Will a human always outsmart a computer? An essay

Adam Olszewski

Pontifical University of John Paul II in Kraków, Poland

Abstract

The title question of the paper has its empirical origin in the form of an individual's existential experience arising from the personal use of a computer, which we attempt to describe in the first section. The rest of the entire paper can be understood as a philosophical essay answering the question posed. First the connection between the main problem of the article and its "premonition" by mankind, which was expressed in the form of ancient myths and legends, is briefly suggested. After shortly discussing the problems that early considerations of AI focused on, i.e. whether machines can think at all, we move on to reformulate our title question, about the possibility of outsmarting AI. This outsmarting will be understood by us in a rather limited way as to prevent a machine from completing its implemented task. To achieve this objective, after softly clarifying the basic terms, an analogy is built between the "outsmarting" of a machine by a human (the target domain) and the playing of a mathematical game between two players (the base domain), where this outsmarting is assigned a "winning strategy" in the certain game. This mathematical model is formed by games similar to Banach-Mazur games. The strict theorems of such games are then proved and applied to the target of the analogy. We then draw conclusions and look for counter-examples to our findings. The answer to the title question posed is negative, and it is not clear how far it should be taken seriously.

> *Much, if not all of the argument for*
> *existential risks from superintelligence*
> *seems to rest on mere logical possibility*
> (Dubhashi and Lappin, 2017).

0. An existential ambient of the problem

The appearance of personal computers—which took place in the middle of 1970s in the world (USA) and a decade later in Poland (due to the "iron curtain" which at that time separated this country from the rest of the world)—brought hope for providing man with a useful tool with versatile applications and capable of performing certain very practical functions. Thus, in those days, from the perspective of an average user, a computer standing on his desk was a *machine*, that is a concrete, experiential and *human-friendly* device, supporting his activities, for example as a typewriter or a memory bank for storing data. Since then, over nearly fifty years, such a notion of a useful machine has considerably evolved. This has happened thanks to the extremely dynamic development of information technology (which included increasing the computational capabilities of machines, their speed and memory and developing new and better programming techniques and languages) and the emergence of the Internet. It seems that the present concept of a *machine* has undergone a substantial modification. This machine, which was a friendly tool supporting man and operating in virtual (non-real) time and sepa-

rated from the outside world, suddenly became a machine operating in real time and additionally started drawing us deeper and deeper into a kind of addiction and making us undertake certain activities, which—although performed through a machine standing on the desk— have far-reaching consequences in real time and, more importantly, in reality. This new machine on our desk is, in fact, a terminal of an extremely vast network (*inter-net*), in which powerful algorithms create the elements of the *machine proper*. This *new machine*, which has taken on *a new meaning*, is also interactive in a double sense. First, it is able to learn in a certain sense strictly defined by computability theory, and second, it is known that it must be managed by or is in the hands of a person or a group of people because it is not an in- dependent entity (or at least we do not know about it yet). We can call such a machine *artificial intelligence* (AI). We deliberately omit here a wide array of issues that need to be clarified or systematically listed[1]; these shortcomings justify using the term *essay* in the title of this work. Besides, this paper is my first attempt at analyzing this somewhat strange subject, so I consider it to be the first of a series of papers and an essayistic outline of my further work on the issue.

1. The difference between an algorithm, a program, a machine, and AI

For the sake of clarifying the approach adopted, it will be good to describe succinctly and in a popular scientific way how we will un- derstand the differences between an algorithm, a program, a machine and AI. The main philosophical issues in the relationship between an algorithm, a program, and a machine are:

---

[1] For some clarifications see (Krzanowski and Polak, 2022).

- the existence and mode of existence of an algorithm;
- the existence and mode of existence of programs;
- mutual relations between an algorithm and a program;
- the implementation of an algorithm of a program on a machine.

