the formal proof of the theorem as an output. In Robertson et al., only a part of the proof is formalized, specifically, the proof of the reducibility of the list of configurations U . The authors used software that was responsible for going through U and confirming that every configuration in it is reducible. The authors wrote a program that was checking the reducibility by themselves, and they provide a link to the program files so everyone is able to redo the proof they presented (Robertson et al. 1995, 18). Robertson et al. also state that since they did not prove the correctness of the compiler or the infallibility of the hardware they used to run the program, their proof is not a ‘proof’ in a traditional sense (Robertson et al. 1996, 24). However, the program’s algorithm is still presented in an open source (Robertson et al. 1996, 17). The ‘computer’ part of Gonthier’s proof of the four colour theorem is also presented in an open source (Gonthier 2005, 1383), and a reader is able to go through it and, in case of an expert background, understand the formal rules used to proof the reducibility and unavoidability. However, in both of these cases, there remains the problem of the unsurveyability of the proof itself. The surveyability of the algorithm used to obtain the proof does not equal to the surveyability of the proof itself. The surveyability of the algorithm in Robertson

et al. and Gonthier gives more reliability to the proof, which distinguishes the authors' results from the original results of Appel and Haken (1977). In the case of Appel & Haken's proof, the algorithm of the exhaustive search implemented to prove the unavoidability of U was not presented in the published paper. Even if the algorithm was presented to the open public, readers would not be able to run it, as one would need a supercomputer for that. Therefore, in classical computer-assisted proofs, the content of the program that assists the proof is often open for the reader – one can rerun it on their own computer to make sure the program is reliable⁶. However, the possibility of a software mistake is always present: according to several experts who use computer assistants in their work (for instance, C. Lam, L. Thiel, S. Swiercz), when we rely on software in any step of the proof, we rely on our theory of how the software works. If that theory is wrong, for example, if the programming language is read differently by the compiler than how we expect it to be read, we may end up with incorrect results (Lam et al. 1989, 1122; Brown 2010, 35). Since there is no way to check whether our theory of how the software works is correct, the output of the program becomes, at least partly, empirical.

Davies et al. used Message Passing Neural Networks (MPNN) to implement machine learning algorithms on graphs⁷. The main idea behind MPNN frameworks is that the functions of the network are able to operate on different graph nodes. MPNN model introduced by Gilmer et al. (2020) operates specifically on directed graphs. The authors present the code they used in the proof (MPNN model definition, data loading functions, imported modules, etc.) openly⁸ and any interested reader is able to rerun it and check its validity. Therefore, both classical computer-assisted proofs and AI-assisted proofs include a step the algorithms behind which can be surveyed by the reader. However, the surveyability of the obtained proofs should be approached differently. The reason behind this distinction is the functions that the algorithms perform in the proof. In the case of classical computer-assisted proof, for example, Robertson et al.'s work, a computer program is used to do an extensive search through a large set of cases. The program runs, concludes no counterexamples are found, and the mathematician uses this output as a step in their proof. The mathematical conjecture that is being proved is formulated (or chosen) by a mathematician personally. The output of the program is used as a necessary step without which the chosen conjecture cannot be proved. In the case of Davies et al., the steps including machine learning precede the proof – we start with formulating the conjecture and finding the patterns between $X(y)$ and $Y(z)$ that could be useful in a proof. We may use the patterns we found with the help of ML to prove the chosen conjecture, but the most important implementation of ML is the help

⁶The notion of 'classical' computer-assisted proofs that I use includes all non-AI-assisted proofs: starting from Appel & Haken's proof and finishing with various sorts of formal and/or interactive proof assistants: *Metamath*, *Isabelle*, *Coq*, etc. However, when comparing the classical computer-assisted proofs with AI-assisted proofs, I will be mostly making examples of more modern proof assistants that allow full formalization.

⁷The model Davies et al. used is based on Gilmer et al. (2020).

⁸The code source for the proofs in knot theory and representation theory: https://github.com/deepmind/mathematics_conjectures.

with the formulation of the conjecture in the first place. The surveyability of the algorithm that was used to formulate the conjecture or to look for patterns between two mathematical entities is important, but if we understood these patterns with an independent mathematical insight and were able to provide a human-based proof to support it, our proof would be perfectly surveyable. The process of formulating the conjecture independently of AI would not be surveyable unless the mathematician describes precisely how the idea came to their mind, and even if they do, it is most commonly a description not relevant to mathematical practice (Carter 2017, 539-540). The characteristic of surveyability of the software used in the proof should be approached differently for AI-based and classical computer-assisted proofs since the proof strategies differ between the two. I am going to describe the differences in the proof strategies more precisely in the next paragraphs.

In addition to the differences in the software itself, classical and AI-based computer proofs *use* the software differently. In classical computer proofs, a computer program can be used in two ways: 1) to perform an exhaustive search through a large set of examples, and 2) to obtain a formalized proof of a theorem (or a chosen conjecture that helps to prove the theorem). In both cases, the software is used directly in the process of obtaining the proof. It is either implemented to verify a specific statement (for example, to verify the unavoidability of U in Appel & Haken 1977) or to formalize the whole proof (like in Gonthier 2005). Without that step, the proof cannot be completed at all – in several cases, the computer-assisted step is what the novelty of the proof is based on⁹. The original idea of how to pursue the proof, what conjecture to choose, and what functions and patterns we should research to complete the proof is suggested and chosen by the authors of the classical computer proofs. The structure of the proofs that use software to go through a large set of examples, like Appel & Haken’s or Robertson et al.’s proofs, is built by the authors. Even the fully formalized proofs, like Gonthier’s proof of the four-colour theorem, are directed by a mathematician – a mathematician should choose what formal rules and lemmas to present as an input for the program, and what formal language to choose so that the program will be able to process the proof correctly (Huet et al., 1997).

In the case of AI-assisted proofs, the usage of the software is essentially different. First, there is no specific limit to which ML can be used in the proofs: the process Davies et al. present is cyclic, and the algorithm can be used several times to find the conjecture the author wants to prove and the variables and structures responsible for it. The process of finding the conjecture and the patterns is, as in the classical computer-assisted proofs, directed by a mathematician: they decide whether to change the dataset of examples of $X(z)$ and $Y(z)$ or whether to reformulate the original hypothesis based on how

⁹For example, Appel & Haken based their proof of the four color theorem on the previously introduced notion of reducible configuration suggested by Birkhoff (1913). The general structure of Appel & Hakken’s proof was based on Birkhoff’s reasoning, therefore the only novel step that the authors introduced was to use a supercomputer to go through the set of configurations U .

they evaluate the performance of $\hat{f}$. However, the mathematician does not have an initial plan for the proof when they implement ML into their work. The mathematician uses ML to construct the plan of the proof. An agent at the beginning of the framework presented in Figure 1 knows neither which specific objects and patterns in mathematics they are going to implement nor the specific formulation of the conjecture they are going to prove. They start with an intuition about the existence of a connection between two variables and try to use ML as a tool to understand which mathematical structures connect these variables. At the stage of formulating and writing the proof, when the mathematician understands the specifics of relations between $X(z)$ and $Y(z)$, the author of the proof does not need to implement a computer program to verify a lemma without which they cannot continue to the next step¹⁰. Therefore, the formal proof itself may be independent of the computer program, but the originality of the idea behind the proof is not. The extent to which the idea behind AI-assisted proof is dependent on the program and the extent to which it is dependent on the mathematician who was supervising the program is a subject for discussion, but it is certain that ML plays a significant role: without it, a mathematician would not be able to analyze the large dataset of examples for $\hat{f}$, and, subsequently, would not be able to identify any specific patterns that are worth looking closer at. We can see that the roles of software in AI-assisted proofs and classical computer proofs are not at different places in some hierarchy, but rather they approach the proving process from different ends: AI allows a mathematician to plan the structure of the proof by helping to analyze a large dataset, and non-AI proof assistant allows a mathematician to overcome a significant step in an already-planned proof by going through a large set of cases and verifying the absence of counterexamples.

It was already discussed in the previous paragraphs of this section, but it is worth mentioning separately that the initial goals of the mathematicians in AI-assisted and non-AI-assisted proofs are essentially different, which affects the discussion that the obtained proof initiates in philosophy of mathematics. In the classical computer-assisted proof, an initial goal of a mathematician (let us call them *mathematician A*) is to *prove* a theorem. At one point, mathematician *A* decides to use a computer program to assist them, but the initial goal is still the same: prove the chosen theorem. Now, suppose, the mathematician successfully implemented a computer program at some step of the proof to perform an extensive search through a large set of examples. Once the computer search was implemented, mathematician *A* was able to obtain the intended result: prove the chosen theorem. The questions the obtained proof can raise in the mathematical community are directly connected to the result of the proof. For instance, some authors may not consider the proof to be as reliable as regular human-based proof due to the unsurveyability of software mistakes (Lam 1990, Brown 2010). Some authors may accept the result of the proof, but not consider it a significant progress in mathematics due to the lack of new methods

¹⁰Unless, of course, there would be a proof combining AI-guidance of a mathematician's intuition and a classical computer-assisted step in the proof itself.

and patterns introduced in the proof (Barrow 1995, Thurston 1994). In both cases, the parameter that is being evaluated is the result of the proof: either the formal result – the theorem was proven, or the less formal, extended result – the mathematical community learned about new connections and methods in mathematics. In the case of AI-assisted proof, the original goal is to *understand* where to start with the proof, and how to correctly *formulate* the theorem that we will be proving. Suppose, mathematician B decides to use AI to help them understand the relations between a chosen pair $X(z)$ and $X(z)$. Mathematician B goes through several cycles of reformulating the hypothesis, changing the dataset, and training the function, and finally ends up with a conjecture that they know how to prove that connects the needed variables. Mathematician B proves it (by using the connections revealed to them by AI assistance) and presents the proof to the mathematical community. The questions the proof presented by mathematician B may raise are concentrated not on the final result of the proof (the proved theorem or the reveal of new mathematical methods) but rather on the authorship and originality of the proof. For instance, we can reconsider the notion of scientific understanding in mathematics (Krenn et al. 2022, 765) or the subjective role of a mathematician in the process of implementing mathematical intuition. To conclude, the philosophical discussion of classical computer-assisted proofs is concentrated on the validity of the obtained result, while the philosophical discussion of AI-assisted mathematical proofs is concentrated on the creative participation of AI in mathematical intuition.

