članci
Articles

UDK 004.896:17
1Wittgenstein, L.
https://doi.org/10.31337/oz.80.1.1
Izvorni znanstveni rad
Primljeno: 28.2.2019.
Prihvaćeno: 4.12.2024.

Some (Wittgensteinian) Remarks on the Ethics of Artificial Intelligence

Borut Cerkovnik[*]

Logic is the Ethics of thinking, in the sense in which Ethics is the bringing to bear of self–control for the purposes of realizing our desires.
Charles Sanders Peirce

Summary

I argue in favor of a distinction between human understanding and machine "understanding". Based on Wittgenstein's view on machines and his considerations on understanding, I aim to demonstrate that no machine with artificial intelligence can reach functional equality with human beings. In particular, this also holds for ethical praxis because it consists of an extremely blurred net of language–games, guided by ethical rules. Therefore, a machine can never have the human ability (disposition) to act ethically and cannot be a moral agent.

Keywords: *Wittgenstein; artificial intelligence; applied ethics; understanding; human; robot; machine*

Introduction

Let us begin the paper[1] with two stories from the mass media on artificial intelligence (AI). The stories shed light on the conceptualization of AI and offer some guidelines for our (Wittgensteinian) examinations of AI and applied ethics (AE).[2]

* Borut Cerkovnik, Ph.D., Assistant Professor, Faculty of Arts, University of Ljubljana. Address: Aškerčeva 2, 1000 Ljubljana, Slovenia. E–mail: *borut.cerkovnik@ff.uni–lj.si*

1 The paper is a result of work done within the research program *Philosophical Investigations* (P6–0252), financed by ARIS, the Slovenian Research Agency.

2 As an anonymous reviewer suggests, a reader must remember that Wittgenstein's views on ethics and AI »were quite specific, if not to some point even idiosyncratic«. Firstly, for Wittgenstein, ethics consists of acting, not theory or interpretation, which is clear from his *Lecture on Ethics* (see Wittgenstein, 2014). Ethics consists in following ethical rules through praxis. This praxis is not uniform, as the cited § 77 of Philosophical Investigations below claims (Wittgenstein, 2009), but is a family of similar yet different practical applications of »good« (see Wittgenstein, 2009, § 66). Secondly, as has been pointed out below, from the cited passage on Turing machines, Wittgenstein refuses the view that they have what could be called original intentionality. If the

The first is from an interview with Slovenian AI expert Marko Grobelnik[3] in *Mladina* (Zgonik, 2018), in which he answers questions about AI's present status and future. As he explains in the interview, his team »gathers about 100 million articles from all over the world, which we then 'compress' into particular events. About one million of them are collected per year« (Zgonik, 2018).[4] For the purposes of this article, his point of view is essential, namely: »Our system enables us to detect and analyze the amount of data that no human on Earth can do. Nevertheless, we are still short on understanding of the depth of the events. We measure these events with thousands of articles that report them, but we do not understand these articles. We can grasp key concepts and relations between factors in texts, but today computers are not able to grasp any subtle meaning. I often like to give the example of the fairy tale, *Little Red Riding Hood*, which every five–year–old child understands, and which a computer with highly developed artificial intelligence is not able to understand« (Zgonik, 2018).

»There is no deep understanding yet«, maintains Grobelnik (Zgonik, 2018). After the journalist has ascertained that AI is better than a human in specific strategic games, he is asked about the areas in which people will preserve their primacy over machines the longest.[5] Grobelnik answers, »where a deep understanding of the world is needed«, and this means »reading and understanding philosophical works« (Zgonik, 2018). After all, he claims that according to the most optimistic plans, computers will be able to understand simple texts by the year 2030 (Zgonik, 2018). It is evident that Grobelnik is very moderate in his estimation of the intellectual power of AI. Unlike some developers of AI technology, who maintain that, for example, a virtual secretary already understands everything, and even more, that a human can. Grobelnik's viewpoints regarding AI contain the following philosophical thesis: »(a) There is no categorical distinction between human understanding and machine 'understanding'. Humans and machines share the same sort of understanding. At the moment, the difference is only in degree.«

Another story on AI is from an interview in the same magazine with the »professor for the ethics and culture of robots and AI« Kathleen Richardson[6]

goal is to let off a stone from upon a sheer slope to reach its foothills, *we* want to reach the goal. The same happens with the machine's routines. What is intentional in them is our intention.