For the sake of further considerations, We will adopt a solution which refers to the solution of the problem of universals formulated by St. Thomas Aquinas. One *universal* (or idea) may exist and take one of three forms, depending on the way it is approached:

- *ante rem*—in the world of Platonic ideas—an algorithm;
- *post-ante rem*—as a construction of the mind—a program;[2]
- *in re*—physically present in the thing—implemented into a machine (computer) in the form of the processes that control the behavior of the physical device.[3]

In other words, the three objects under consideration are the same object which manifests itself in three different forms or phases. We are not necessarily going to defend the above distinction at all costs, but we believe that it is easy to understand and will facilitate further reasoning although it will not be essential for the core of our considerations.

Since artificial intelligence (AI) is an artefact that exists in reality, its definition should have the character of a *real definition*. However,

---

[2] *Post-ante* is a strange term which is to indicate that from one point of view a program appears before the *in re* phase and after the *ante rem* phase, while from another point of view it appears after the *in re* phase.

[3] Implementation is an extremely interesting concept in the philosophy of computer science. It allows us to transform something abstract into something physical, which requires thorough consideration. There is a certain similarity here with the criticism of Platonism, where one of the objections against Platonism is the impossibility to move from the abstract to the physical.

due to the fact that AI was brought into existence not so long ago, and the fact that its properties are still subject to almost constant fluctuations, it is difficult to give it an adequate real definition. As a result, most often the definitions given by authors of AI are the result of his/her individual decision and/or consist in choosing from a range of detailed and rigorious definitions available in the literature, which we will also do below. On the other hand, the property of intelligence that we attribute to an artificial system is of natural origin. After all, in the original sense, intelligence, as a property, is attributed to man or, by analogy, to an animal. Defining this property in humans is also of a reporting nature. According to researchers of the issue, from a methodological point of view, the general notion of goal is crucial for the definition of AI[4], as the following table summarises (Bringsjord and Govindarajulu, 2022; cf. Russell and Norvig, 2021):

	Human-Based	Ideal Rationality
Reasoning-Based:	Systems that think like humans.	Systems that think rationally.
Behavior-Based:	Systems that act like humans.	Systems that act rationally.

Table 1: Four Possible Goals for AI According to (Russell and Norvig, 2021).

For our purposes, the following definitions of AI are of particular value by pointing to this artificial component of AI as a *machine* or *computer system* or simply a *computer*; and presupposing the relevant *programmes* (algorithms) implemented in them.

---

[4] More about this see e.g. (Dodig-Crnkovic, 2022).

>The capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this.[5]

>Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision.[6]

The above definitions of AI are schematically based on the following *quasi-equalities*:

- E_1: AI = (System computer + Simulation + Data).

However, some progress has been made in recent years and has taken the form of so-called *adaptive systems*. For them, the quasi-equality looks slightly different:

- E_2: AI = (System computer + Simulation + Data + Interaction with Data).[7]

These two variants of AI are not, it seems, functionally equivalent, but it is clear from their presentations that E_1 is 'part of' E_2.[8] Given this important distinction, the reader should be forewarned that our considerations will concern E_1, and consequently AI will be understood abstractly as a single algorithm. However, to reassure the reader, let us note that:

---

[5] Cf. *Oxford English Dictionary*, entry: „artificial intelligence" (*Artificial Intelligence*, 2022).

[6] Cf. (Burns, Laskowski and Tucci, 2022). One can also add other goals here like: decision-making, translation between languages, visual perception; and others.

[7] „Adaptive AI can change its own code to incorporate what it has learned from its experiences with new data" (Kopera, 2021).

[8] This matter od adaptive AI was brought to my attention by an anonymous reviewer for which I thank him.

A. Any task that can fulfill system E_1, can also fulfill system E_2. (from the above observations);

B. E_1 will win the game with a human i.e. will accomplish its task. (based on the results from section five);

C. Ergo: E_2 will win the game with a human. (from A. and B.).

This reasoning should show that, despite limiting our considerations to E_1-type systems, we do not limit the resulting conclusions to E_1, but they also apply to E_2.

