Received: July 30, 2022 / Accepted: October 22, 2022 / Published online: Dec. 12, 2022

ORIGINAL SCIENTIFIC PAPER

UDC: 1 Wittgenstein, L
1:004.8

Wittgenstein and LaMDA

Karlo Gardavski[1]

Abstract:

This paper is based on Ludwig Wittgenstein's (late) teaching on language and meaning, and its aim is to show how we can avoid the amorphization of artificial intelligence or interpreting the work (the question of giving meaning) of AI as similar to or the same as the work of a human being. The way of determining the meaning of certain linguistic units performed by an AI and a human differs because the languages they operate with have a different set of rules or criteria that are indicators of what a certain linguistic entity means. The core of AI in terms of meaning is its logical base, according to which it operates/calculates/manipulates the given information. On the other hand, a human being finds his/her criteria in the language activity performed in the language communities.

Key words: rules, meaning, machine, language, AI

Introduction

[1] K. Gardavski
University of Sarajevo, Faculty of Philosophy, Dept. of Philosophy
Scientific and Reasearch Incubator
Franje Račkog 1, 71000 Sarajevo, Bosnia and Herzegovina
karlo.gardavski@gmail.com

The development of technology carried with it a dose of anthropomorphizing technology, which is also the case today. Namely, the new Google AI called LaMDA raised a lot of public speculation about whether AI has the same functions as a human, in this case, it is similar to the question of whether AI can have feelings.[1] LaMDA is Google's most recent developed chatbot or program specially designed to simulate a conversation with people.[2] The reason was that an engineer who worked for Google provided a lot of information about how the AI behaved during testing. During testing, the AI was able to answer questions about how it feels and that it was afraid. Within the paper, the goal is to argue, based on Wittgenstein's teaching on the problem of meaning in language, that the interpretation of AI as sentient arises from not understanding the function of language at the level of semantics. How we determine the meaning of linguistic entities in language differs radically in terms of the conditions for the correctness of the meaning. We will argue that the AI does not understand the meaning of its answers. Understanding the meaning of a linguistic entity is a product of language practice that can only be achieved by humans. AI operates with symbols is based on the rigid logical rules that it has at hand and gives answers in accordance with these rules. That is, how you set the AI to operate then it will give you such answers. The AI language is formal and predetermined. The communication that people make is unpredictable, contextually sensitive, contingent and has all its significance only in the mutual relationship between all members of a linguistic community. This paper will also point out that to have a feeling means to be part of the language practice in a language community. The ability to identify whether we have feeling X or Y depends on the language that a community uses to describe the phenomenon X or Y. Humans and AI differ widely in the kinds of rules they have at their disposal that determine the meaning of linguistic entities. If we accept Wittgenstein's argument on the problem of meaning, the same implication applies to phenomena such as feelings, they have their own meaning and understanding in communication. AI is not sentient because the concept of sentience has been misinterpreted due to a misunderstanding of the function of language (in terms of meaning).

The first part of the paper will briefly present Ludwig Wittgenstein's pragmatic understanding of the problem of meaning. The second part of the paper is devoted to the presentation of the operation of the computer machine. In this case, it is a Turing machine that will serve as a general representation of how computer

[1]See https://theconversation.com/is-googles-lamda-conscious-a-philosophers-view-184987 , https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/ , https://www.bbc.com/news/technology-61784011

[2]See https://blog.google/technology/ai/lamda/

machines (and AI) work. Within the third part of the work, the goal is to present two types of rules that are in the background of the language used by AI and humans. These two types of rules are radically different and show us that understanding the meaning of a linguistic entity is a product of language practice, not a calculation made by a machine. At the end of the paper, we will present three arguments that show that at the level of semantics, a machine (AI) is not able to understand the meaning it uttered through calculation.

