https://doi.org/10.70265/IDTC7345

THE DEVELOPMENT OF SOLUTIONS AS A TEST FOR THE PRESENCE OF ARTIFICIAL INTELLIGENCE (AI): A PROPOSAL FROM A SECURITY PERSPECTIVE

Plamen Atanasov

Summary: *The development of solutions indistinguishable according to its creator is proposed in the work as a fundamental criterion for the presence of artificial intelligence (AI), no matter whether it is of a digital or other type. The goal is not to justify or reject the information-dominated nature of modernity, but to increase the effect of cognitive and instrumental efforts in the direction of implementing the computer metaphor in the management processes of labour, with which a person actively fits into the environment. The positives and negatives are examined in the context of security, with an emphasis on the Safety Principle.*

Keywords: *Artificial Intelligence (AI); Decision making; Security; Detection criteria*

INTRODUCTION

The article purposefully seeks an answer to the question of the balance between the advantages and dangers that accompany the use of AI in the activity that a person develops in the context of his security. In connection with this goal, the research focus is directed in the direction of the decision-making process, and the hypothesis is investigated:

The inability of a human to distinguish whether a solution proposal is prepared by another human or by an AI-enabled device is a sufficient, but not necessary, criteria for the presence of AI. From the perspective of security, which in modern times is understood as a stable and dynamic equilibrium, a problem arises due to the contradiction that, unlike humans, AI cannot fully capture the determining factors of the environment. Therefore, the non-recognition of AI-generated decisions by humans leads to the threat of making decisions inadequate for humans, which cause conflicts in the action related to the search, achievement and preservation of security.

Undeniable scientific and technological progress, and especially advances in recent years, require a critical rethinking of the vision of artificial intelligence (AI) and linking it to the place and role that this phenomenon occupies in the human decision-making process. Building on this premise, the article focuses on the understanding that decisions form the basis of meaningful human action by which people intervene (change and fit into) the world and which are defined by the term “labour”. Through work, people actively adapt to their environment, modelling it as much as possible according to their needs, intentions and capabilities. Due to its importance and the asymmetrical positioning of man in relation to the world, the intervention is large and inevitably brings its challenges, benefits and negatives. One of the vectors on the decision-action-outcome axis (changing the environment) orients scientific interest in AI issues and its ability to generate solutions that seem acceptable to humans. A view through the prism of the fundamental understanding of security is proposed, which is from the point of view of man, and which provokes not only topical practical, but also philosophical questions. With its resemblance to human thinking, human communication, and human information work, AI is increasingly leading us to question whether humans have a competitor when it comes to fitting into the world around them. To the extent that this eventual competitiveness is seen as a challenge, understanding security provides a promising field for successful problem research because it allows for a view from different directions. On the one hand, according to the scientific vision established in the 21st century – for example, that of the Copenhagen School – security is a social construct understood as a stable yet dynamic equilibrium, which is specific because it has meaning only for the person, although it affects the world. On the other hand, such an acceptance prompts reflection in the direction of pseudo-human constructs, one of which is artificial intelligence. It resembles the human and can replace a part of the intrinsic human activities, which part, thanks to the factors mentioned above, acquires an increasingly substantial expression in a limited but infinite continuum of material and immaterial dimensions of human action. We are no longer discussing the application of artificial intelligence, but rather its pervasive presence in daily life of the human being, of society, of the State and of international aggregates (interstate unions, religious confessions, political alliances, transnational companies, etc.)

In the context of security, these unknowns concern the state, with its entry into the order of world relations, society and its civilizational (in the sense of Voltaire[1]) understandings, moods and attitudes regarding power, civil liberties and democracy. Last but not least, there is the national dimension of security, which concerns military affairs and the protection of the achievements of a country and society in it.

Subjected to logical concept analysis and presented in the light of security, the object-object connectivity shown above between decisions and

[1] Voltaire understood civilization as a philosophical construct that is bound by reason and science, tolerance and human freedoms and rights, work and morality. The philosopher developed these views in his “Dictionnaire philosophique” and “Candide ou l'Optimisme”.

actions (the subject) and AI (the object) leads to a critical rethinking of the concept of artificial intelligence, its binding to and the elaboration of a unified understanding, which is clear both for the philosophical definition of the construct and for the principles related to its implementation.