2. Myths as 'premonitions' of mankind

The existential situation outlined above may—and sometimes does—cause anxiety in some members of the community of human users, as is amply demonstrated in literature. However, it is difficult to point to any facts to which we have access in the form of empirical verification, on the basis of which a rigorous narrative could be constructed regarding the justification of existential anxiety. On the argumentative side, it is difficult to find such premises that would make it possible to justify a conclusion regarding a future threat from AI. Therefore, literary works devoted to this issue are based on speculations and some even treat it as science fiction, while others treat it as a purely logical possibility (cf. article motto). We, however, do not downplay this premonition of humanity (expressed in literary terms), and in this section we will try to explain where this anxiety may come from. Our explanation refers to myths that appeared in human history a long time ago. Mankind, as a species, through its representatives, has created strange stories called *myths*. I call these stories strange because, being created at a very early stage in the development of our species, they speak of problems that have been continuously accompanied it in its

history. For example, we have the myth of the Sphinx, a hybrid winged creature with the body of a lion and the head of a woman, who killed travelers heading to Thebes if they failed to solve a riddle. A similar pattern can be seen in the case of the dragon of Wawel Castle from the legend of Krakus and of many other creatures from various legends. The motif shared by many of these stories is an attempt to defeat or outwit an evil creature, which poses a threat to the normal functioning of the community in which it appears. The origin of this creature is in some sense beyond the natural. Most often, in these stories an attempt to outwit it is a rational act, such as solving the riddle of the Sphinx or giving a dragon a fake sheep filled with sulfur. Some believe—and these are not only so-called *ordinary people* but also distinguished scientists—that in certain situations AI is *per analogiam* appears as the embodiment of a *mythical creature*, which threatens the normal functioning of a society, violating the freedom or privacy of all or some of its members. And, consequently, there is the question of outwitting the creature. Let us repeat that these myths, being a kind of common heritage of mankind, are the real cause of humans' fears and anxieties.

3. Turing's and Searle's tests—AI's first issues

In 1936, the groundbreaking year in the creation of AI,[9] two formal models of computability of effectively computable functions were published: one in the work by the American mathematician Alonzo Church (lambda calculus) and the other in the work by the American

---

[9] Here we are making a mental shortcut, because strictly speaking it was crucial for AI to create a computer as a machine that practically realises the mathematical idea of computability.

logician—born in Augustów (now in Poland)—Emil Post (machine). A year later the most famous work in the area was published: the one written by the English mathematician and logician, Alan Turing (machine). Very soon Turing's theoretical machines found implementation in the form of real machines, which were called *computers*. This started a discussion about the capabilities of such machines. Turing (1950) proposed what is known today as *the Turing test*, in which a human evaluator judges whether his interlocutor in a conversation held in a natural language is a human or a machine (AI). Turing's intention in this test was to try to answer the question whether a machine can think like a human. He considered the theological argument (one of the arguments against machine thinking), according to which God created man as the only thinking being in the universe, and thinking was a function of the human soul. Without going into the intricacies of the problem here, let us note that according to Catholic theology, thinking belongs to God and to His angels, thus it is not a function of the body or brain as I think Turing probably believed.[10] Turing wanted to convince his contemporaries that a machine can think like a human being in order to contradict underestimation of a machine's capabilities in this regard, widespread at that time. The second test—called the Chinese Room argument—comes from John Searle (1980), and was intended to demonstrate that no digital computer has a kind of "mind" or "consciousness" even if it functionally bears a far-reaching

---

[10] From a certain point of view, calculation does not intrinsically belong to thinking, because, for example, according to Catholic theology, God, although he thinks, does not calculate, because calculation is the manifestation of a certain kind of ignorance. Cf. for example Isaiah 55:8-9: "For my thoughts are not your thoughts, neither are your ways my ways", declares the Lord. "As the heavens are higher than the earth, so are my [. . .] thoughts [higher] than your thoughts." Some people believe (basing this belief on the Bible) that God's words in which He speaks of cause and effect justify attributing thinking to Him.