3 See his official web page, where he is introduced as »an expert in the areas of analysis of large amounts of complex data with the purpose of extracting useful knowledge« (Grobelnik, s. a.). »His main achievements are from the field of Text–Mining (analysis of large amounts of textual data).«

4 All translations from the Slovene are have been done by the author.

5 In the article, the term *machine* is used as meaning "artificial intelligence machine", except if otherwise indicated. The same applies to *robot* and *computer*.

6 See her official web page (DMU, s. a.). Lately, her work is targeting sex robots, for — as she maintains — »sex robots exists only because of the ubiquitous rendering objects of women« (Kocijančič, 2018). On her web page, she claims that: »Moreover, my work is about showing how persons are still seen as property, things, and objects, and about new debates that promote AI and robots as persons and the ongoing attack on human subjectivity. I maintain in my work a profound ontological difference between persons and things and reject that they are equivalent and can be measured in relation to each other.«

(Kocijančič, 2018). She characterizes the classical understanding of artificial intelligence as pure (science) fiction. This fiction, stemming from Čapek's R.U.R., understands machines as conscious beings, which develop self–consciousness and feelings. She compares the idea of such robots with the idea of goblins or wizards: »In reality, this will never happen« (Kocijančič, 2018). According to Richardson, we fantasize about robots in this way because people »can change people into property« (Kocijančič, 2018). Therefore, one thinks that *the reverse* would also be possible — »that an object will become animated, that robots will develop consciousness« (Kocijančič, 2018). I would add that Asimov's famous *Three Laws of Robotics* could be applied only in such fictional contexts and therefore do not have any philosophical relevance. Therefore, as I will show below, this apparent and fictional likeness between humans and machines will never lead to real equality, particularly in ethical contexts.

In the interview, Richardson stresses that, »of course, robotics and development of artificial intelligence make progress and robots, meant to interact, without doubt will become more sophisticated, but they will never be like humans because humans certainly are not objects« (Kocijančič, 2018). Richardson advocates an ontological difference between humans and robots. Humans are persons and robots are things. She maintains that »the human language is too complex for robots to master in the way we humans do, let alone robots being able to help overcome feelings of loneliness or psychological disorders« (Kocijančič, 2018).

It is then obvious that the so–called deep understanding of AI is not a human understanding and therefore AI will never attain it. The thesis (a) is not true. One of the deepest parts of human understanding is, of course, ethical understanding. Therefore, the ethics of AI does not exist — machines can be neither moral agents nor moral patients.

1. Wittgenstein's view on machines

Wittgenstein defends the thesis that there is an ontological difference between humans and machines. Moreover, he defends the thesis that there is a categorical difference between human understanding and machine "understanding," which is elucidated in the next chapter. Here is one of his most cited fragments on this topic from *Philosophical Investigations*: »Could a machine think? — Could it be in pain? — Well, is the human body to be called such a machine? It surely comes as close as possible to being such a machine. But surely a machine cannot think! — Is that an empirical statement? No. We say that only of a human being and what is like one that it thinks. We also say it of dolls; and perhaps even of ghosts. Regard the word 'to think' as an instrument!« (Wittgenstein, 2009, §§ 359 & 360).

Wittgenstein discusses the same question in the *Blue Book*. He uses it to object to the view that thinking is something that happens in our head and that our head is »the real seat of the activity of thinking«. For Wittgenstein »thinking essentially consists in operating with signs«, but it can mislead us to think that

thinking is a mental activity; that »It is not material, but an event in private consciousness«. If this were so, then machines could also think and thinking would show itself by some processes within them. He proposes an analogy with the statement "A machine can't have a toothache" and stresses that the impossibility of a machine thinking (or having a toothache) is a logical and not an empirical one (Wittgenstein, 1969, 15–16). Later in the book, he returns to the similar question, "Is it possible for a machine to think?" and adds in brackets »whether the action of this machine can be described and predicted by the laws of physics or, possibly, only by laws of a different kind applying to the behaviour of organisms«. Again, he ascertains that this possibility is excluded by logical and not empirical reason, »The trouble is rather that the sentence, 'A machine thinks (perceives, wishes)': seems somehow nonsensical. It is as though we had asked 'Has the number 3 a colour?'« (Wittgenstein, 1969, 47).