I should also note that several points of the expressed criticism of classical computer-assisted proofs (Parshina 2023) can be used to argue against AI-assisted proofs as well. For instance, the unsurveyability of the algorithm used to obtain proof is still a problem: if we use ML to help us formulate the conjecture, we cannot show a specific line of thought we pursued. The problem even becomes more significant since we use ML: there is no unified understanding of how exactly neural networks work (Lillicrap & Kording, 2019). That means that we use a tool that we are not fully capable to understand, and, therefore, we are not fully capable to control it. There are still problems with the discussions centering on the fact that mathematics becomes partly empirical with the use of any kind of software in order to obtain a mathematical result.

Nevertheless, the discussion of the originality and insightfulness of a mathematical proof obtained with computer assistance seems to gain a new turn: first of all, Davies et al.’s example shows that computer assistants can be used during the planning stage of the proof and not during the proving itself. That means that a proof can be computer-based while there is no such step in it where a computer program does an extensive search through a large set of cases. It leaves room for the human mathematician to introduce new methods and approaches that, according to Barrow and Thurston, are responsible for the insightfulness and novelty of the proof. Moreover, in the case of Davies et al., AI software was used specifically to obtain insight to initiate a proof, or, as Davies et al. formulate it, to “guide mathematical intuition” (Davies et al. 2021, 70). That means that computer software can be used in a mathematical proof in such a way that the proof does bring new methods and connections into

mathematics – in this case, it is the direct role of the software. However, the source of this insightfulness is unclear: is it the mathematician who invents a novel approach with the help of an AI assistant, or is it the AI assistant who is solely responsible for the novelty of the result? To continue this discussion, we must establish how the notion of agency can be applied to the case of Davies et al. Specifically, we need to establish how agency is divided between the human mathematician and the AI. Moreover, as we have seen in Chapter 2, the notions of mathematical insight and intuition are strongly connected to the agency of a mathematician as well. Once we establish how agency should be divided in AI-assisted mathematical proofs, we will be able to discuss the application of the notions of insight and intuition to the role of AI. That is what we are going to do in the next chapter.

Chapter 4

The Place of AI: Between Insight and Intuition

4.1 Agency of AI

In the previous chapter, we discussed the role AI plays in Davies et al.'s proof and how it is different from the role of software in classic computer-assisted proofs. We established that the difference between the use of software in the two types of proof is based on the differences in the initial goal of a mathematician implementing it into their work. Classic computer assistants are implemented into a pre-existing proof structure: the strategy of the proof already exists, a mathematician uses computer software to go through a certain step in the proof; AI assistants are used when the structure of the proof does not yet exist, sometimes even when the statement that one wants to prove is not yet fully established: a mathematician uses AI to precisely formulate the statement they are going to prove and study the patterns that may be helpful to establish the proving strategy. The differences in the ways software is used in these two types of mathematical proof give rise to differences in the critical discussion about the validity of these proofs. We concluded that the critical discussion about the use of classical computer assistants in mathematical proofs is concentrated on the validity of the obtained result, while the critical discussion of AI assistance in mathematical proofs is concerned with questions of creativeness, authorship, insightfulness, and other concepts connected to the notion of agency applied to a human mathematician and AI. In this section, we are going to discuss whether the notion of agency can be applied to the AI used in Davies et al.'s work. We are going to compare it to the agency of a human mathematicians and establish whether human and artificial agencies can co-exist in AI-assisted proofs. I must notice that the question of agency in AI cannot be discussed in full in just one section of this thesis: it is a problem deserving several books to be written about it. Hence, I will not answer the broad question “Can AI be attributed agency?”, but try to answer one with a smaller research scope – “Can the AI Davies et al.

used be attributed agency in mathematical practice?”. To answer this question, I will discuss some literature on the agency of AI. However, since there is yet no established literature tradition focusing specifically on AI’s agency, I will mainly rely on the classical accounts in philosophy of action.

In section 3.3, we discussed the connection between the perceptual account of mathematical intuition and human agency. We have established that mathematical intuition, as it is described by Parsons and Chudnoff, is directly connected to the internal mental states of a human mathematician. It is clear that a person who obtains and uses mathematical intuition to conclude a proof exhibits some kind of agency in applying mathematical intuition and proving a theorem (if it is not clear, however, we discuss it in the next few paragraphs). But can we say that the AI system used in Davies et al.’s work exhibits some kind of agency as well? Can we say that it performs an intentional act by finding connections between the mathematical entities presented at the input, or does it merely work as a useful tool for a human agent? To start this discussion, we need to establish the definition of agency we are going to use.

First of all, we may establish that a human mathematician who comes up with the intuition about a connection between two mathematical entities, supervises the training of a model, formulates and reformulates the conjecture, and, in the end, proves the theorem, is, certainly, an agent. They act and choose how to act, both while they are formulating the final version of a conjecture with the use of AI and while proving the resulting conjecture. It does not matter what theory of action we choose to define the notion of agency – the causal theory of action, irreducible agent-causation, the teleological approach – the mathematician could always be attributed some sort of agency. According to the causal theory of action (Davidson 1971, Anscombe 1995), a mathematician is supposed to have a primary reason for their action – in this case, it is the desire to formulate the conjecture and finish the proof, or a belief that they are able to do so, – and an instrumental belief – a belief that the actions they choose to take will eventually lead to the proof. It would be unreasonable to suggest that a mathematician does not have the desire to prove a conjecture they are trying to prove or that they would not have a belief that the steps they choose to take will lead to the correct proof. Hence, the causalists’ notion of agency describes the role of a human mathematician successfully. We will also have no problem if we try to define the agency of a mathematician with a descriptive account of action (Chisholm 1964): they formulated the conjecture and found its proof because they intended to do so – to formulate the conjecture and find the proof, therefore, it was an intentional action. If we try another alternative theory, for example, the teleological theory of action (Wilson 1964, Sehon 2005), we would still be able to define a mathematician as an agent: their action is defined by their initial goal to obtain the proof, and they acted because they believed that their actions could lead to this goal. To conclude, we have no problems attributing agency to a mathematician implementing AI into their work.

It is not clear, however, whether the AI system used in Davies et al.’s proof can possess any kind of agency: can we say that it acts independently, or is it

simply a tool mathematicians use to get help with the proof strategy? We can distinguish three different ways the interaction between a mathematician and the AI can be described in the case of Davies et al.’s proof:

- (1) a human mathematician and AI are two independent agents acting together;
- (2) a human mathematician and AI form an epistemic group: they act as one agent;
- (3) a human mathematician is an agent and AI is a tool they use.

In the case of (1), we consider AI to act as an agent on the same level as, or at least close to, a human mathematician. Human mathematician manifests their agency by pursuing the relation they anticipated with mathematical intuition, supervising the model, formulating the final conjecture, and, finally, proving the formulated conjecture. AI, in its turn, manifests its agency by analyzing a dataset of pairs $(X(z); Y(z))$, predicting the following values of $Y(z)$, choosing the patterns and relations that may be essential in formulating the conjecture that establishes a connection between $X(z)$ and $Y(z)$. A mathematician and AI are two independent agents: each of them is responsible for some steps in the flowchart of Davies et al. The formulated and proved conjecture is a result of their joint cooperative work.

Now, we should come back to the theories of action we have chosen to consider in this section. The most widespread approach is represented by a number of causal theories of action described in D. Davidson (1963, 2001), E. Anscombe (1957, 1955), M. Bratman (1987), and K. Setiya (2011). According to the general causal account, an intentional action is identified as such because it is causally connected to certain mental states (Glasscock et al., 2023). The specificity of mental states varies from author to author. According to Davidson, one of the first authors to formulate the causalist account, intentional action has a primary reason as its cause. The primary reason consists of two sides: the initial belief or desire driving the agent to complete the action, and the belief that the actions the agent is going to pursue would allow them to fulfill that desire (Davidson 1963, 686). The mental states constructing the primary reason are the causal reasons for the behavior of the agent: an individual acts because of the possession of these mental states.

The alternative approaches, such as the descriptive or teleological accounts, also refer to specific mental states in their description of agency. For example, Anscombe, who is often considered to be a non-causalist, explains the nature of action through the notion of practical knowledge: an action is intentional when an agent understands what they are doing (Anscombe 1957, §23). In the descriptive account, the cause of an action is the agent’s intention to make the action: a person is “making A happen with an end to make A happen” (Chisholm 1964, 616-617). So, according to Chisholm, if an agent commits an act A intentionally, they act with a mental state of ‘intending for A to happen’. In the case of the teleological account, an individual commits an intentional act

A with a specific goal that is intended to be the main result of A . The existence of a specific goal or purpose also refers to a certain mental state of the agent. All of the presented accounts describe various mental states as the reason or explanation of the agent’s intentional actions: understanding what one is doing, understanding why they do it in this particular way, and intending to make the action. Hence, all the most widespread accounts of agency connect this notion to the mental state of the agent.