resemblance to human conversational behaviour. Searle's argument dampened the enthusiasm of optimistic proponents of AI and its unlimited possibilities. It can be said that the texts mentioned in this section share a common goal: to take a stand in the argument about machine thinking at a time when, in popular understanding, a machine's skills were not appreciated. From the point of view of this paper, both texts (Turing's and Searle's) have lost a great deal of relevance since their publication, as so much has changed in this area. Summing up what has been said so far briefly and succinctly: today no one asks whether machines can think but rather what machines can do in terms of thinking and intelligence and where the upper limits of their capabilities lie.

4. Preliminary formulation of the fundamental question

Taking into account what has been outlined above, we can say that, from a particular point of view, AI can appear as an element of reality—as an artefact—which poses a threat to society on the way to the unrestricted realization of its development. Therefore, we should look for means by which we could 'outsmart' this *creature*. In different words, we can ask if a human can stand up to AI that controls him.[11] This is another way of phrasing the question posed in the title of the paper.

---

[11] An extreme case of this is fictionally considered in the plot of the 2009 film "Echelon Conspiracy".

5. Preparatory analysis to adopt a model for consideration

Due to the generality of the problem under consideration and the lack of precision, we are forced to adopt a theoretical model that will at least allow us to answer a part of the question. In the initial paragraph of the paper, we mentioned two senses of interactivity of AI. The second sense, referring to a need for AI to be managed by a human, will not be addressed here, since the anxiety linked with AI concerns the case when AI acquires self-awareness and escapes any human power over itself.[12] Second, we will assume that such AI essentially remains a machine. Subject literature devoted to this area is extensive, especially after the publication in 2015 of the famous open letter "An Open Letter: Research Priorities For Robust And Beneficial Artificial Intelligence", which is now signed by about eight thousands of people involved in science, mostly AI professionals.[13] The authors write there: "[w]e recommend expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial: our AI systems must do what we want them to do [...]". An important work, though long forgotten, is (Good, 1965), in which its author introduces the key concept of *singularity*, understood as the point in human history when an ultraintelligent machine will appear. We think there is one term that frequently appears in subject literature used

---

[12] My attention has been drawn to a film entitled "Saturn 3" (1980), whose plot considers a similar case.

[13] The letter was signed by world-famous scientists and experts, including Elon Musk and Steven Hawking.

to describe the state of our knowledge on the functioning of AI (or certain algorithms), namely *opacity*.[14] Webster's dictionary gives the following definitions of this term:

- obscurity of sense;
- the quality or state of being mentally obtuse.

This state of opacity affects a large area of the phenomena of social life and raises concerns in some members of the global society. This opacity results in a lack of information and—in a sense—makes the entire area of reality epistemically inaccessible. However, this situation is not unique because similar situations appear in the case of the area of knowledge about the spiritual realm, in particular about God, and also in the microworld studied by quantum mechanics. Although God and the microworld are radically different from the world in which we live, we can sometimes find certain similarities, which we call analogies, between them. It is when we find ourselves in a situation which is analogous but epistemically limited with respect to some reality that we often use a research method called the argument from similarity. Its scheme may look like this:

- An object (of type) X and an object (of type) Y are similar (which is symbolically denoted as $X \approx Y$);
- The similarity follows from P;
- Theorem T holds about object X (symbolically: $T(X)$);
- Therefore: theorem T holds about Y (symbolically: $T(Y)$).