For Wittgenstein, the problem of machine thinking (understanding, perceiving, wishing, feeling, etc.) is not hypothetical, which would mean that we describe (or maybe build) a machine that thinks, either in a naturalistic (physicalistic) or behavioristic view whatever we define as *thinking*. Wittgenstein's approach is conceptual: »Don't think, but look!« (Wittgenstein, 2009, § 66). This means looking at all the contexts (language games) where the word *thinking* is used without adopting some theory in advance. It will be found that *thinking* is a family resemblance concept and can not be coped with a definition. Krkač (2012, 23–25) calls this sort of investigation philosophical morphology. The concept of form of life plays the most crucial role in it: »to perform like humans do perform, to use things, to act, to live the life of a human being, regardless of the aspect (everyday action, language, knowledge, consciousness, technology, history, art, and religion) means to have as a context and as an interpretative background (which is *a rough ground as a valley of foolishness*), to be a part of, and to co–create and co–invent particular forms of life« (Krkač, 2012, xxv). According to him (here he agrees with other scholars, especially Hacker), Wittgenstein's position is neither Cartesian nor behavioristic. Wittgenstein starts from ordinary human interactions that can be labeled as a »*first person, plural* (*we*)« perspective (Krkač, 2012, 159). Ignoring this perspective is what forms the fundamental error of cognitive science (Ule, 2008, 157) and AI as well. Concerning ethics, it seems that we can somehow reduce *our dispositions to act ethically* not only to the states of our brains but also to certain internal states of machines with artificial intelligence. (Somehow in the same way as we reduce the solubility of salt in the water to its chemical structure.) Because our moral dispositions stem from our intersubjective praxis, both reductions are impossible.

I elucidate below what exactly this perspective means. Nevertheless, it would be worth emphasizing that only from it can a form of life become somehow *perspicuous*. From this perspective, humans and machines can never be similar or equal, because they cannot share a form of life. (Besides: How could robots, as dead things, even have some form of *life*?) The similarity between humans and machines is seemingly possible only from a misleading behavioristic perspective.

Of course, we can *imagine* that robots can perform some actions as people do, that they communicate *as* we do, that they play games (language–games and others), etc. — with us or without us. Hypothetically, if we can imagine all of these, then we can imagine machine thinking (understanding, perceiving, wishing, feeling, etc.). In a similar way, we can look at puppets, etc. and in a way even at stones (Wittgenstein, 2009, § 283). However, this is just a sort of fiction, which leads »to this: that only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. 'But in a fairy tale a pot too can see and hear!' (Certainly; but it *can* also talk.) 'But a fairy tale only invents what is not the case; it does not talk *nonsense*, does it?' — It's not as simple as that. Is it untrue or nonsensical to say that a pot talks? Does one have a clear idea of the circumstances in which we'd say of a pot that it talked? (Even a nonsense poem is not nonsense in the same way as the babble of a baby.) We do indeed say of an inanimate thing that it is in pain: when playing with dolls, for example. But this use of the concept of pain is a secondary one. Imagine a case in which people said *only* of inanimate things that they are in pain; pitied *only* dolls! (When children play trains, their game is connected with their acquaintance with trains. It would nevertheless be possible for the children of a tribe unacquainted with trains to learn this game from others, and to play it without knowing that it was imitating anything. One could say that the game did not make the same kind of *sense* to them as to us)« (Wittgenstein, 2009, §§ 281 & 282).

All that Wittgenstein says here about these human activities also holds *mutatis mutandis* for human ethical enterprises. They belong *only* to us. Speaking of the ethics of AI would be speaking of ethics in some secondary, not genuine sense, speaking of it in a genuine sense is simply nonsense. As Hacker (1990, 161) says, »the fact that we now apply a limited range of epistemic verbs to our gadgets no more shows an enlightened 'semantic momentum' stemming from insight into the true nature of the mental than does the fact that we have always applied these expressions to dolls or spirits and ghosts (PI §282)«. Ule ingeniously emphasizes the tension by claiming that we cannot look at humans as if they are alive or as if they think but reckon them as living beings, thinking beings, etc. — this is their nature. Of course, we can look at them as dolls, but this is not their nature. »Similarly, I cannot look at a doll, as if it is a simulation of humans, but it (for us) is a simulation of humans. I can look at it as if it is a living, thinking being, but this is not its nature« (Ule, 2017, 115).