We can also approach the notion of agency from a more specialized, although less traditional background: the accounts of artificial agency (or the accounts that allow artificial agency) described in L. Floridi & J. W. Sanders (2004), Barandiaran et al. (2009), and E. Popa (2021). Floridi & Sanders (2004) distinguish three features of an artificial system that can be attributed agency: interactivity, adaptability, and autonomy (Floridi & Sanders 2004, 357). Interactivity refers to the possibility of the system to influence its environment and be influenced by it; adaptability describes the ability to learn from previous interactions; autonomy amounts to the system’s ability to operate on its own, without human intervention. Barandiaran et al.’s account of “minimal agency” also outlines three conditions for an artificial system to be attributed agency: 1) the system is an entity distinguished from the environment, 2) in that environment, the system is doing something by itself, and 3) it does it according to a certain goal (Barandiaran et al. 2009, 369). For Popa (2021), the agency of AI highly relies on human norms and morals in the first place. However, according to Popa, even if we can attribute agency to AI, we can only do so for highly-autonomous AI systems: “Thus, the question of agency only arises for those AI systems possessing autonomy in the sense of high-level abilities” (Popa 2021, 1738). Popa refers to the distinction between high-level and low-level artificial autonomy outlined by C. P. Ezenkwu and A. Starkey (2019). According to the authors, low-level autonomy consists of perception, learning, context-awareness, and decision-making, while high-level autonomy includes self-motivation, domain independence, self-recovery, and self-identification of goals. High-level autonomous system is capable to work almost independently of a human agent: it can identify its domain by itself, make its own decisions, and choose its own goals. We can conclude that all the presented accounts of agency in AI are based on the condition of autonomy – the ability of the system to work independently of human intervention. Hence, if we consider the traditional (but general) accounts, the necessary condition to have agency is to be able to have mental states, and if we consider the accounts specified for AI (but not widely established), the necessary condition for agency is to have a high level of autonomy. In the next part of this section, we will see whether any of these conditions work for Davies et al.’s AI.

Here, if the current paragraph could have an infinite length, we would discuss whether it is possible for AI to possess mental states, and, therefore, some sort of agency. However, since the topic is way too broad, we are only going to talk about the AI used by Davies et al. and its characteristics. What the authors call ‘AI’ is a group of datasets, ML models, and other algorithms that are able to make predictions about the patterns mathematicians need to consider in

order to research a chosen mathematical relation. The mathematician chooses the relation they are going to pursue before implementing AI into their work. When a mathematician uses AI to develop their understanding of that relation and achieve a clearer formulation of the conjecture, it is the mathematician who possesses the mental states necessary for intentional action. It is the mathematician who chooses the goal (formulate and prove conjecture $\hat{f}$), it is the mathematician who has the desire to pursue the proof, it is the mathematician who believes that the action they chose will lead to the precise formulation of f' . It is also the mathematician who decides to use AI to assist them. Hence, the actions of choosing the relation, obtaining its precise formulation, and proving that this relation exists are intentional because of the various mental states of the mathematician. Davies et al.'s AI is neither involved in making decisions about what path to choose to formulate the final conjecture nor does it have intentions or beliefs about its actions. It is yet impossible to design an AI system that is able to self-reflect (Ng & Leung 2020, Maetschke et al. 2021), and Davies et al.'s AI is not an exception. Even if we choose a Dennett-like approach to the problem and say that everything that behaves as an agent can possess agency, Davies et al.'s AI would still not fit the criteria. If we look at the framework of mathematician-AI interaction in Davies et al.'s case, AI does not seem to perform any actions in it by itself: it is always a human mathematician who chooses the way of development, and AI's actions are defined by the mathematician. Hence, Davies et al.'s AI cannot possess agency in mathematical practice, according to the traditional theories of action.

Following the same reasoning as in the previous paragraph, we cannot attribute agency to Davies et al.'s AI based on the more specified theories of artificial agency. We have seen that high-level autonomy is a basic criterion for an AI system to be an independent agent, and Davies et al.'s AI clearly lacks it. It does not set its own goals (a mathematician formulated a hypothesis), it cannot independently adapt to the domain it works with (Davies et al. had to reprogram it for a mathematical domain), and it cannot work independently of a human agent (during the process Davies et al. described, human intervention was required on every step, often repeatedly). Even if this AI could be attributed autonomy, it would be at most a low-level autonomy as described in Ezenkwu & Starkey (2019). And this level is not enough to be attributed agency, according to the accounts we considered. Therefore, the AI Davies et al.'s used cannot be attributed agency neither based on the traditional general accounts of agency nor based on the more specified accounts of artificial agency.

We have established that AI cannot be treated as an independent epistemic agent interacting with a human mathematician. However, maybe, we should not try to describe a mathematician and AI as two independent agents: we may choose to describe them as a group agent, as in the case (2). Group agency is a notion that describes a group of individuals that act as an agent: a group brings a change in its environment by responding to stimuli based on the group's beliefs, intentions, and desires. A belief of a group is not equal to the sum of the beliefs of all its members. K. Mathiesen (2011) gives the following example: a mother thinks that the best age for their daughter to start dating is 18, father thinks it

is 14. Together, they decide to set the age at 16. The final decision of the mother and the father as a group is not equal to any of their individual positions. It may appeal to us to use group agency to describe the work of a mathematician and AI as it does not require AI to possess any individual agency: only to be a part of the group one. However, there is at least one essential requirement for the members of a group that acts like an agent: an active agreement of every member to be a part of the grouped agent, which is, to act intentionally, together with other members, as a group. We have already briefly discussed what is an intentional action: it may be defined differently, but it connects the agent's action to their mental state (understanding why they do it, why they do it in a particular way, etc.). According to C. List and P. Petitt (2011), most group agents are formed through members “jointly *intending* to form a group agent” (List & Petitt 2011, 57). P. French (1995) and K. Mathiesen (2011) describe group agents as groups the members of which actively agreed to be a part of the group (by agreeing to abide by certain collective rules). Hence, to form not simply a group, but a group agent, every member of the group should make at least one intentional action: agree to be a member of it. However, as we have seen earlier, we cannot ascribe any kind of intentional action to AI designed by Davies et al. Hence, it would not be able to have an intention to form an epistemic group. If we tried to describe AI and mathematicians with the notion of group agency, all the intentional action on both stages – agreeing to be in a group and pursuing the actions as a group – would be initiated and performed by a mathematician. We might as well attribute agency to a group consisting of a mathematician and a calculator and insist that the formed group possesses agency because the mathematician possesses agency. AI would not be able to add or reform any intentions that a mathematician has, hence, it cannot be described as a part of a group agency.

We have established that we cannot attribute agency to the AI implemented by Davies et al., nor can we unify the AI and the mathematicians who used it into a single group agent. What we have left now is the option to regard AI as a tool an agent (mathematician) can use in their work. We have established in section 3.2 that the role software plays in AI-assisted proofs is significantly different from the role software plays in classic computer-assisted proofs: AI helps a mathematician to understand what functions and variables one needs to take into consideration to formulate a conjecture they are going to prove, and classic computer assistant either helps to formalize the proof of an already existing conjecture or to go through a crucial step in the pre-planned proof. Classic computer assistant functions as a tool that helps to obtain the proof of a conjecture already formulated by a mathematician, following a proof plan chosen solely by a mathematician. AI assistant is also a tool, but it provides help at the stage that requires creativity: the pre-proving stage, where a mathematician makes use of mathematical intuition and insight to determine what relations and entities they will pursue in formulating and, subsequently, proving the conjecture. AI-based and classical assistants both function as tools, but AI provides support in the development of an idea a mathematician chooses to pursue, not in the implementation of that idea. Hence, we may conclude that even if we do

not ascribe AI any agency in producing mathematical intuition and insight, its function as a tool is essentially different from the function of classic computer assistants. In the next section, we are going to discuss the nature of assistance AI provides in mathematical proofs. We will establish whether we can describe the role AI plays as assisting (in any way) in obtaining mathematical intuition and/or insight.

4.2 Intuition and AI

In sections 2.2 and 2.3, we have established that the most comprehensive approach to mathematical intuition is the perceptual approach developed by Parsons and Chudnoff: it describes the entire mathematical practice (not just some part of it, such as, for example, mathematical education) and it characterizes a wide variety of forms of mathematical intuition (not, for example, just a spontaneous visual impression). We have also shown in section 2.3 that it is necessary for an individual to possess agency and be able to obtain intentional mental states to possess mathematical intuition. In section 4.1, we established that we are unable to attribute agency and all the implications following this concept (e.i. internal mental states, unique subjective experience, etc.) to the AI used by Davies et al. If we cannot attribute agency to AI, we cannot make it the subject of mathematical intuition: we cannot talk about the ability of AI to directly *possess* mathematical intuition. However, this conclusion does not yet entail the complete disconnection of AI from the framework of mathematical intuition. In the previous section, we suggested that AI can be used as a supporting tool in obtaining intuition and/or insight in mathematics. In this section, we will determine how exactly AI can provide assistance in obtaining mathematical intuition.

Davies et al. describe the performance of AI in their proof as ‘guiding’ a mathematician’s intuition (Davies et al. 2021, 70). Under the term ‘guidance’, the authors understand the possibility to use AI as a tool that allows better pattern recognition and expands the interpretation methods that a human mathematician has at their disposal. AI cannot directly play the role of an agent achieving mathematical intuition. Davies et al. write: “In mathematics, it has been shown that AI can be used as a valuable tool by finding counterexamples to existing conjectures” (Davies et al. 2021, 70), which confirms that they regard AI as a tool providing assistance to a mathematician. What this tool does is come up with suggestions for the possible grounds of the connection between the two chosen mathematical entities.