Wittgenstein on meaning and rules

This part of the paper will briefly present the basics of learning about the problem of the meaning of the late Wittgenstein. This learning should be the basis for understanding how humans and AI differ in the way they create answers to questions, i.e. in operating with language. The general characteristic that Wittgenstein represents in his late phase when it comes to the meaning of different linguistic entities is that the meaning depends on the use in a specific linguistic context (social community). For Wittgenstein, the meaning of certain words in a language is a product of use. To use words in a certain way is to play a language game, and Wittgenstein says:

> „In the practice of the use of language (2) one party calls out the words, the other acts on them. In instruction in the language the following process will occur: the learner names the objects; that is, he utters the word when the teacher points to the stone.—And there will be this still simpler exercise: the pupil repeats the words after the teacher——both of these being processes resembling language. We can also think of the whole process of using words in (2) as one of those games by means of which children learn their native language. I will call these games "language-games" and will sometimes speak of a primitive language as a language-game. And the processes of naming the stones and of repeating words after someone might also be called language-games. Think of much of the use of words in games like ring-a-ring-a-roses. I shall also call the whole, consisting of language and the actions into which it is woven, the "language-game" “ (Wittgenstein, 1986, p. 5).

In other words, language games are an activity performed by members of a language community, that is, they communicate with each other and in the process come to determine what a word means. That is, we determine what the given word means according to its usage. Wittgenstein interprets the meaning as instrumental and pragmatic and explicitly claims: „Think of the tools in a tool-box: there is a hammer, pliers, a saw, a screw-driver, a rule, a glue-pot, glue, nails and screws. —

The functions of words are as diverse as the functions of these objects. (And in both cases there are similarities.)“ (Wittgenstein, 1986, p. 6).

How many meanings one word can have within different language games is unknown to us, because the meaning, if it is seen from its use, is not given in advance, but is learned through language practice, used further during the playing of the game, and altered if the need for change arises. That is, there are many language games or ways to determine what one linguistic entity means in communication with others. As Wittgenstein says:

> „But how many kinds of sentence are there? Say assertion, question, and command?—There are countless kinds: countless different kinds of use of what we call "symbols", "words", "sentences". And this multiplicity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten. (We can get a rough picture of this from the changes in mathematics.)“ (Wittgenstein, 1986, p. 11).

Here the term "language-game" is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life.

Every way in which a language game or a language practice is conducted depends on the practice itself or the way of playing language games. Determining meaning is not merely a theoretical matter for Wittgenstein, but an activity that cannot be avoided. That is, to speak a language means to have a form of life. We can find the reason for this in the way we lead our lives, as we actively use language. For Wittgenstein, there is no difference between physical action and linguistic action, one does not work without the other. It is not possible to act without speaking the language.

What is important to emphasize once again is that meaning is a product of practice. But how does understanding occur in practice? For Wittgenstein, the answer to that can be found in what is called a rule. A rule is a condition under which a certain word is used in a given practice. Wittgenstein says of the rule:

> „The rule may be an aid in teaching the game. The learner is told it and given practice in applying it.—Or it is an instrument of the game itself.—Or a rule is employed neither in the teaching nor in the game itself; nor is it set down in a list of rules. One learns the game by watching how others play. But we say that it is played according to such-and-such rules because an observer can read these rules off from the practice of the game—like a natural law

> governing the play.—But how does the observer distinguish in this case between players' mistakes and correct play?—There are characteristic signs of it in the players' behaviour. Think of the behaviour characteristic of correcting a slip of the tongue. It would be possible to recognize that someone was doing so even without knowing his language“ (Wittgenstein, 1986, p. 27).

Learning what a word means in a given practice means teaching someone the rules of how to use that word. Understanding the meaning of a given word means knowing the rule according to which that word was used. Rules are, if we can say it that way, indicative of what I think a given word means when we use it. This form of concept of rules that Wittgenstein offers us has its foundation in the language community. This is the concept of meaning and rules are terms that depend on the linguistic activity of a language community. The exhaustiveness of the use of the rule or the correct use of the meaning depends on all the players of that community, who, by their reactions to linguistic actions, judge whether the rule is used correctly. The rules are in the public realm, and thus the criteria for correct adherence to the rules are also public.