It is worth noting that, in general, we may be dealing with two areas of reality, one of which, i.e. a source to which object X belongs, is well known to us, while the other, i.e. a target to which object Y belongs,

---

[14] For example, a recent talk by P. Stacewicz at the *Homo informaticus 8.0* conference explicitly dealt with opacity in the context of AI.

is not well known to us. The reasons for this may vary, and in our case they are opacity or lack of information. Thanks to the similarity we have previously found, we believe that we can cognitively *invade* area Y. More precisely, this means that we can transfer our previously acquired knowledge about X into 'knowledge' about Y.[15]

AREA A	AREA B
1. Object X: game $G(A)$;	1. Object Y: using of a program;
Similarity P:	Similarity P:
2. Player **I**; 3. Player **II**; 4. The game consists of moves and leads to a result. 5. Winning: the play belongs to set A; 6. Theorem T(X) holds: the game $G(A)$ will be won by Player **I**.	2. Algorithm (program); 3. A human; 4. It consists of actions (elementary steps) and leads to an effect. 5. Winning: what the algorithm (machine) "wants" is accomplished; 6. Theorem T(Y) holds: what the algorithm has planned will happen.

Table 2: A summary of the points on which the analogy is based.

Of course, in general, the argument from similarity is not deductive. Its Achilles' heel lies in establishing similarity between objects. In our case, we assume it on the basis of the above sketchy and introductory considerations, while we leave the in-depth investigation of the issue for the future. That is why our assumption and thus our model can be accused of lack of soundness with respect to the phenomenon we model. Our response to this accusation is that our

---

[15] These issues of similarity and argument are closely related to the theory of analogy, but we will not address them here.

approach nevertheless possesses methodological merit because it is at least an attempt to approach the issue. We will now describe area A in detail, on the basis of the above mentioned similarity.

6. Description of the theoretical model, i.e. Area A

Banach-Mazur games were first described in 1930 and were accompanied by a description of a certain problem by the Polish mathematician from the mathematical school of Lvov, Stanisław Mazur, which is recorded in the "Scottish Book" under the number 43. At present, the game is described in the following way.[16]

We will consider an infinite two-person game with complete information, which we denote by the symbol $G(A)$, where $A \subset \omega^\omega$, that is A is a subset of the set of all infinite sequences of natural numbers with zero. The symbol $\omega^{<\omega}$ denotes the set of all finite sequences of natural numbers. The *empty sequence* is denoted by $\langle \rangle$, and the *length* od the finite sequence s by $|s|$. We also have two players: Player **I** and Player **II**, who take turns making *moves*, i.e. choices of natural numbers.

Player **I**:	x_0	x_1	x_2	$\ldots$		$\ldots$	
Player **II**:		y_0	y_1	y_2		$\ldots$	$\ldots$

Table 3: A graphic representation of the game.

We use the moves of both players to create one infinite sequence of the form: $z := \langle x_0, y_0, x_1, y_1, \ldots \rangle$, which we call *a play* in *a game* $G(A)$. Definitions of players' strategies play a key role:

---

[16] The formulations of the given definitions and statements mainly after an excellent exposition given by (Khomskii, 2010) and (Soare, 2016).

Definition 1. Strategy σ for Player **I** is the function:

$$\sigma : \{s \in \omega^{<\omega} : |s| \text{ is even}\} \to \omega. \square$$

Definition 2. Strategy τ for Player **II** is the function:

$$\tau : \{s \in \omega^{<\omega} : |s| \text{ is odd}\} \to \omega. \square$$

Definition 3. Let σ be a strategy for Player **I**, and the sequence $y = \langle y_0, y_1, y_2, \ldots \rangle$ be an infinite sequence of *moves* of Player **II**, then:

$$\sigma^* y = \langle x_0, y_0, x_1, y_1, x_2, y_2, \ldots \rangle,$$

where:

$x_0 = \sigma(\langle \rangle);$
$x_{i+1} = \sigma(\langle x_0, y_0, x_1, y_1, \ldots, x_i, y_i \rangle). \square$