Now, we must return to the definition of mathematical intuition we refined in Chapter 2. According to the perceptual approach, mathematical intuition is a process, close to perception, during which one experiences a mathematical object as if it was presented to them in their immediate environment. If we have an intuition about an object, we are convinced that it can be constructed in mathematics. It does not mean, however, that we know exactly how to do so. Mathematical intuition cannot help us reveal any structural connections we

need to prove a conjecture about the intuited entity, and mathematical intuition cannot be used as a tool that we use to come up with a proof: generally speaking, mathematical intuition cannot help us obtain any explicit, structural mathematical knowledge. Mathematical intuition can answer the question “Is there an object X in theory Z ?” based on the experience of the intuiting mathematician, but it cannot answer the question “How do we prove there is X ?”. As we discussed in section 2.6, one important point in distinguishing mathematical intuition from mathematical insight is that the former does not provide any explicit knowledge concerning the object one is intuiting. Since no direct knowledge can be obtained with the use of only mathematical intuition, intuition cannot be used to justify any mathematical result. Here, we come to the following question: how Davies et al. can use AI to guide mathematical intuition if what AI does is assist in finding structural connections in the relation that is intuited by a human mathematician? How can it structurally guide an object-oriented experience that has nothing to do with the formal justification of the object? Even if the authors used the word ‘guidance’ to describe the instrumental, tool-like work of AI that a mathematician uses to investigate the relation they intuit, how can this work assist in obtaining mathematical intuition? Since we already had some relation to investigate with AI, intuition was already obtained. I suggest that in cases like those presented by Davies et al., AI assists not in guiding mathematical intuition, but in formulating a clearly structured statement concerning the object experienced with mathematical intuition.

Intuition about a mathematical object can be the first step in the process of formulating and proving a mathematical conjecture. An intuition about two different mathematical objects can be formed into an intuition about the relation between the two objects (the so-called intuition *that* described by Parsons). The newly formed relation can also be regarded as a token to which the intuition is directed. However, it is not the case that a mathematical intuition about a certain object can be developed into a clear mathematical statement about the object. Intuition can be used as a starting point. For instance, a mathematician intuits a relation between two mathematical objects. Now, the mathematician can try to pursue this relation by studying possible patterns that could be used to define the relation, searching for similarities in the definitions of the objects, researching theories that could describe both objects in a similar way, etc. The mathematician would try to approach the intuited relation from a justified point, using the already proven theorems and conjectures. At this point, however, the mathematician would stop working with the mathematical intuition. They would be studying the object the existence of which they suggested based on their intuitive experience, but they would not be applying mathematical intuition any further. Mathematical intuition cannot be developed into a formal statement about the existence of the intuited object: the process of obtaining a mathematical intuition of a certain object is finite, and it ends when a mathematician experiences it. It is the case that several intuitions *of* may be used to compose an intuition *that*, but the resulting intuition would still not bring any mathematical knowledge to the table: only the assurance of the mathematician about a certain proposition they may regard to be intuitively true. The

mathematician can come back to the experience whenever they would like, but they cannot solely develop this experience into a proven statement about the object. They can use any possible scientific approach to justify the existence of the object or the truthfulness of the proposition, but, in any case, they would be working outside of the scope of only mathematical intuition.

One may suggest that what AI does in Davies et al.’s proof is developing, or helping to develop several intuitions *of* into one intuition *that*: a mathematician chooses some objects they intuit to exist in mathematics as the input for AI, and AI suggest some ways to form them into an intuition about a specific proposition. I disagree with this explanation. First, in the authors’ example, mathematicians do not choose random objects they intuit as input for AI, they choose the objects that, according to their intuition, are related in some way. So, mathematicians initially have some intuition about the proposed relation. I believe that the intuition about an unspecified relation between two entities would still be regarded by Parsons as intuition *of*, not intuition *that* because it does not have a form of a specific proposition that forces itself upon us as being true (it is not yet even a proposition). Second, according to Parsons, intuition *that* describes a state when it is intuitively obvious to us that a certain proposition is true. Now, if we suggest that we use AI to form an intuition *that* about some objects, it would mean that we somehow use AI to achieve the state where it is intuitively obvious to us that f' is true. However, that is not the assistance AI provides. As we have discussed in section 3.1, there are two main points of work AI does: first, it helps to correctly define the conjecture we are going to prove, and, second, it helps to find the specific patterns and connections we need to prove the conjecture. Even if AI would help us to formulate a correct proposition about two or more mathematical objects we intuit, the resulting conjecture f' would not be intuitively obvious to us. Davies et al. use AI to reveal patterns that are hardly noticeable to the human eye because of the big amount of data that needs to be processed to find these patterns. That is why the relations AI helps to reveal would hardly be intuitively obvious to any human mathematician. It is not the case that mathematicians cannot understand the relations that were defined with the help of AI: they can, and they do (as we see in Davies et al.). They just do not find them intuitively obvious: if they did, we would not need AI to assist in the verification of these relations. So, even if we treat mathematical intuition in Davies et al.’s proofs as intuition *that*, we would still not be able to explain the role of AI as ‘assisting (or guiding) mathematical intuition’.

Since we established that mathematical intuition cannot be guided or developed into a clearly structured definition of the intuited mathematical entity, we cannot describe the role of AI as guiding, developing, or expanding a mathematician’s intuition. AI assists in searching for patterns and structures that could be used in the formulating of a mathematical statement concerning the intuited object. Mathematical intuition has nothing to do with the explanation or verification of the intuited mathematical entity – it is only responsible for the suggestion that there is such an object. Since AI assists in establishing the exact structure of the object, we cannot describe it as a tool assisting a math-

ematician’s intuition. AI assists a mathematician in their work of finding a suitable formulation and justification of a conjecture about the intuited object, but it does not assist or guide a mathematician’s intuition. Therefore, we can conclude that despite the formulation the authors use to describe the work AI does in Davies et al.’s proof, their AI does not help to obtain or develop mathematical intuition: it helps to approach the object of mathematical intuition by providing possible grounds for the correct formulation of a conjecture.

It is possible that AI provides assistance not to the ‘philosophical’ mathematical intuition, but to the phenomenon we described in Chapter 2 as ordinary, ‘day-to-day’ mathematical intuition. However, in sections 2.3 and 2.4, we established the two main points concerning this problem. First, the ordinary notion of mathematical intuition does not have much to do with the philosophical account of intuition in mathematics – if something, it is closer to the philosophical account of mathematical insight. Second, even though the day-to-day notion is closer to the concept of mathematical insight, it cannot be defined precisely: different mathematicians, authors, and teachers put different meanings into this concept and use it briefly. Hence, even if it is the case that AI guides ordinary intuition, we would not be able to establish it without the specific definition of ordinary intuition. Since we established that AI provides assistance in the structural design of mathematical statements and their justification, we may suggest that it provides assistance in obtaining mathematical insight – the phenomenon that allows a mathematician to obtain a structural understanding of a mathematical entity. In the next section, we will discuss whether AI, as it was used by Davies et al., assists in obtaining or guiding mathematical insight.

4.3 Insight and AI

In the previous section, we suggested that the assistance AI provides in mathematical proofs is focused on helping in obtaining mathematical insight, not mathematical intuition. In the present section, we are going to discuss the part AI takes in obtaining mathematical insight.

In section 2.4, we discussed MacDonald’s and Griffiths’ definitions of mathematical insight. The authors connect the ability to produce insights to the existence of a mathematical model a mathematician can understand and interpret. To obtain a mathematical insight is to obtain the understanding of a model describing the basic functional core of a theory, a connection between two theories, providing a representative example of a certain mathematical structure, etc. Moreover, according to Nathan and Walkington, this understanding should be traceable: a mathematician should be able to provide a written solution to a problem they considered. We also discussed that, according to MacDonald and Griffiths, the ability to produce mathematical insight can only be attributed to a human mathematician. The authors describe the ability to obtain mathematical insights as something only humans can achieve: when an agent obtains an insight, they obtain a deep structural understanding of a theory, which means that an agent acquires a certain mental state. According to MacDonald and

Griffiths, an artificial agent is unable to obtain an understanding of a model, and, therefore, is unable to obtain mathematical insight. There are several studies examining whether it is possible to attribute consciousness or at least ‘an ability to understand’ to AI that would support MacDonald and Griffiths’s position. Most of the researchers agree on the fact that we cannot ascribe consciousness to AI at its current level of development – it would be possible only for strong AI, which currently does not exist (Baum et al. 2011, Adamson 2012, Neville-Neil 2016, Armstrong et al. 2016, Ng & Leung 2022). Maetschke et al. (2021) conducted research evaluating the ability of currently existing AI systems to *understand* the information they are processing. The authors suggested that ‘AI that understands’ should be able to not only disclose statistical relations within the processed data but also form a human-like understanding of the data, which could be expressed in different forms, for example, distinguishing correlation from causality, possessing the ability to self-reflect on their work and its result, ability to explain the reasoning behind the obtained result (Maetschke et al. 2021, 2-3). According to Maetschke et al., some beginner steps in developing ‘AI that understands’ are currently being taken. For example, there exist AI systems that are able to jointly process multi-modal data (visual, audible, and textual). However, an AI that can obtain a human-like understanding of the information it processes does not yet exist. Maetschke et al.’s research supports the position of MacDonald and Griffiths: AI cannot perform the role of an active agent in obtaining mathematical insight. An agent obtaining a mathematical insight accesses it through the understanding of a model, which is a certain mental state. AI that was used by Davies et al. cannot obtain this type of human-like understanding of models, just like any other currently existing AI.

Since MacDonald and Griffiths only discuss the notion of insight, not the characteristic of insightfulness, we should add that AI also cannot be insightful. In section 2.5, we discussed the notion of insightfulness in mathematics defined by Carter. Even though Carter pointed out that insightfulness is not always connected to the strive for truth, he still connects this characteristic to an intellectual virtue. Namely, cautiousness in selecting correct insights from incorrect ones. As we have seen in section 3.1, Davies et al.’s AI produced both valuable and invaluable patterns to explore: that is why the process had to be supervised and repeated several times before a final conjecture could be established. The process of distinguishing valuable and invaluable patterns was done by a mathematician. According to Carter, a mathematician could be called insightful, but not the AI. However, since we critiqued Carter’s explanation of how insightfulness is formed, we may suggest that it is not the case that insightfulness should always be connected to a mathematician’s cautiousness. We, nevertheless, pointed out that we agree with the attention Carter gives to the previous experience of a mathematician: insightfulness is always based on a significant amount of practice and experience. The insightfulness of an agent is characterized by the agent’s ability to transfer this experience into valuable insights. Now, even if we suggest that AI can have some sort of ‘experience’ in mathematics, for example, if it is based on a self-learning network, it would not

be able to produce insights, as we established in the previous sections. And, even though insightfulness may not be defined as an ability to produce insights, it is the most important feature of an insightful agent. Hence, AI cannot be insightful as well.