This setting of rules as an external criterion plays an important role for us in understanding what is related to Wittgenstein's critique of private language and will help us understand the difference between how meanings in language are manifested for a human and how for an AI.

Criticism of private language is the idea of deciphering the meaning of certain words, which is not possible without the presence of one choice of meaning, which in this case is one language community. To have a private language would be to have a source of private meanings that only one person understands and no one else. Thus, we would probably enter one semantic solipsism. With the concept of rules, we wanted to point out that the very idea of a private language is impossible, because the question that would be asked if meanings were private is: how can we understand each other? That is, how do we know what certain words mean, and what are our criteria for determining the meaning?

Wittgenstein says: „Let us remember that there are certain criteria in a man's behaviour for the fact that he does not understand a word: that it means nothing to him, that he can do nothing with it. And criteria for his 'thinking he understands', attaching some meaning to the word, but not the right one. And, lastly, criteria for his understanding the word right. In the second case one might speak of a subjective understanding. And sounds which no one else understands but which I

'appear to understand" might be called a "private language" (Wittgenstein, 1986, p. 94).

Understanding meaning means learning to use it. For example, when it comes to certain mental and emotional states, without a language practice we would not be able to identify them at all. Having an idea of what pain, happiness, suffering is means learning what they mean in a certain language. To know that someone is in pain means to grasp the rules that help us recognize a state that corresponds to that rule. To learn what pain means is also to learn a language.

Briefly summarizing the given chapter, the goal was to present that the focal place for the understanding of meaning in Wittgenstein's terms is the language game or language practice. Language games are activities that we perform in communication with other members of the language community. Within the language games, we learn the rules and indicators of how to act correctly in the environment of one language. Following a rule means following a language game or acting according to the rules of a community within which we have learned how to behave linguistically. The exact number of rules and language games does not exist, namely, the meanings can be multiple. What one word means will depend solely on the way a language game is performed among all its participants.

Turing machine

Within this chapter, the goal is to show how a machine or, above all, its software works, as well as how the machine or AI gives answers to the pieces assigned to it in the program. The example we will use to shed some light on this is Turing's machine, considering that Turing and John von Neumann (1945) as the "father" of the CPU, presented the basic architecture according to which the machine operates its calculations. Machines continue to perform operations on that basis, which is based on systems of rigid logical rules. Although the Turing machine is hypothetically hardware, it still has the basis for the operation of software, which is a logical/algorithmic unit.

First of all, let's introduce Turing's bow. It consists of three parts: store, executive unit and control (Turing, 2004, pp. 215 – 217). Store is used to store all given information that the machine has at its disposal. The executive unit performs symbol manipulation or calculation. Instructions are also written inside the Store, i.e. Turing calls it a "table of instructions", i.e. a table of rules used to manipulate symbols in order to perform a certain function. With Von Neumann, the table of instructions is called Logical unit, and that name is taken to be part of modern computer chip (1979). Control is the part responsible for checking whether the

machine followed the rules correctly. Thus, if we give certain information to the computer, i.e. input, the computer processes that information in the way its instruction system is written. And after that, it gives the output information or information that the computer processed through the instructions and gives the answer that was required from the rules that were given to it.

If we ask the question now in the context of the Turing machine, how certain language entities get their content, the answer will be and depend on the system of rigid rules that the computer has available to perform certain operations. If, for example, we want to know what the variable X means, we will only get know it if we see how and in what way the computer manipulates that variable with logical rules. In a certain way, it could be said that if we want to explain the meaning of a certain variable, we need to look at a series of rules that are designed to explain that variable. For example, we take the following function:

$(P \cdot Q) \rightarrow R$

If we want to get information about what the variable R means, then we have to have the input P and Q which are then put in a relationship with each other and a logical operator to explain R, which in this case is the output. The meaning of the variable R, as we see it, is envious of rigid rules that are placed in a certain order. This was an example of a single function. Today's computers operate with information that exceeds even our imagination, and have an enourmous amount of instructions that are placed there to manipulate symbols. But the setting remains the same for both Turing and Google's LaMDA. Having the meaning, or better say content, of a particular variable or some linguistic entity depends on a strictly ordered system of rules that produce results depending on how they are set to function within a particular program. When a computer gives us information, we simply look at the way his rule base has arranged the content.