Definition 4. Let τ be a strategy for Player **II**, and the sequence $x = \langle x_0, x_1, x_2, \ldots \rangle$ be an infinite sequence of *moves* of Player **I**, then:

$$x^* \tau = \langle x_0, y_0, x_1, y_1, x_2, y_2, \ldots \rangle,$$

where:

$y_0 = \tau(\langle x_0 \rangle);$
$y_{i+1} = \tau(\langle x_0, y_0, x_1, y_1, \ldots, x_i, y_i, x_{i+1} \rangle). \square$

If $A \subset \omega^{\omega}$ is a pay-off set, then:

- Strategy σ is **a winning strategy** for Player **I** in Game $G(A)$ iff for all $y \in \omega^{\omega}$, we have: $\sigma^* y \in A$.
- Strategy τ is **a winning strategy** for Player **II** in Game $G(A)$ iff for all $x \in \omega^{\omega}$, we have: $x^* \tau \notin A$.

With the above definitions, we can formulate the axiom of determinacy (**AD**):

(**AD**) For each set $A \subset \omega^\omega$, Game $G(A)$ is determined i.e. exactly one of the players has a winning strategy for $G(A)$.$\square$

This axiom contradicts the axiom of choice in the sense that the axiom of choice implies the existence of an undetermined infinite game.

7. Useful theorems and Theorem T

Theorem 7.1. *Let $A \subset \omega^\omega$ be a countable set, then Player **II** has a winning strategy in the game $G(A)$.* (Khomskii, 2010, p.14)[17]

A slight modification of this concept of the game $G(A)$ is the Banach-Mazur game $G^{**}(A)$, where players alternately choose finite sequences of numbers. In our case, we can treat these two game concepts as equivalent, thanks to the adoption of some coding process of finite sequences.[18] For the second concept we have theorems:

Theorem 7.2. *Player **I** has a winning strategy in the Banach-Mazur game $G^{**}(A)$ iff A is comeager* (Soare, 2016, p.213).

Corollary 7.3. *Player **II** has a winning strategy in the Banach-Mazur game $G^{**}(A)$ iff A is meager* (Soare, 2016, p.213).

The meager and comeager sets relate to Baire space ω^ω. Intuitively "comeager sets are large. They form a filter, are dense, uncountable, and are closed under countable intersections. Meager sets are small. They form an ideal, and countable sets are meager." (Soare, 2016, p.212). I mention this because the matter may be of interest to philosophers.

---

[17] In parentheses I give places from the literature where the proofs of these theorems can be found. When there is no such indication then the claim with the proof comes from me.

[18] I owe my attention to this issue and its clarification to Prof. Yurii Khomskii.

Corollary 7.4. *Let $A = GR_1$, then there is a winning strategy for Player **II**.*

Proof. Let $A = GR_1$, i.e. A is the set of all unary general recursive functions. The set GR_1 is countable. Therefore, by virtue of theorem 7.1, there exists a winning strategy for Player **II**. $\square$

Corollary 7.5. *Let $A \subset \omega^\omega$ and $A = \{f : f \in GR_1 \text{ and } f(n) = s, \text{ for even } n\}$. Then Player **II** has a winning strategy.*

Proof. $|A| < |\omega^\omega|$, then by virtue of theorem 7.1 we have the thesis. $\square$

Corollary 7.6. *Let $A \subset \omega^\omega$ and $A = \{f : f(2n) = s\}$. Then Player **I** has a winning strategy.*

Proof. Since $|A| = |\omega^\omega|$, we cannot use Theorem 7.1. The winning strategy for Player **I** consists in continuously choosing a constant s, i.e, $\sigma(\langle x_0, y_0, x_1, y_1, \ldots, y_i \rangle) := s$, for any i. Let any $y \in \omega^\omega$ be a sequence of the moves of Player **II**. Then for each y, $\sigma^* y = z$, where $z := \langle x_0(= s), y_0, x_1(= s), y_1, \ldots \rangle$. Sequence z has such a form that for each i, $z(2i) = s = x_i = f(2i)$ for any function $f \in A$. $\square$

Theorem T. Let $A \subset \omega^\omega$ and $A = \{f : f|_P \in GR_1, \text{ where } f|_P \text{ is the restriction of function } f \text{ to set } P \text{ of all even numbers}\}$, then Player **I** has a winning strategy.