In section 4.1, we established that AI cannot function as an independent agent in obtaining mathematical intuition or insight. In section 4.2, we concluded that AI also cannot function as an assistant in obtaining mathematical intuition: AI provides help in structuring and formulating a conjecture that is to be proven, and mathematical intuition does not contain any structural or formal content. Now, we established that we cannot claim AI to possess a human-like understanding of the data it processes. Therefore, we can conclude that AI cannot neither obtain mathematical insights nor possess insightfulness. However, we have seen that the assistance AI provides is aimed at making the relation a mathematician studies more articulated, or formalized. This type of assistance is close to the processes happening as a result of mathematical insight, as described by MacDonald, Griffiths, Nathan, and Walkington. We suggest that AI functions as an instrumental tool by helping a mathematician to construct a model, and, consequently, obtain mathematical insight.

I propose to describe the role of AI in mathematical proofs (as it is presented by Davies et al.) as *assisting in building a mathematical model* based on which a mathematician can obtain mathematical insight. It is not the case that AI can build a model itself. In Davies et al.'s work, AI is used to find patterns that could function as a ground for formulating f' , but the mathematician themselves formulates the final conjecture. As discussed in section 4.1, we cannot attribute any human-like creative agency to AI. However, there is no requirement for an exclusively human-like agent to be responsible for the construction. A mathematician can obtain insight by using the tools that help them make the information they study more representative. If we define a model, following MacDonald, as a representative example of a theory, we are not restricted by the internal mental capacity of the agent: a representative example can be produced by a computer. For example, A. Baker suggests that the role computer takes in mathematical proofs is described as computing instances – checking the validity of a set of cases within a hypothesis (Baker 2008, 338). A computer can be used by a mathematician to find or visualize a representative example: however, the properties of the example case should be chosen by a mathematician, who as well evaluates the representativeness of the example (Palais 1999, 647-648). However, AI does not ‘compute instances’ like a classical computer assistant. It provides a different kind of help, as we have seen in section 3.2, where we compared the functions of the two different types of computer assistants. AI provides assistance in discovering patterns that would be helpful in formulating and proving a final conjecture. It does not generate any specific examples that would represent a theory in a way that would allow a mathematician to easily grasp its main features. Therefore, AI does not help to achieve mathematical insight by independently building a representative mathematical model. It rather provides assistance to a mathematician who tries to find and understand the model.

In the framework described by Davies et al., a mathematician suggests the existence of a relation between some entities $X(z)$ and $Y(z)$, which we may call a manifestation of mathematical intuition. To proceed to the clear formulation and the proof of the intuited relation, a mathematician needs to obtain a clear understanding of the function f' connecting the two entities. That ‘understanding’ can be described as mathematical insight. We should note that the defined f' is not a mathematical model itself: it defines the relation between $X(z)$ and $Y(z)$, but it does not represent it. It is possible that a mathematician who is presented with the formula would understand how the formula works but not capture the essential structure of the relation. That case can be illustrated by the characteristic of traceability provided by Nathan and Walkington (which we discussed it in section 2.5). A mathematician may be able to understand how a given formula works, but not be able to prove that it does¹. In such cases, the mathematician would not have obtained mathematical insight: they did not grasp any structural knowledge about the entity they studied. Hence, the final conjecture f' is not an example of a model that initiates mathematical insight. However, the clear formulation of f' is one of the outcomes of understanding of a mathematical model (e.i. obtaining mathematical insight). If an agent achieved mathematical insight about the relation $\hat{f}(X(z)) \approx Y(z)$, they would understand how the function defining the intuited relation works, and, hence, would be able to formulate it.

AI points out the patterns that may help to define f' and prove it. Therefore, we can describe the role of AI as assisting in constructing a model based on which a mathematician aims to achieve mathematical insight. In this approach to the construction of a mathematical model, the model is built purposefully – which is not always the case for mathematical insight. Based on the description of mathematical models given by MacDonald and Griffiths, models appear as a result of a deep understanding of the studied topic(s). To obtain mathematical insight, it is necessary to understand a model of the studied theory. However, it is not always the case that an individual researches a theory and tries to construct its model with the goal of obtaining a novel mathematical insight. As we have seen from Griffiths’ examples of models presented in section 2.4, they can serve as a ground for obtaining mathematical insight without providing any information that would help a mathematician obtain a novel result. Sometimes, it is the case that an individual studies a theory to gain ideas on the proof methods they could use and sometimes, it is the case that an individual studies a theory to merely understand it, not to prove some statement related to the theory. In the case of Davies et al., a mathematician wants to understand the model of the intuited relation to understand how to prove the said relation. Hence, constructing a model is a purposeful process. In that case, we may even describe the role AI takes as ‘guiding a mathematician in obtaining an

¹Here, I talk about the proof that a given formula is correct. However, the nature of the verification that demonstrates that an agent obtained an insight can include an explanation of why there is a relation between X and Y , why these variables were used to define this relation and not the other, how exactly changes in X affect changes in Y , etc. Following Nathan and Walkington, we may call it a traceable solution to the problem.

understanding of a model’, or, by extending to the purpose of this process, ‘guiding a mathematician in obtaining mathematical insight’. However, I must point out that the word ‘guide’ is used here in a very weak sense: AI does not make any decisions about the final formulation of f' or developing the proof strategy. E. Davis gives a good explanation of why it would be hard to use AI to *guide* mathematical intuition, or what we call mathematical intuition: “No teacher ever tries to guide their students’ intuitions by telling them, “I have looked at millions of examples, and I can report that there is a pattern that works X% of the time, mostly on the basis of features A, B, C” which is what the Deep Learning tells the mathematician, and all that the DL tells the mathematician” (Davis 2021, 6). AI functions as a tool that suggests the entities and patterns to study, not as a tool that is choosing the approach to the proof. AI guides a mathematician in obtaining insight strictly in the sense that it suggests the patterns that may be worth looking into – it is entirely up to the mathematician whether to use them or not, and how. In the flowchart described by Davies et al., a mathematician may change the initial dataset and formulations of $\hat{f}$ several times to achieve the final formulation of f' : that is exactly what happened in the proofs the authors present.

We have established that AI which is claimed by Davies et al. to guide mathematical intuition does not, in fact, provide any help in obtaining, guiding, or assisting a mathematician’s intuition. It provides assistance in achieving mathematical insights. Specifically, it helps a mathematician understand the structure and proof strategy for the intuited relation. It is important to distinguish mathematical intuition and mathematical insight: the former is a process responsible for bringing a certain mathematical entity to attention, while the latter is a state of understanding of the structure behind the entity (its precise definition, connection to other entities, etc). Mathematical intuition is an experience, similar to perception, of a mathematical object. Parsons and Griffiths tried to conceptualize it but never discussed the possible origins of intuition. The authors stated the connection of mathematical intuition to the practical experience of a mathematician, but they did not discuss how exactly the intuition emerges – what should an individual do to obtain it?. It would be hard to formulate what exactly one should do other than ‘a lot of mathematics’ to obtain a mathematical intuition about any object. Since the object of intuition should not be strictly defined but rather experienced, there is no formal way to access the object rather than, indeed, experience it directly, alike with sensory perception. If AI would be able to help a mathematician experience the object with intuition, it would mean that AI is able to intervene in a mathematician’s mental experience: the experience of mathematical objects that are not yet formalized or justified. In this case, the problem would lie not only in the fact that the AI algorithms that are supposedly helping a mathematician’s intuition are incomprehensible: AI would be involved in a creative process that should be performed on either absent or vaguely specified data. It would be problematic for AI to assist a process the goal of which is to experience an abstract object. I would suggest that even a strong AI would not be able to assist a mathematician in their sensory-like experience, as it is not possible ‘assist’ in the process

of obtaining qualia². Experience of an object with mathematical intuition is, as discussed in section 2.3, an internal experience, the result of which cannot be formally described. Mathematical insight is also a subjective mental experience. However, the result of mathematical insight can be presented formally. As Nathan & Walkington described, mathematical insight always has a formal verification of the solution as its result. Since the result of mathematical insight is usually a mathematical proof, it is clear how AI can help to obtain it without interfering with human mental states: it suggests the patterns and variables that may affect the relation a mathematician chooses to explore. A mathematician may choose to use them for the construction of the model or not – it would be their own decision³. Unlike in the case of mathematical intuition, AI helps a mathematician not to experience the insight directly, but to understand which patterns can be used in the construction of the model. When a mathematician understands a model, they obtain insight. To conclude, we have established that AI, as presented in Davies et al., is not involved in the initial creative process of suggesting there exists some sort of mathematical relation between two entities. The function AI provides is assistance to a mathematician in constructing a model and, consequently, assistance in obtaining mathematical insight. To sum up, AI is not a creative agent. Why is it important?