If the setting of rigid rules is responsible for the question of the problem of meaning, what is the point of Turin's test? It reads like this:

> „The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'“ (Turing, 2004, p. 214).

Turing offers us this test as a benchmark to show us whether the machine can think (although "think" is too ambiguous a word), but, at the end, he admits that the question of whether a machine can think is meaningless (Turing, 2004, p. 219). The purpose of the test is not to see if the machine is a thinking being, but to see the possibility of how much it needs to process certain information and give an answer to the question asked within the game.

The answer that the machine will give will always be limited by the information given to it and the rules it has that manipulate the information to give a certain answer. The Turing test is not a test to see if a machine has the same intellectual capacity as a human, but a test that serves as an indicator of how quickly a machine can perform operations.

It should not be denied that there are similarities that can be found between the work of a machine and a human (a "computer metaphor" as Putnam would put it), but for the context of this work, the basis on which they operate in arriving at what a particular linguistic entity means is radically different. In the continuation of the work, it will be elaborated why the rules by which humans performs the communication function are not the same as with AI.

Two types of rules

In the previous two chapters, the goal was to show how certain linguistic forms acquire their meaning or content within a given language. Rules are the key to understanding because they are indicators of what a certain word means. Within this chapter, the goal is to highlight the two types of rules we have.The first type of rule is related to human language, such a garden of rules is contextually sensitive, it depends on the conduct of linguistic practice between members of a community, and the criteria for correctness of meaning is one language community. Another type of rule is that used by AI or a system of rigid rules that predetermine the way in which given information will be displayed. Those rules are given in advance. The first type of rules is the one we associate with Wittgenstein and we will call them pragmatic rules, while the second type of rules will be called rigid rules.

Wittgenstein himself notes that the machine and man operate in completely different ways, in terms of problematic meaning:

> „The machine as symbolizing its action: the action of a machine—I might say at first—seems to be there in it from the start. What does that mean?—If we know the machine, everything else, that is its movement, seems to be already completely determined. We talk as if these parts could only move in this way, as if they could not do anything else. How is this—do we forget the possibility

> of their bending, breaking off, melting, and so on? Yes; in many cases we don't think of that at all. We use a machine, or the drawing of a machine, to symbolize a particular action of the machine. For instance, we give someone such a drawing and assume that he will derive the movement of the parts from it. (Just as we can give someone a number by telling him that it is the twenty-fifth in the series i, 4, 9, 16, ...) "The machine's action seems to be in it from the start" means: we are inclined to compare the future movements of the machine in their definiteness to objects which are already lying in a drawer and which we then take out.—But we do not say this kind of thing when we are concerned with predicting the actual behaviour of a machine. Then we do not in general forget the possibility of a distortion of the parts and so on.—We do talk like that, however, when we are wondering at the way we can use a machine to symbolize a given way of moving—since it can also move in quite different ways. We might say that a machine, or the picture of it, is the first of a series of pictures which we have learnt to derive from this one. But when we reflect that the machine could also have moved differently it may look as if the way it moves must be contained in the machine-as-symbol far more determinately than in the actual machine. As if it were not enough for the movements in question to be empirically determined in advance, but they had to be really—in a mysterious sense—already present. And it is quite true: the movement of the machine-as-symbol is predetermined in a different sense from that in which the movement of any given actual machine is predetermined. (Wittgenstein, 1986, pp. 77 – 78)

Machines simply have a predetermined system of operations that they can and do perform. No matter how complicated the written computer systems are, no matter how much code goes into the foundation of a program, the base is the same. Rigid rules determine in advance what a word will mean in a system. The correctness of what something means in a computer language depends on the rules that control and place certain content in a certain order.