Proof (sketch). The proof runs along the line of the proof of the previous corollary, except for the fact that for each y, $\sigma^y = z$, where $z := \langle x_0, y_0, x_1, y_1, \ldots, \rangle$. Sequence z has such a form that for each i, $z(2i) = g(i)$ for some fixed function $g \in GR_1$. $\square$

8. Transfer of Theorem T to Area B—T(Y)

In this section we will perform the final step announced in section five, which is the argumentation step we are entitled to by virtue of the argument from similarity presented above. Undoubtedly, this step is quite problematic. We assume that we have established similarity between certain infinite cases of Banach-Mazur games and the use of an algorithm. Again, we omit here the somewhat complicated matter of implementing an algorithm on a machine, and (making a shortcut) we will also talk that a human uses a machine (computer). A machine and a human are understood here as players, where an algorithm (machine) is Player **I** and a human is Player **II**. The moves of a machine consist in giving orders, while a human responds to them by performing some operation on the machine. From a certain point of view, we can look at a machine as a place where a game is played and where an algorithm 'meets' the human mind. We model both types of moves as alternating choices of natural numbers by both players in the form of one infinite sequence of natural numbers. Theorem T formulated above precisely expresses an intuition that for Player **I** it is sufficient to always generate a recursively enumerable recursive sequence in the game. Let us now turn to objects similar in terms of the relation of similarity, i.e. to the counterparts of Player **I** and Player **II**, namely an algorithm and a human. As a result of its action, an algorithm should always generate a sequence which is recursively enumerable. We assume this on the basis of Church's thesis and believe that an algorithm cannot actually generate a sequence other than a recursively enumerable sequence.[19] Hence, the sequence resulting from the game

---

[19] This passage requires a longer explanation, but due to lack of space, it is not provided here.

will always have such a recursively enumerable set on even positions. This state of affairs will make it possible to accomplish what was coded in the program (algorithm).

Let us now formulate Theorem T(Y), which—for obvious reasons—will not be precise in the case considered in the paper:

Theorem T(Y). Any program (algorithm) implemented on a machine will accomplish the goal written into the program, and an *ordinary user* cannot change it, which means that the user will always lose.

9. Crackers

Unlike Theorem T(X), which is a mathematical theorem that is always true, Theorem T(Y) is empirical, and thus a counter-example can be found for it. A counter-example in such situations can be generated because of an unforeseen gap in the understanding of basic terms. A group of unusual computer users, called *crackers*, generate counter-examples to Theorem T(Y). The unusual nature of these users of a machine, somewhat akin to hackers, is expressed in their setting a goal for themselves to overcome the limitation implied by Theorem T(Y).[20] There are essentially two ways in which crackers operate: breaking into a program and breaking into a server. Crackers are not ordinary computer users, they are often very knowledgeable and competent in certain areas of computer science, and their inexperienced followers are called *script kiddies*. Crackers break firewalls, i.e. these features of a program which are to ensure victory in the game to the algorithm, thus, essentially they break the rules of the game. Banach-Mazur games do not provide for such cases, although from

---

[20] Note that a Banach-Mazur game does not allow players to break the rules of the game.

the mathematical side this does not change anything because only the assumptions of a theorem become unfulfilled, and the theorem becomes empty satisfied. Thus, in the context of the main question of the paper, we can make this optimistic prediction:

- For any algorithm, if an algorithm follows a program written by a human, then its *cracker* exists.

A pessimistic prediction referring to the notion of *singularity* would sound like this:

- There exists such an algorithm, not necessarily written by a human, that its *cracker* does not exist.