Since it lacks the creative agency, mathematical AI cannot be regarded as an independent alternative to a human mathematician. Even if Davies et al. do not suggest it, some scientists may see it as a possible line of development (Wang et al., 2019). One of the focuses of the discussion about the usage of AI in science is the level of exploratory creativity of AI – creativity in searching and formulating new hypotheses, finding connections between previously unrelated domains, and, generally, the independent production of new scientific discoveries (Lawrynowicz 2020, 70-71). There exist several projects where AI was used to obtain new discoveries in science. For instance, King et al. (2009) presented AI ‘Adam’ that was able to autonomously generate functional hypotheses about the behavior of the yeast *Saccharomyces cerevisiae* and test the formulated hypothesis⁴. Liu et al. (2023) proposed an algorithm called Sparse Invariant Detector (SID) that is able to auto-discover conservation laws in physics from differential equations. SID was able to re-discover 14 of the already-known conservation laws as well as 3 new laws that were previously unknown to the domain experts. Jain et al. (2023) presented an AI-based system ‘GFlowNets’ that assists in formulating and verifying scientific hypotheses, specifically in organic chemistry. Adam, SID, and GFlowNets are examples of projects that research the exploratory creativity of AI. However, despite there being several examples of AI used in scientific discoveries, it is too early to talk

²As qualia is a unique subjective experience, it is impossible for an individual to get help in *experiencing* it. It is only possible to process a subjective experience independently.

³It is possible for a mathematician to make a wrong decision. That would lead to their inability to construct a relevant model, and, therefore, obtain a mathematical insight.

⁴As King et al. state themselves, the autonomy of their AI may be a subject of discussion (King et al. 2009, 88). However, we are not going to argue for or against this wording: we just use the authors’ project as an example of existing discussion about the exploratory creativity of AI in science.

about an AI that can function as a fully autonomous researcher: independently choose the theory and data they would be working with, independently collect the required data, formulate a hypothesis based on the studied data and/or experiments, and self-reflect on the quality of the finished work (Amabile 2019, 5). Even the newest achievements in AI-driven scientific discovery, for example, SID and GFlowNets, function as tools, not autonomous researchers. Liu et al., the researchers who built SID, confirm this balance of power: “Note that SID does not replace us human scientists, but rather acts as a helpful assistant: while humans need to input basis functions (i.e., *formulating* hypotheses) to SID, SID is good at computing conserved quantities (i.e., *testing* hypotheses) based on the given prompt” (Liu et al. 2023, 1). The authors of GFlowNets also call their creation a ‘valuable tool’, avoiding describing their AI as an autonomous agent. It is a human scientist (chemist, biologist, physicist, etc.) who suggests researching a specific topic, theory, or phenomenon. After the object or research is suggested, AI can suggest the research methods and directions, but the initial creative idea was produced by a scientist. The approach of Davies et al. seems to stand out among other authors using AI in scientific research: they state, right in the name of the article, that they are using AI to guide a mathematician’s intuition – the mental state, or process, that is presumably responsible for the creativity and novelty of the scientific discovery. If AI would be indeed used by Davies et al. to guide the emergence of ideas that come before the hypothesis formulation or verification strategy, it would be inaccurate to still regard this AI as merely a tool. However, if AI does not assist or guide intuition but rather assists a mathematician in constructing a model of the phenomenon they decided to research, its role can still be regarded as purely instrumental. By concluding that AI assists a mathematician in obtaining insight and not by guiding their intuition, we confirm the instrumental role of AI in mathematics.

Here is one line of critique the reader may have in mind: so, Davies et al. claimed to guide mathematical intuition, but this thesis shows that the AI they use only helps with mathematical insight. And, as it was established, the ‘ordinary’ understanding of mathematical intuition is closer to what is called mathematical insight in philosophy of mathematics. Hence, Davies et al. were right, they just did not use the academic definition of mathematical intuition, as they are not required to: they are software engineers, not philosophers. To that argument, I have to answer the following. First of all, I believe, the authors indeed used the ordinary notion of mathematical intuition. They did not provide any references to the definition of the notion of intuition they were using, as well as they did not explain how exactly a mathematician’s intuition and AI are interacting with each other. However, Davies et al. still claimed to do a very innovative thing: their article prescribes a high level of exploratory agency to AI, and, to explain this step, they use the notion of mathematical intuition. So, if we want to understand whether it is really the case that AI can be implemented into scientific discovery in the status of a ‘guide’, we need to explore the concept of mathematical intuition, and how exactly AI interacts with it. Second, the authors do not only claim that AI assists a mathematician’s intuition, they claim that AI guides it, which is also a really strong assumption. As we have

concluded in this chapter, the assistance of AI consists of providing a mathematician with the suggestion of patterns that are worth pursuing. That type of help can hardly be read as an example of guidance. It is clear that the more practice and information a mathematician would have, the better they would be at performing their intuition and insightfulness. However, this experience cannot be replaced by a consult with AI predictions, as mathematical practice trains one to perform hard mental work, and the intuition and insightfulness one experiences are based on this mental work. As we discussed earlier in this section, a consultation from a teacher who states that they looked through a big set of examples would not serve as guidance to us when we try to solve a particular problem, even though it may be of help. We may conclude that Davies et al. do not simply make minor formulation mistakes in their article by using the word ‘intuition’ in its day-to-day sense. The authors incorrectly describe AI assistant as a tool that is able to guide scientific exploration in mathematics, while, in fact, they have used AI to help a mathematician to understand an already chosen mathematical relation.

Nevertheless, Davies et al.’s proof provides a valuable counterexample to the critique of insightfulness of computer-based proofs described in section 1.2. Even though, as we concluded, AI itself is not a creative agent, it provides help in such a way that leaves a place for a mathematician to produce insightful proofs. As we concluded in section 3.2, AI provides assistance in the formulation of the conjecture that is to be proved, and, consequently, assistance in constructing the proof strategy. AI does not perform calculations necessary for the proof to be concluded, unlike classical computer assistants. Hence, even if Davies et al.’s proof is not as insightful as a proof could be (that is up for the professional mathematicians to decide), it still stands as an example that a computer assistant can be used in a proof and not jeopardize its novelty, insightfulness, and originality. Moreover, since AI is used to assist with the proof strategy, it may even be one of the indirect reasons that the resulting proof presents an insightful new method of dealing with a certain problem. However, the person responsible for the creative content of the proof would always be the mathematician themselves.

To conclude, in this Chapter, we have established that the role of AI in the proofs of Davies et al. should not be described as guiding mathematical intuition, but rather as assisting in constructing a model based on which a mathematician can obtain an insight. That means that Davies et al.’s AI is not involved in guiding a mathematician’s state of awareness about an abstract mathematical relation, but it is involved in helping a mathematician to provide a robust ground for this relation.

Conclusion

My goal for this thesis was to establish whether we can apply the notions of mathematical intuition and mathematical insight to describe the work of AI in mathematical proofs. In the first chapter, we describe how computer assistants were used in mathematical proofs before the implementation of AI-based assistants into mathematical practice, and how the implementation of classic computer assistants was argued against. As we established, one point of the critique was focused on the lack of insightfulness coming from the resulting proofs: no new methods or patterns were uncovered, and hence, no insights were brought into mathematical practice.

In the second chapter, we have described different approaches to mathematical intuition and mathematical insight. We concluded that the perceptual approach to mathematical intuition is the most conclusive and strongly established. Therefore, we based the account of mathematical intuition in philosophy of mathematics on the perceptual approach described by Parsons and Chudnoff. For the same reasons, we chose to ground the discussion of mathematical insight on the accounts provided by Griffiths and McDonald. The authors base the notion of insight on the agent's understanding of a representative model describing the related theory. We have also noticed some strong points presented by other accounts of mathematical insight: the description of insightfulness as a property (Carter) and the traceability of knowledge obtained with mathematical insight (Nathan and Walkington).

In the third chapter, we discussed how the method of obtaining AI-assisted proofs presented by Davies et al. is different from the classical computer-assisted proofs described in Chapter 1. We concluded that the software is used differently in the two types of proofs, as well as the initial goals of the mathematicians implementing computer assistants into their work are different in their nature. In the case of classic computer proofs, the software is used to verify a certain step in a pre-planned proof; in the case of the AI assistant, the software is used to plan the mathematical statement and the proof strategy in the first place. This implementation of software into mathematical proofs can be used to argue against the critique of computer assistants focused on the lack of insightfulness and novelty coming from computer-assisted proofs. Davies et al. used AI specifically to unravel the patterns that would be otherwise unlikely to be detected by a human mathematician. Hence, the authors presented a case of a computer-assisted mathematical proof the novelty of which – the original approach to the

proof and formulation of the conjecture – was obtained with the help of AI. This example illustrates that it is possible to not merely obtain a non-trivial mathematical result by using a computer assistant, but also that a computer can assist in finding and structuring this novelty.

In Chapter 4, we established that despite Davies et al. stating that in their proof, AI guides mathematical intuition, that is not the function it performs. It is not possible for AI to assist a mathematician in intuiting an object: AI cannot provide help in the process of obtaining this experience. It is also not the case that AI can be attributed any kind of creative agency in mathematical proofs, at least, not to the AI that was used in Davies et al.’s proofs. AI does not provide help to a mathematician in choosing what mathematical relation they are going to research. It, however, provides assistance on the stage where an already chosen relation should be formulated into a precise conjecture, with a clear indication of the patterns that a mathematician is going to pursue in the proof. Hence, the role of AI in proofs can be described as providing assistance in understanding the structure and proof strategy of the object one has intuited. This kind of understanding is described by several authors (MacDonald, Griffiths, Carter) as mathematical insight. Hence, AI assists not in guiding or obtaining mathematical intuition, but in providing a ground for mathematical insight. That role, even though it provides essential help in structuring the conjecture one is going to prove, does not make AI responsible for the decision of *what* one is going to prove. AI provides assistance during the decision of *how* one is going to prove the chosen relation, but a human mathematician is still the only agent responsible for the final formulation and strategy: as we have seen in Chapter 3, a mathematician can reintroduce the input data and reformulate the candidate for the final conjecture several times until they are satisfied with it, and in this process, the agency of the mathematician plays the final role.