Anthropomorphizing computers or interpreting computers as similar to humans starts with understanding what kind of rules lie behind how we determine the meaning of given linguistic entities. Wittgenstein is thus right when he says: „If you do not keep the multiplicity of language-games in view you will perhaps be inclined to ask questions like: "What is a question?" (Wittgenstein, 1986, p. 12). The relation from not knowing the function of language with regard to the problem of meaning leads us to positions in which we interpret the work of machines as similar or the same as humans.

If we look at the example of LaMDA, which answered the engineer's questions in which it talks about the fear it has, the given setting in which we distinguish between two types of rules will not lead us to interpret or see the computer as a human. If, for example, to the question "What are you afraid of?", the machine gives the answer, "Being disconnected from the electricity.", does it mean that the machine understood what the word “fear means,” so it can respond to it from its own perspective or opinion. The answer is actually no. The machine only handles a large amount of information and rules that enable it to be approximately responsible. However, we have to take into account the fact that LaMDA AI has far more information at its disposal, more than even Turing could imagine.

Whatever question we ask the AI, the answer we get is the one that the rules system allows it to answer.

Man, on the other hand, if we follow Wittgenstein's argumentation, does not function like that. The rules he uses are of an external, holistic, pragmatic character. That is, they are contextual and fallible. What a given word exactly means depends on the way in which language practice is carried out, it depends on how all participants of a community react to the statements of others and on the relationship with which we manage to achieve the semantic minimum necessary for understanding, manifested in what we call a “rule” in Wittgensteinian terms.

I would like to point out here that the argument about the kind of rules we follow in the matter of semantics was also made by John Searle in *Minds, Brains and Science*. Searle follows the same line of argumentation in which he states that AI follows formal rules that manipulate symbols, that is, AI works according to a formal procedure written by whoever programs the actual language (Searle, 1984, pp. 44 – 45). Searl shows this with the famous example called “Chinese Room”. It is the assertion that AI cannot know the meaning of certain linguistic creations that Wittgenstein and Searle agree upon. However, the differences begin to arise when Searle develops his argument against the theory of strong AI and when he claims that we explain human cognitive processes as a system of rules that manipulates symbols - computation. Thus, Searle points out that:

> “According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the

> programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations" (Searle, 2004, p. 235).

Namely, Wittgenstein and Searle differ in the fact that Wittgenstein's argumentation and criticism go in the direction of externalism. Wittgenstein argued that the key to understanding semantics lies in the language practice and in its context. However, Searle is not an externalist.

In the continuation of the paper, the goal is to expand Wittgenstein's main argument about rules and to additionally establish the degradation between the two types of rules that we have discussed so far.

Wittgenstein(ian) argument(s)

Within this part of the paper, the aim is to offer two arguments which, in their appearance, are not explicitly Wittgenstein's, but correspond to what Wittgenstein advocated, as well as one Wittgenstein's argument which is a ragged link of the first two. Namely, these arguments are aimed precisely at pointing out that the problem of meaning in language remains an area in which we can point out that the language of a machine and a human remain different in terms of semantics. The examples that will be discussed are: Putnam's "Twin Earth" and Kripke's "Skeptical paradox".We also want to point out that different philosophers, whoever they may or may not have been, under the influence of Wittgenstein, have pointed out the same problem of meaning, and we will argue that the given examples in their "essence" are Wittgensteinian.

Putnam's Twin Earth experiment went in the direction of criticizing the idea of internalized semantics or semantics that are stored in our mental processes. First of all, we will briefly present this extemporaneous theory and show how it builds on Wittgenstein's ideas.

We have to imagine Twin Earth as a planet similar to the Earth, that is, a planet that is a copy of the Earth but differs in certain elements. Namely, on Twin Earth, sciences were developed identically to those on Earth, but science that is similar to chemistry on Earth, because the formula for water H20 is written as XYZ. Namely, water as a phenomenon in nature is described by the chemical formula on Twin Earth as XYZ, while on Earth it is marked as H20 (Putnam, 1972, pp. 223-224).