10. Conclusions

Let us finally take a brief look at the entirety of the argument presented in this paper to help the reader grasp its structure. We will list it in points:

   i. We began with some people's existential anxiety about AI;
  ii. We outlined the definitions of the key terms:
 iii. We pointed out the role of myths and legends in the analyzed issue;
  iv. We posed the problem;
   v. We analyzed the stages of building a model based on similarity;
  vi. We described a mathematical model in the form of Banach-Mazur games;
 vii. We formulated a fundamental theorem on mathematical games;
viii. We transferred this theorem to the area of computer (algorithm)-human relations in the form of:

  a. If an algorithm is correctly defined, a human, as an ordinary user, is unable to prevent it from completing the task written in the program;

 ix. We provided a counter-example to the theorem—the cracker problem.

What are the conclusions of the paper? First, the main conclusion is that the computer, understood as an algorithm, will always win in a "confrontation" with an average representative of the human race. Second, based on experience, we know that there are users, specially educated, who are able to outsmart the computer. The third conclusion is that the adequacy of the theoretical model in the form of Banach-Mazur games for the considered problem should be further discussed and this model—as it seems theoretically promising—deserves further investigation.

Bibliography

Artificial Intelligence, 2022. Available at: <https://www.oed.com/viewdictionaryentry/Entry/271625> [visited on 4 January 2023].

Bringsjord, S. and Govindarajulu, N.S., 2022. Artificial Intelligence. In: E.N. Zalta, ed. *The Stanford Encyclopedia of Philosophy* [Online]. Fall 2022. Stanford, Calif.: Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/> [visited on 4 January 2023].

Burns, E., Laskowski, N. and Tucci, L., 2022. *What is Artificial Intelligence (AI)? Definition, Benefits and Use Cases*. Available at: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence> [visited on 4 January 2023].

Dodig-Crnkovic, G., 2022. In search of common, information-processing, agency-based framework for anthropogenic, biogenic, and abiotic cognition and intelligence. *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)*, (73), pp.17–46.

Dubhashi, D. and Lappin, S., 2017. AI Dangers: Imagined and Real. *Communications of the ACM* [Online], 60(2), pp.43–45. https://doi.org/10.1145/2953876.

Good, I.J., 1965. Speculations Concerning the First Ultraintelligent Machine. In: F.L. Alt and M. Rubinoff, eds. *Advances in Computers* [Online]. Vol. 6. New York: Academic Press, pp.31–88. https://doi.org/10.1016/S0065-2458(08)60418-0.

Khomskii, Y., 2010. *Infinite Games. Summer Course at the University of Sofia, Bulgaria* [Online]. Available at: <https://www.math.uni-hamburg.de/home/khomskii/infinitegames2010/Infinite%20Games%20Sofia.pdf> [visited on 4 January 2023].

Kopera, G., 2021. *How Adaptive AI Outpaces Traditional AI Capabilities*. Available at: <https://www.thoughtai.org/post/how-adaptive-ai-outpaces-traditional-ai-capabilities> [visited on 4 January 2023].

Krzanowski, R. and Polak, P., 2022. The Meta-Ontology of AI systems with Human-Level Intelligence. *Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)*, (73), pp.197–230.

Russell, S.J. and Norvig, P., 2021. *Artificial Intelligence: A Modern Approach*. Fourth, *Pearson series in artificial intelligence*. Hoboken: Pearson.

Searle, J.R., 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences* [Online], 3(3), pp.417–424. https://doi.org/10.1017/S0140525X00005756.

Soare, R.I., 2016. *Turing Computability* [Online], *Theory and Applications of Computability*. Berlin; Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31933-4.

Turing, A.M., 1950. Computing Machinery and Intelligence. *Mind* [Online], 59(236), pp.433–460. https://doi.org/10.1093/mind/LIX.236.433.