For the purpose of argumentation in the section that follows, two parts of paragraphs 281 and 282 of the *Philosophical Investigations* are essential: (1) The rhetorical question "Does one have a clear idea of the circumstances in which we'd say of a pot that it talked?" that will be turned into the analogous answer that no one has a clear idea of the circumstances in which we would say of a robot that it acts ethically. (2) »Only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious«, which is true for acting ethically,

too. From Wittgenstein's remark on Turing's machines — »Turing's 'Machines'. These machines are *humans* who calculate« (Wittgenstein, 1980, § 1096) — and (2) follows that robots (machines) would be moral agents only if they were like humans. It will be shown that because of (1) it is obvious that they are not.

2. Understanding of understanding

There is only one mention of ethics in *Philosophical Investigation*, but it is significant. It appears in the series of §§ 65–88 that Baker and Hacker (2005) in their *Analytical Commentary* label with "Family resemblance, determinacy of sense, and the quest for essence". In this part of Investigations Wittgenstein shows that we cannot define what a language–game or language is, because language–games form a family. »And the upshot of these considerations is: we see a complicated network of similarities overlapping and criss–crossing: similarities in the large and in the small« (Wittgenstein, 2009, § 66). These similarities he calls "family resemblance".

He stresses that we use the concept of (language) game in such away that its extension »is not closed by a boundary«, »the concept of a game is a concept with blurred edges« (Wittgenstein, 2009, § 68 & 71). Someone's knowledge of what (language) game consists of brings forward paradigmatic examples of a "game"; »my knowledge, my concept of a game [is] expressed in my describing examples of various kinds of games, showing how all sorts of other games can be constructed on the analogy of these, saying that I would hardly call this or that a game, and so on« (Wittgenstein, 2009, § 75). What happens if someone wants to define such a blurred concept and draw a sharp boundary, where none exists? Wittgenstein uses an analogy, as follows: if we want to replace a blurred red rectangle with a sharp one, »several such sharply delineated rectangles could be drawn to correspond to the blurred one. — But if the colours in the original shade into one another without a hint of any boundary, won't it become a hopeless task to draw a sharp picture corresponding to the blurred one? Won't you then have to say: 'Here I might just as well draw a circle as a rectangle or a heart, for all the colours merge. Anything — and nothing — is right.' — And this is the position in which, for example, someone finds himself in ethics or aesthetics when he looks for definitions that correspond to our concepts« (Wittgenstein, 2009, § 77).

According to Wittgenstein, we can refer to something simple, primitive, or extraordinarily complex with the term "language–game", and finally even to the whole of language. It is essential that with it we simultaneously refer to the form of life, also, because »to imagine a language means to imagine a form of life. The word 'language–*game*' is used here to emphasize the fact that the *speaking* of language is part of an activity, or of a form of life. What is true or false is what human beings *say*; and it is in their *language* that human beings agree. This is agreement not in opinions, but rather in form of life« (Wittgenstein, 2009, § 19, 23 & 241). The basis of living with others is an agreement in the form of life. Praxis of human life is the ultimate context in which language stands.

Like other human activities, ethical praxis is also interwoven with the language of ethics. Both are part of the human form of life. Knowing ethics means knowing *the circumstances* in which we would say of a human that they act ethically. It is unnecessary (and often superfluous) that we can describe the circumstances. Still, we must practically know them in the sense of Ryle's knowledge–how. Ethical praxis consists of an extremely blurred net of language–games, guided by ethical rules. We are trained to understand, what is "good", what "we should do" and "what is not allowed".

In paragraphs 143–202 of *Philosophical Investigations*, Wittgenstein considers understanding and rule–following. He argues that understanding is not some sort of inner psychological state, but some sort of disposition (ability) which manifests itself by applying a word or a rule: »The grammar of the word 'know' is evidently closely related to the grammar of the words 'can', 'is able to'. But also closely related to that of the word 'understand'. (To have 'mastered' a technique.) But there is also this use of the word 'know': we say 'Now I know!' and similarly, 'Now I can do it!' and 'Now I understand!'« (Wittgenstein, 2009, § 150 & 151).