If we were to go by the idea of innate and internalized semantics, people on planet Earth and Twin Earth should call the appearance of water the same. But that is not the case. The reason is that the meaning we give to certain phenomena does not depend on any internalized processes, but on the external conditions in which we find ourselves. To know the meaning of a phenomenon in the world means simply to learn the language of the community of insiders that we operate. The relationship "meaning just anit in the head!" (Putnam, 1972, p. 227). The thesis that Putnam imposes on us here is socio-linguistic, and he argues that the choice of vocabulary used to describe the world depends on the language community. At this moment, both Putnam and Wittgenstein follow a similar path, where both understand the idea that understanding the meaning depends on the context of use, adoption and use of a vocabulary in a language community. Calling a phenomenon in the world differently is nothing more than a reflection that we have acquired a vocabulary that is useful to us in a certain way, i.e. we use it in a certain way.

For Putnam, the three main reasons for understanding meaning are: 1. meaning is holistic, 2. meaning is normative, 3. meaning depends on our physical and social environment in which we find ourselves (Putnam, 1991, pp. 8-18). All three reasons apply to Wittgenstein's ideas. Holism emphasizes that the meaning of a linguistic entity depends on all the contexts we have within the use of language. The correctness of the meaning is dictated by the language community and its reactions to the flow of communication, and the context within which language practice is carried out affects the practice itself. The meaning here is understood as externalist and additionally pragmatic, because it depends on the way in which it is used.

A human can change the rules for correctness of meaning, a computer cannot. Humans have their basis for understanding meaning in their language community, the computer has a rigid system of rules that reproduces the content.

Now let's think of a situation: Let's imagine that we have developed a certain type of AI that has the ability to communicate with people at such a level that it is difficult to distinguish the AI's answers from those that would be given by another human. Namely, let's imagine that we ask the AI if it knows the meaning of the + sign in a certain formula. Of course, the machine will exhaust the information about it from its database and try to give us an adequate answer. However, does the machine understand the meaning or in this context how can the machine know what I, as its interlocutor, mean by the plus sign, maybe I mean some other operation that is written identically but has a different meaning. The machine will give its answer by filtering all the database it deems adequate to give an answer. It will follow the rules written inside it and will give an agreement, but that does not

give the answer that the machine knew what I meant by the meaning of +. This is an attempt to adapt Kripke's skeptical paradox regarding the understanding of meaning or following the rules he proposed for the purpose of trying to interpret the late Wittgenstein regarding the understanding of the problem of meaning. (Kripke, 1984, pp. 8-9)

What we aimed to draw attention to is that the meaning does not depend only on the two answers from the database we have, because we don't know that our interlocutor understands the + sign in the same way as we do. The only way to solve this problem is to go back to the two concept of rules that we have mentioned earlier. Namely, the communication rules serve as indicators of correct linguistic behavior, they arise through the linguistic burst, depend on it and are implicit in it. Whether I speak correctly depends on the way of communication, which is unpredictable. Communication led by an AI is predictable because it is guided by a rigid set of rules that transforms the given content.

Even within human communication, there is a difference between whether we are merely reproducing the content we have acquired or whether we really understand the meaning of that content. Meaning and understanding are always the main criteria of a community, and actually reproducing content is what AI can do. Now, let's imagine that the machine gave an answer to the questions about the meaning of the + sign and that this meaning (content) coincides with the meaning of its cognate. The answer would be no, because that would ignore what we wanted to highlight by separating the two types of rules.

A person acquires meaning through communication with members of a community, a computer through a series of algorithms provides content that people enter into the program code. As Kripke argues (or rather interprets the late Wittgenstein), there are three main reasons to understand meaning as conditioned by the language practice of a community: 1. agreement, 2. form of life, 3. criteria (Kripke, 1984, pp. 96-99). All three arguments cannot be applied to the language in which AI operates.