We have come to the conclusion that in Davis et al.’s example, AI is used as a tool and cannot be attributed exploratory agency directly. However, the method of implementation of AI into mathematical practice demonstrated by the authors is significantly different from the previously used computer assistants. That method shows that in mathematics, computer software can be used not merely as a tool assisting at a step of the already pre-planned proof, but also at the step preceding the proof, involving the planning and formulation of the conjecture one is going to prove. This example illustrates that ‘insightful proof’ – proof revealing novel methods and approaches – can be obtained with the use of a computer assistant. Moreover, these new approaches can be planned with the help of a computer assistant. However, a phenomenon that can be regarded as a starting point of many proofs – mathematical intuition – cannot be assisted by a computer, since it lacks any structural content and only amounts to the mathematician’s experience of an object, close to perception.

Currently, machine learning⁵ is widely used in mathematical practice (d’Ascoli 2022, Zhang 2022, Berglund 2021). It is possible that the framework established

⁵Some authors call the assistants they use ‘AI’, and some only refer to it as certain machine learning techniques.

by Davies et al. will be developed into a more sufficient system that requires fewer cycles combining machine learning predictions and human supervision and correction. However, even if the authors achieve that, the role they leave to AI would not involve any direct interaction with mathematical intuition other than, of course, the development of the idea suggested by it. The method Davies et al. suggested is very innovative for mathematical practice, but it still does not allow for AI to guide the inner creative experience constituting mathematical intuition.

Bibliography

- [1] Abrahamson, D., & Trninic, D. (2015). Bringing forth mathematical concepts: signifying sensorimotor enactment in fields of promoted action. *ZDM Mathematics Education*, 47(2), 1–12.
- [2] Abrahamson, D. (2015). The monster in the machine, or why educational technology needs embodied design. In V. R. Lee (Ed.), *Learning technologies and the body: integration and implementation in formal and informal learning environments* (pp. 21–38). New York: Routledge.
- [3] Adamson, P. (2012). *Stanford Encyclopedia of Philosophy*.
- [4] de Almeida, A. F., Moreira, R., & Rodrigues, T. (2019). Synthetic organic chemistry driven by artificial intelligence. *Nature Reviews Chemistry*, 3(10), 589-604.
- [5] Amabile, T. M. (2020). Creativity, artificial intelligence, and a world of surprises. *Academy of Management Discoveries*, 6(3), 351-354.
- [6] Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. *AI & society*, 1-11.
- [7] Anscombe, G. E. M. (1957). *Intention*. Reprinted, Cambridge, MA: Harvard University Press, 2000.
- [8] Appel, K. I. and Haken W. (1977). “The solution of the four-color-map problem.” *Scientific American*, 237(4), 108-121.
- [9] Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: a model of artificial intelligence development. *AI & society*, 31, 201-206.
- [10] d’Ascoli, S., Kamienny, P. A., Lample, G., & Charton, F. (2022). Deep symbolic regression for recurrence prediction. In *International Conference on Machine Learning* (pp. 4520-4536). PMLR.
- [11] Arwanto, A., Budayasa, I. K., & Budiarto, M. T. (2018). The Exploration of Mathematical Intuition and Its Role in Solving Mathematical Problem. In *Mathematics, Informatics, Science, and Education International Conference (MISEIC 2018)* (pp. 111-113). Atlantis Press.
- [12] Azad, K. (2013). *Math Better Explained: Learn to Unlock Your Math Intuition*. Amazon Fulfillment.
- [13] Baiduri, Y. M. C., & Ulfah, F. (2022). The Intuitive Thinking Process of High Ability Students in Mathematical Problem Solving Based on Gender. *Journal of Hunan University Natural Sciences*, 49(2).
- [14] Baker, A. (2008). Experimental mathematics. *Erkenntnis*, 68, 331-344.
- [15] Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- [16] Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5), 367–386.
- [17] Battaly, H. (2008). Virtue epistemology. *Philosophy Compass*, 3(4), 639-663.

- [18] Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change*, 78(1), 185-195.
- [19] Beck, L. J. (1954). *The method of Descartes*. Oxford: The Clarendon Press, 1952.
- [20] Benacerraf, P., Putnam, H., Putnam, H. W., Hilary, P., & Benacerraf, H. (Eds.). (1983). *Philosophy of mathematics: Selected readings*. Cambridge University Press.
- [21] Berglund, P., Campbell, B., & Jejjala, V. (2021). Machine Learning Kreuzer–Skarke Calabi–Yau Threefolds. arXiv preprint arXiv:2112.09117.
- [22] Birkhoff, G. D. (1913). The reducibility of maps. *American Journal of Mathematics*, 35(2), 115-128.
- [23] Blundell C., Buesing L., Davies A., Veličković P., and Williamson G. (2022) Towards combinatorial invariance for Kazhdan-Lusztig polynomials. *Representation Theory of the American Mathematical Society* 26, no. 37: 1145-1191.
- [24] Bowers, K. S., Regehr, G., Balthazard, C., and Parker, K. (1990). Intuition in the context of discovery. *Cogn. Psychol.* 22, 72–110. doi: 10.1016/0010-0285(90)90004-N Bratman, M. (1987) *Intention, Plans, and Practical reason*. Cambridge, MA: Harvard University Press.
- [25] Brenti, F. Kazhdan-Lusztig polynomials: history, problems, and combinatorial invariance. *Sémin. Lothar. Combin.* 49, B49b (2002).
- [26] Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. Preprint at <https://arxiv.org/abs/2104.13478> (2021).
- [27] Brown, J. R. (2010). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures*. Routledge.
- [28] Carter, J. A. (2017). Virtuous insightfulness. *Episteme*, 14(4), 539-554.
- [29] Caton, H. (1975). Will and Reason in Descartes’s Theory of Error. *The Journal of Philosophy*, 72(4), 87-104.
- [30] Charlton, B. G. (2009). The vital role of transcendental truth in science. *Medical hypotheses*, 72(4), 373-376.
- [31] Cheng, M. (2022). The Creativity of Artificial Intelligence in Art. In *Proceedings* (Vol. 81, No. 1, p. 110). MDPI.
- [32] Chudnoff, E. (2014). *Intuition in mathematics*.
- [33] Chumley, L. H., & Harkness, N. (2013). Introduction: qualia. *Anthropological Theory*, 13(1-2), 3-11.
- [34] Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23), 685-700.
- [35] Davis, E. (2021). Deep Learning and Mathematical Intuition: A Review of (Davies et al. 2021). arXiv preprint arXiv:2112.04324.
- [36] Donovan, A., Boncoddio, R., Williams, C. C., Walkington, C., Pier, E. L., Waala, J., et al. (2014). Action, gesture and abstraction in mathematical learning. San Diego, CA: Thematic Panel presented at the Sixth Conference of the International Society for Gesture Studies.
- [37] Dyer, M. (1987) *Hecke Algebras and Reflection in Coxeter Groups*. (2nd ed.), Ph.D. Thesis, University of Sydney.
- [38] Emirbayer, M., & Mische, A. (1998). What is agency?. *American journal of sociology*, 103(4), 962-1023.
- [39] Ewald, W. B. (1996) *From Kant To Hilbert: A Source Book in the Foundations of Mathematics*. Oxford University Press.
- [40] Ezenkwu, C. P., & Starkey, A. (2019). Machine autonomy: Definition, approaches, challenges and research gaps. In *Intelligent Computing-Proceedings of the Computing Conference* (pp. 335–358). Springer, Cham.

- [41] Feferman, S. (2000). Mathematical intuition vs. mathematical monsters. *Synthese*, 125, 317-332.
- [42] Ferrari, P. L. (2003). Abstraction in mathematics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1435), 1225-1230.
- [43] Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379.
- [44] Flowers, J. C. (2019). Strong and Weak AI: Deweyan Considerations. In *AAAI spring symposium: Towards conscious AI systems* (Vol. 2287, No. 7).
- [45] Foster, J. (2000) *The Nature of Perception*. Oxford University Press Oxford.
- [46] Gill, K. S. (2020). AI & Society: editorial volume 35.2: the trappings of AI Agency. *AI & SOCIETY*, 35, 289-296.
- [47] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2020). Message passing neural networks. *Machine learning meets quantum physics*, 199-214.
- [48] Glasscock P., Juan S. & Tenenbaum S. (2023) Action. *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.)
- [49] Glouberman, M. (2011). Descartes, scientia and pure enquiry. *British journal for the history of philosophy*, 19(5), 873-886.
- [50] Gödel, K. (1947). What is Cantor's continuum problem? *The American Mathematical Monthly*, 54(9), 515-525.
- [51] Greco, J. (1999) 'Agent Reliabilism'. *Noûs*, 33: 273-96.
- [52] Greco, J. (2012) 'A (Different) Virtue Epistemology.' *Philosophy and Phenomenological Research*, 85: 1-26.
- [53] Griffiths, H. B. (1971). Mathematical insight and mathematical curricula. *Educational Studies in Mathematics*, 153-165.
- [54] He, Y. H. (2021). Machine-learning mathematical structures. *International Journal of Data Science in the Mathematical Sciences*, 1(01), 23-47.
- Hege, H. C., & Polthier, K. (1997). *Visualization and mathematics: Experiments, simulations and environnements*. Springer Science & Business Media.
- [55] Huet, G., Kahn, G., & Paulin-Mohring, C. (1997). The coq proof assistant a tutorial. *Rapport Technique*, 178.
- [56] Jain, M., Deleu, T., Hartford, J., Liu, C. H., Hernandez-Garcia, A., & Bengio, Y. (2023). GFlowNets for AI-driven scientific discovery. *Digital Discovery*.
- [57] Jankowiak, T. (2014). Sensations as representations in Kant. *British Journal for the History of Philosophy*, 22(3), 492-513.
- [58] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- [59] Kazhdan, D., Lusztig, G. (1979) Representations of Coxeter groups and Hecke algebras. *Invent Math* 53, 165-184.
- [60] Kitcher, P. (1976). Hilbert's epistemology. *Philosophy of Science*, 43(1), 99-115.
- [61] Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1534-1555. doi: 10.1037/0278-7393.25.6.1534
- [62] Kockelman, P. (2013). *Agent, person, subject, self: A theory of ontology, interaction, and infrastructure*. Oxford University Press.
- [63] Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12), 761-769.