Similarly, Ryle (2009, 229) says: »Knowing a rule is knowing how. It is realised in performances which conform to the rule, not in theoretical citations of it.« Rule–following is executed in action, where simultaneously understanding — our ability to follow the rule — of it is expressed. As an expression of a rule can appear a sign, an order (Wittgenstein, 2009, § 206), a table (Wittgenstein, 2009, § 86), a signpost (Wittgenstein, 2009, § 35), but »a person goes by a signpost only in so far as there is an established usage, a custom« (Wittgenstein, 2009, § 198). The use of (language) signs executes rule–following and exhibits the intentionality of human action since following a rule is always an action performed for certain reasons, which are the rules themselves.[7]

As already mentioned above, the first and third person perspectives are not sufficient for the understanding of (intelligent) understanding. The *we–perspective* is also necessary. Since, for example, »making a move in chess doesn't consist only in pushing a piece from here to there on the board — nor yet in the thoughts and feelings that accompany the move: but in the circumstances that we call 'playing a game of chess', 'solving a chess problem', and the like« (Wittgenstein, 2009, § 198). In this sense, computers do not play chess, for they cannot have built–in stated circumstances, which can only grow out of the intersubjective praxis of playing. These circumstances form a family of chess games, which is incorporated in the human form of life.

The same *mutatis mutandis* holds even more true for ethical games because acting ethically does not only consist of moving and/or uttering in a particular way, nor in states and procedures of system built into a robot that accompany

7 To fulfill the anonymous reviewer's proposal, I must add that the ability to follow the rules (= knowledge–how) establishes itself through its applications as every other ability (skill). Full knowledge–how is distinct from other kinds of abilities since it is the ability to apply rules in acting (playing chees, winning a game, making decisions...). In this sense, it is different from unintentional skills, such as, for instance, digesting, and innate skills, such as, for example, finding a nest.

its moving and uttering: but in circumstances that we call »doing good«, and the like. However, machines do not have the *we–perspective*; they do not master the circumstances. »Doing good« is not the reason for their behavior. Therefore, they cannot take an active part in the extremely blurry family of ethical games. Machines can only seem like humans; actually, they are akin to animals for »word learning studies provide evidence that at a very early age children come to understand intentional action. And human learning is what it is — namely, cultural learning — because human beings, even when quite young, are able to understand the intentional and mental states of other human beings. Through this understanding, cultural processes take human cognition in some directions not possible in other species — and make human cognition an essentially collective enterprise« (Tomasello, 2004, 58).

The intentions of human actions are inherent to them, but the intentions of machines come from without — from our community — therefore they are our intentions. Thus, we can interpret a machine's action only as one in accordance with some rules, not as an action performed for reasons — no machine can follow rules like humans do. In more Rylean terms one can say that a machine can have some built–in skills (maybe even skills, which can produce new skills) but can never reach a knowledge–how. According to Wittgensteinian (and Rylean) view, which is represented in this article, there are differences between full–blooded knowledge–how that is a sophisticated skill and plain skill. These differences manifest in that (1) plain skill is formed causally or conditionally. Meanwhile, knowledge–how is a result of (intersubjective) learning; (2) when subjects have a certain knowledge–how, they are following a rule, but they can freely choose to follow another rule; plain skill is, in this sense, absolutely determined in advance; (3) a rule is the reason for intentional action, which is manifested in knowledge–how, a plain skill is not acting because of some reason (although it is sometimes explained in this way).[8] So, knowledge–how is possible only within the human form of life. Therefore, robots (machines) would be moral agents only if they were like humans. Obviously, they are not.

Conclusion

Those who argue for autonomous robots' ethics accept the thesis (a) that humans and machines share the same understanding. I consider that this thesis was satisfactorily rejected. Loh's (2018) article is a good survey of their enterprise. She cites Wallach & Allen (2009, 69), namely that »computer systems may be capable of functioning as if they understand the meaning of symbols without actually having what one would consider to be human understanding«. It seems that she agrees with the thesis (a), when she says that »the boundary between func-

8 The border between these two sorts of skills is blurred, but there exist paradigmatical cases of both. One can, of course, say that a migratory bird knows how to reach its home in spring and autumn, although it is clear that it has merely an evolutionary built–in skill to do it.

tional morality and full moral agency to be gradual with respect to a certain type of autonomy«, but she is skeptical about how an artificial system might achieve a functional equivalent to full moral agency in the near future (Loh, 2018, 53). I think that this achievement is possible only in fiction.[9]

Bibliography

Baker, Gordon Park; Hacker, Peter (2005). *Wittgenstein, Understanding and Meaning: An Analytical Commentary on the Philosophical Investigations 1: Part II: Exegesis §§ 1–184*. Oxford: Blackwell.