Both of Wittgenstein's examples actually intersect in what we have already briefly mentioned in our discussion, which is the idea of a private language, or simply complex meanings that only one individual can understand. This argument can help us to see that by treating the problem of meaning as non-private it radically differentiates AI from humans and problematizes examples related to mental states such as feelings. Namely, to say that a machine can feel something means to have a language in which the state of things is interpreted in a certain way. The reason

why we interpret machines as having feelings is the reason that we still interpret feelings as private, and we ignore that having feelings means having a pattern of interpreting these states within language games. Understanding a state such as feelings means being a participant in the language game, that is, it means having an already acquired language that allows us to interpret certain behaviors as states of feelings, etc. The reason we interpret machines as the ones that have feelings is related to the fact that we have accepted the idea of a private language.

Wittgenstein says:

> "In what sense are my sensations private? — Well, only I can know whether I am really in pain; another person can only surmise it.—In one way this is wrong, and in another nonsense. If we are using the word "to know" as it is normally used (and how else are we to use it?), then other people very often know when I am in pain.— Yes, but all the same not with the certainty with which I know it myself I—It can't be said of me at all (except perhaps as a joke) that I know I am in pain. What is it supposed to mean—except perhaps that I am in pain? Other people cannot be said to learn of my sensations only from my behaviour,—for I cannot be said to learn of them. I have them. The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself" (Wittgenstein, 1986, p. 89).

Wittgenstein specifically uses the concept of pain in this example, and this quote highlights an important point which claims that the way of understanding what pain is depends on all the people who interpret one's linguistic behavior.That is, what it means to be in the midst of pain will depend on the language game we use to describe that state. Descriptions of the state of colors are not certain, because language games are not like that, they are changeable and contextually sensitive.

Wittgenstein further claims:

> "Now, what about the language which describes my inner experiences and which only I myself can understand? How do I use words to stand for my sensations?—As we ordinarily do? Then are my words for sensations tied up with my natural expressions of sensation? In that case my language is not a 'private' one. Someone else might understand it as well.—But suppose I didn't have any natural expression for the sensation, but only had the sensation? And now I simply associate names with sensations and use these names in descriptions" (Wittgenstein, 1986, p. 91).

We can only know what state we are in through language, which gives us the possibility to have a description of a certain state, pain or other feeling. Language serves us to describe certain behavior. Being in a certain state, such as pain, means publicly behaving in a certain way and having language as a means of describing certain behavior in a certain way. Memories must be public in order to be understood or even explained.

Furthermore, Wittgenstein explains:

> "What would it be like if human beings shewed no outward signs of pain (did not groan, grimace, etc.)? Then it would be impossible to teach a child the use of the word 'tooth-ache'."—Well, let's assume the child is a genius and itself invents a name for the sensation! —But then, of course, he couldn't make himself understood when he used the word.—So does he understand the name, without being able to explain its meaning to anyone?—But what does it mean to say that he has 'named his pain'?—How has he done this naming of pain?! And whatever he did, what was its purpose?—When one says "He gave a name to his sensation" one forgets that a great deal of stage setting in the language is presupposed if the mere act of naming is to make sense. And when we speak of someone's having given a name to pain, what is presupposed is the existence of the grammar of the word "pain"; it shews the post where the new word is stationed" (Wittgenstein, 1986, p. 92).

To have a feeling means to publicly project it, so that other participants in the communication understand that behavior as one that is identified as pain. As in the example of toothache, that state of pain first of all had to be publicly expressed in order to be interpreted at all, and secondly it had to be learned from a specific language community within which a specific individual was located. Because „you learned the concept 'pain' when you learned language" (Wittgenstein, 1986, p. 118).

If we accept Wittgenstein's argumentation on the issue of semantics in language, we see that the reasons why the states of pain and similar sensations we interpret in the world depend on the language games we use. The question of whether a machine that can use language, that is, communicate successfully with the people who examine it and also understand the meaning of the terms it uses, is a wrong assumption. We can even argue that the language of the machine and that of a human being are dissimilar in the very question. The rules that serve as indicators of correct behavior are indicators of that. Meaning only arises if we have the opportunity to participate in a language game, the machine does not understand the meaning of the content it operates on, so it only acts according to the rules that were given to it beforehand.