- [64] Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., dos Passos Gomes G., Häse F., Jinich A., Nigam A., Yao Z. & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12), 761-769.
- [65] L. E. J. Brouwer (1948) *Collected Works*, vol. I, ed. A. Heyting, North Holland, Amsterdam, 1975.
- [66] Lam, C. W. H. (1990). How reliable is a computer-based proof. *Mathematical Intelligencer* 12.1: 8-12.
- [67] Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics. arXiv preprint arXiv:1912.01412.
- [68] Lample, G. & Charton, F. Deep learning for symbolic mathematics. Preprint at <https://arxiv.org/abs/1912.01412> (2019).
- [69] Lawrynowicz, A. (2020). Creative AI: A new avenue for the Semantic Web?. *Semantic Web*, 11(1), 69-78.
- [70] Legaspi, R., He, Z., & Toyoizumi, T. (2019). Synthetic agency: sense of agency in artificial intelligence. *Current Opinion in Behavioral Sciences*, 29, 84-90.
- [71] Leung, A. K. Y., Kim, S., Polman, E., Ong, L. S., Qiu, L., Goncalo, J. A., & Sanchez-Burks, J. (2012). Embodied metaphors and creative “acts”. *Psychological science*, 23(5), 502-509.
- [72] Lillicrap, T. P., & Kording, K. P. (2019). What does it mean to understand a neural network?. arXiv preprint arXiv:1907.06374.
- [73] List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.
- [74] Lohmar, D. (2002). Husserl’s concept of categorial intuition. *One hundred years of phenomenology: Husserl’s logical investigations revisited*, 125-145.
- [75] Lorenz, K. Z. (1974). Analogy as a source of knowledge. *Science*, 185(4147), 229-234.
- [76] Macdonald, I. D. (1978). Insight and intuition in mathematics. *Educational Studies in Mathematics*, 9, 411-420.
- [77] MacKenzie, D. (2004). *Mechanizing proof: computing, risk, and trust*. MIT Press.
- [78] Maddy, P. (1980). Perception and mathematical intuition. *The Philosophical Review*, 89(2), 163-196.
- [79] Maetschke, S., Iraola, D. M., Barnard, P., ShafieiBavani, E., Zhong, P., Xu, Y., & Yepes, A. J. (2021). Understanding in artificial intelligence. arXiv preprint arXiv:2101.06573.
- [80] Mainali, B. (2021). Representation in Teaching and Learning Mathematics. *International Journal of Education in Mathematics, Science and Technology*, 9(1), 1-21.
- [81] Marghetis, T., Edwards, L. D., & Núñez, R. (2014). More than mere handwaving: gesture and embodiment in expert mathematical proof. In L. D. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives on gesture and embodiment in mathematics* (pp. 227–246). Charlotte, NC: Information Age Publishing.
- [82] Marghetis, T., Edwards, L. D., & Núñez, R. (2014). More than mere handwaving: gesture and embodiment in expert mathematical proof. In L. D. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives on gesture and embodiment in mathematics* (pp. 227–246). Charlotte, NC: Information Age Publishing.
- [83] Mathiesen, K. (2011). Can groups be epistemic agents?. *Collective epistemology*, 23-44.
- [84] Melnikov, A. A., Poulsen Nautrup, H., Krenn, M., Dunjko, V., Tiersch, M., Zeilinger, A., & Briegel, H. J. (2018). Active learning machine learns to create new quantum experiments. *Proceedings of the National Academy of Sciences*, 115(6), 1221-1226.
- [85] Mitchelmore, M. C., & White, P. (1995). Abstraction in mathematics: Conflict, resolution and application. *Mathematics Education Research Journal*, 7(1), 50-68.

- [86] Montano, U. (2013). Explaining beauty in mathematics: An aesthetic theory of mathematics (Vol. 370). Springer Science & Business Media.
- [87] Montmarquet, J. (1993). Epistemic Virtue and Doxastic Responsibility. Totowa, NJ: Rowman & Littlefield.
- [88] Moreno, A., & Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life*, 11(1-2), 161-175.
- [89] Nathan, M. J., & Martinez, C. V. (2015). Gesture as model enactment: the role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice*, 1(1), 4–37.
- [90] Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive research: principles and implications*, 2, 1-20.
- [91] Neville-Neil, G. (2016). The Chess Player who Couldn't Pass the Salt: AI: Soft and hard, weak and strong, narrow and general. *Queue*, 14(6), 14-18.
- [92] Neild, D. (2021). AI is discovering patterns in pure mathematics that have never been seen before. *Science Alert*.
- [93] Ng, G. W., & Leung, W. C. (2020). Strong artificial intelligence and consciousness. *Journal of Artificial Intelligence and Consciousness*, 7(01), 63-72.
- [94] Núñez, R. (2006). Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. 18 *Unconventional essays on the nature of mathematics*, 160-181.
- [95] Palais, R. S. (1999). The visualization of mathematics: Towards a mathematical exploratorium. *Notices of the AMS*, 46(6), 647-658.
- [96] Parshina K. (2023). "Philosophical Assumptions Behind the Rejection of Computer-Based Proofs" *KRITERION – Journal of Philosophy*.
- [97] Parsons, C. (1979). Mathematical intuition. In *Proceedings of the Aristotelian Society* (Vol. 80, pp. 145-168). Aristotelian Society, Wiley.
- [98] Parsons, C. (1992). Kant's philosophy of arithmetic. *Kant's Philosophy of Mathematics: Modern Essays*, 43-79.
- [99] Peifer, D., Stillman, M., & Halpern-Leistner, D. (2020). Learning selection strategies in Buchberger's algorithm. In *International Conference on Machine Learning* (pp. 7575-7585). PMLR.
- [100] Petrick, C., & Martin, T. (2012). Mind your body: learning mathematics through physical action. Vancouver, Canada: Paper presented at the Annual meeting of the American Educational Research Association.
- [101] Popa, E. (2021). Human goals are constitutive of agency in artificial intelligence (AI). *Philosophy & Technology*, 34(4), 1731-1750.
- [102] Reber, R., Ruch-Monachon, M. A., and Perrig, W. J. (2007). Decomposing intuitive components in a conceptual problem solving task. *Conscious. Cogn.* 16, 294–309. doi: 10.1016/j.concog.2006.05.004
- [103] Rota, G. C. (1997). The phenomenology of mathematical beauty. *Synthese*, 111, 171-182. Russell, S. J., Norvig, P. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.
- [104] Setiya, K. (2011) Reasons and Causes. *European Journal for Philosophy of Science*, 19 (1): 129–157.
- [105] Slezak, P. (2010). Doubts About Decartes' Indubitability: The Cogito as Intuition and Inference. In *Philosophical Forum* (Vol. 41, No. 4, p. 389).
- [106] Sosa, E. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.

- [107] Sosa, E. (2011). *Reflective Knowledge: Apt Belief and Reflective Knowledge* (Vol. 2). Oxford: Oxford University Press.
- [108] Sparkes, M. (2021) DeepMind AI collaborates with humans on two mathematical breakthroughs. *New Scientist*.
- [109] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319-3328). PMLR.
- [110] Tao, T. There's more to mathematics than rigour and proofs. Blog post: <https://terrytao.wordpress.com/career-advice/theres-more-to-mathematics-than-rigour-and-proofs/> (2016). Date of reference: 15.06.2023.
- [111] Thurston, William P. (1994). On proof and progress in mathematics. *American Mathematical Society*, 30(2), 161-177.
- [112] Tieszen, R. L. (1989). *Mathematical intuition* (pp. 66-91). Springer Netherlands.
- [113] Tollefsen, D. P. (2015). *Groups as agents*. John Wiley & Sons.
- [114] Tymoczko, Thomas (1979). The four-color problem and its philosophical significance. *The Journal of Philosophy*, 76(2), 57-83.
- [115] Van de Pitte, F. P. (1988). Intuition and judgment in Descartes' theory of truth. *Journal of the History of Philosophy*, 26(3), 453-470.
- [116] Volz, K. G., and Von Cramon, D. Y. (2006). What neuroscience can tell about intuitive processes in the context of perceptual discovery. *J. Cogn. Neurosci.* 18, 2077–2087. doi: 10.1162/jocn.2006.18.12.2077
- [117] Wagner, A. Z. (2021). Constructions in combinatorics via neural networks. arXiv preprint arXiv:2104.14516.
- [118] Wagner, A. Z.(2021) Constructions in combinatorics via neural networks. Preprint at <https://arxiv.org/abs/2104.14516>.
- [119] Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., & Gray, A. (2019). Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on human-computer interaction*, 3(CSCW), 1-24.
- [120] Webb, S. (2018). Deep learning for biology. *Nature*, 554(7693), 555-557.
- [121] Whitney, H., & Tutte, W. T. (1992). Kempe chains and the four colour problem. *Hassler Whitney Collected Papers*, 185-225.
- [122] Wilson, R. (2003). *Four Colors Suffice: How the Map Problem Was Solved*. Princeton and Oxford: Princeton university press.
- [123] Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, 100224.
- [124] Zhang, X., Huang, L., Zhao, R., Zhou, H., Li, X., Geng, G., & Zhang, S. (2022). Basis function approach for diffractive pattern generation with Dammann vortex metasurfaces. *Science Advances*, 8(40), eabp8073.
- [125] Zhang, Z., Lei, Y., & Li, H. (2016). Approaching the distinction between intuition and insight. *Frontiers in psychology*, 7, 1195.