DMU (s. a.). De Montfort University, Professor Kathleen Richardson. *De Montfort University*. https://www.dmu.ac.uk/about–dmu/academic–staff/technology/kathleen–richardson/kathleen–richardson.aspx (12.10.2024.)

Grobelnik, Marko (s. a.). Short BIO. *AiLab*. https://ailab.ijs.si/marko_grobelnik/ (12.10.2024.)

Hacker, Peter (1990). *Wittgenstein, Meaning and Mind: An Analytical Commentary on the Philosophical Investigations 3*. Oxford: Blackwell.

Kocijančič, Gregor (2018). Kathleen Richardson: Če si žalosten, si raje kot robota omisli domačo žival. *Mladina*, 48, 30 November. https://www.mladina.si/188466/kathleen–richardson–ce–si–zalosten–si–raje–kot–robota–omisli–domaco–zival/ (12.10.2024.) [Interview with Kathleen Richardson in Slovene.]

Krkač, Kristijan (2012). *A Custodian of Grammar: Essays on Wittgenstein's Philosophical Morphology*. Lanham: University Press of America.

Loh, Janina (2018). On building responsible robots. In Kristijan Krkač & Borna Jalšenjak (eds.), *Applied Ethics and Artificial Intelligence* (pp. 46–59). Zagreb: Zagreb School of Economics and Management.

Ryle, G. (2009). Knowing How and Knowing That. In *Collected Papers: Volume 2: Collected Essays 1929–1968* (pp. 222–235). Oxford: Routledge.

Tomasello, Michael (2004). Learning Through Others. *Daedalus*, 133(1), 51–58.

Ule, Andrej (2008). Circles of Analysis. Münster: LIT.

Ule, Andrej (2017). Wittgensteinska kritika zamisli o mislečih strojih. *Analiza*, 21(3–4), 109–120. ["Wittgensteinian Criticism of Ideas on Thinking Machines."]

Wallach, Wendell; Allen, Colin (2009). *Moral Machines: Teaching Robots Right from Wrong*. Cambridge: MIT.

Wittgenstein, Ludwig (1969). *The Blue and Brown Book*. Oxford: Blackwell.

Wittgenstein, Ludwig (1980). *Remarks on The Philosophy of Psychology: Vol. 1*. Oxford: Blackwell.

Wittgenstein, Ludwig (2009). *Philosophische Untersuchungen: Philosophical Investigations*. Oxford: Blackwell.

Wittgenstein, Ludwig (2014). *Lecture on Ethics*. Malden: Wiley Blackwell.

Zgonik, Staš (2018). "Podatki so nova nafta. Kdor ima dostop do podatkov, lahko rešuje probleme." *Mladina*, 39, 28 September. https://www.mladina.si/187556/podatki–so–nova–nafta–kdor–ima–dostop–do–podatkov–lahko–resuje–probleme/ (12.10.2024.) [Interview with Marko Grobelnik.]

9 I am grateful to the participants of the *Scientific Conference on Applied Ethics and Artificial Intelligence*, 8th of June, 2018, Zagreb School of Economics and Management. Their criticism and suggestions were very helpful in making my argument more clear.

Neke (wittgensteinovske) napomene o etici umjetne inteligencije

Borut Cerkovnik[*]

Sažetak

U ovom radu zagovaram razliku između ljudskoga razumijevanja i strojnoga "razumijevanja". Iz Wittgensteinova stajališta o strojevima i njegovim razmišljanjima o razumijevanju pokazujem da ni jedan stroj s umjetnom inteligencijom ne može doseći funkcionalnu jednakost s ljudskim bićima. To osobito vrijedi i za etičku praksu jer se ona sastoji od iznimno nejasne mreže jezičnih igara, koje vode etička pravila. Prema tomu, stroj nikada ne može imati ljudsku sposobnost (dispoziciju) etičkoga djelovanja i ne može biti moralni agent.

Ključne riječi: *Wittgenstein; umjetna inteligencija; primijenjena etika; razumijevanje; čovjek; robot; stroj*

[*] Doc. dr. sc. Borut Cerkovnik, Filozofski fakultet, Sveučilište u Ljubljani. Adresa: Aškerčeva 2, 1000 Ljubljana, Slovenija. E–adresa: *borut.cerkovnik@ff.uni–lj.si*