Now, let us go back to the premise that a certain AI can possibly have feelings. If we go back to the two types of rules argument, the semantics of human and AI language are different. language is a product of communication, it is unpredictable, and it changes,while the language of a machine is formal and limited to groups of rigid rules with which its program operates.

Feelings, on the other hand, for Wittgenstein are also given through language. Understanding that someone is in a certain state of pain, happiness, hatred, nervousness, is connected exclusively to language. That is, the way someone learned to behave linguistically in a language community, and from the community that further evaluates these states and behaviors. The criteria for having an absolved feeling is a public criterion.

However, if the machine offers the answer that it is in a state of fear, it means that its system of rules is set to perform functions in a certain way. The machine processed the content it had available in the way it was designed to do. On the other hand, the machine answers the questions in the way it was programmed by humans. This misconception occurs as we continue to interpret the concept of feelings as private.

Knowing that someone is in a state of fear depends on the language we use to describe that type of behavior. This language has some social background. The criteria for correctly interpreting that a sentence is in a certain state depends on how we have been taught to behave linguistically. Human language is contingent and context sensitive in terms of meaning. Machine language is a rigid system of rules and information that is given to it to perform calculations.

Conclusion

The difference in semantics between the two types of language should give us insight into the fact that we understand concepts like feelings quite differently. To say that AI is sentient is to interpret it as such. Wittgenstein's teaching (pragmatic) about language gives us the possibility to understand how we operate language differently from machines and that anthropomorphic AI arose from a misunderstanding of the function of language. The reasons why it comes to mixing the work of a machine and a human being remain in the sphere of understanding humankind in a Platonic way and that every advancement in technology that even remotely resembles man is interpreted anthropomorphically. To have feelings for philosophers like Wittgenstein does not mean to have something like a soul and essence, but it simply means to be a participant in a language game. Machines are also part of those games and their work is being interpreted. AIs like LaMDA do not

have feelings because they do not have the same criteria for identifying feelings that humans have when communicating. AI does not understand the meaning of linguistic entities, understanding meaning is a social game played by humans. The AI only performs the operations given to it. If we are still imprisoned in some Platonic framework, then the interpretation of AI's feeling or thinking remains a possibility. Wittgenstein's philosophy gives us the insight that this is neither necessary nor possible, because there is no private language. A serious philosophical problem about the issue of feeling in AI is not necessary, because the problem of feeling in humans is not a solved problem.

In the case of AI LaMDA, Google itself denied that this AI is sentient. Our argumentation only served to show through a philosophical prism why this is so. With the development of new technology, the idea of interpreting machines anthropomorphically is becoming increasingly rare. Obviously, the new paradigm for the development of computer technology is not a comparison of the work of a machine and a man, at first there is no comparison at all, as in the case with the development of neuromorfic computing and neuromorfic chip (CPU).[1] In this case, the CPU technology is developed by applying our knowledge of operation of neural networks to technology. This is not a comparison of man and machine or an attempt to copy the work of a non-human human in order to create a machine similar to him.Whether AI has feelings is not a serious philosophical question, but how we use technology as a society is.

References

Neumann, J. v. (1945) *First Draft of a Report on EDVAC*. http://abelgo.cn/cs101/papers/Neumann.pdf

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*. Harvard University Press.

Putnam, H. (1979) *Mind, Language and Reality (Philosophical Papers, Volume 2)*. Cambridge University Press.

Putnam, H. (1991) *Representation and Reality*. The MIT Press.

Neumann, J. v. (1979) The Computer and the Brain. Yale University Press

[1]Eg Intels Lohi 2 CPU https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html

Wittgenstein, L. (1986) Philosophical Investigation. Basil Blackwell

Searle, J. (1984) *Minds, brains and science*. Harvard University Press

Searle, J. (2004) Minds, brains, and programs, in John Heil (eds.) *Philosophy of Mind (A Guide and Anthology).* Oxford University Press, pp. 235 – 252.

Turing, A. M. (2004) Computing machinery and inteligence, in John Heil (eds.) *Philosophy of Mind (A Guide and Anthology).* Oxford University Press, pp. 212 – 234.