The reason that lends significance to the hypothesis examined in the present article is, that at least in the first quarter of the 21st century, the understanding of AI has been conceptualized in different directions – psychology-oriented; technique-oriented; digitally-oriented; cyber-oriented, etc. At their root, however, they all reach the unknown, to what extent the phenomenon of "Intelligence" is permissible to move beyond the space of human cognition. This diversity benefits from a variety of perspectives in the development and use of AI, but at the same time provokes the possibility of concealing interventions in established social attitudes. In line with the hypothesis presented, this article develops the view that the philosophical understanding of AI provides a sufficient basis for researching the problem.

The rapid development in the direction of information and its operation shows that there are already enough reasons for the scientific focus of the work to allow solving at least the following tasks: 1) To answer the question: Why is the ability to develop workable solutions a determining criterion for the existence of artificial intelligence?

2) To assess the adequacy of the emerging arguments regarding the need to rethink the term security from the point of view not of man, but of the philosophical construct "Artificial Intelligence".

3) To indicate threat vectors that follow the admission of artificial intelligence to the generation of human solutions. Scientific correctness requires clarification that the limitations of the proposed topic are dictated by the focus on the philosophically oriented understanding of AI intelligence, according to which it is recognizable by the factors of communication and encoding of information, as indicated by the concepts of Alan Turing and John Searle.

1. INFLUENTIAL CONCEPTS

Undeniable scientific and technological advances in recent decades make AI necessary and probably indispensable for human everyday life in the 21st century – for manufacturing; about information connectivity – from the one with the closest people to the one between the countries; for medicine; for military affairs; for the economy, etc.

Influential countries in the world also agree that AI is an important factor for development in the 21st century, that it leaves a specific imprint on the security of states and their relations, i.e. on international security (UKNCSC; MFAPRC) (understood as interstate (global), regional and national security).

The focus on the impact of AI on the national security of different countries is also not a novelty for researchers and practitioners, and it is more often a topic of research (Baev, 2025). However, scientific correctness requires us to think about whether AI does not put such an imprint on the ideal expression of security[2], which means that it is necessary to broaden the view of the theory according to which we understand and defined the concept.

1.1. The Safety Principle and the Understanding of the Notion of Security

The above circumstances require not only the use of AI to be regulated internationally (since the activities to place the cyberspace within the geographical boundaries of countries are not spectacularly successful), but also to rethink our understanding of security in the direction of the part, for the person, in his capacity as an actor for security implementations.

Modern scholars have no doubt that security is a social construct (Buzan, Wæver & Wilde, 1998) and represents a stable but dynamic as well as specific equilibrium (Yonchev, 2014). Its specificity is due to the clarification that security is a concept that holds meaning exclusively for the human being (Yonchev, 2014). From this starting point we search, achieve to eventual extent, developed and preserve the security. Especially the element of the unintended consequences of allowing computers (various types of digital devices and networks) to make human-active decisions on their own, requires a closer look at the scientific understanding of security, and in its part about man, as the main actor for the search and achievement of security.

We are witnessing errors in the application of AI, which It is unacceptable for this view to escape the fact that the understanding of security has been present for millennia in the organization of human activity. The classical principle proves this circumstance, whose author is Hippocrates, we described by the Latin maxim "Primum non nocere" ("Above all it does no harm"). Today, this postulate we interpreted far beyond the scope of medicine or government and is understood as the guiding rule that an action (scientific discovery, technology, construction, etc.) we not taken unless it is certain that it will not harm a person. It is obvious that from the pedestal of their knowledge to the corresponding historical period, people cannot determine exhaustively what action harms and what does not, but this does not at all render the principle meaningless. In the context of the current study, the big unknown is about AI. We are

[2] From the point of view of the way in which man makes sense of the environment and security, the securities construct can be seen as three cognitively formed visions: ideal (the philosophical essence of the concept), sought (what people see as security and want to achieve) and achieved (the form that the social aggregate has achieved in the realization of the security sought). It is specific that ideal security is unattainable, and the sought and achieved often differ due to the uncertainties of the environment. (Atanasov, 2025).

witnessing errors in the application of AI, which we, as humans can assume to result of a lack of knowledge on the part of the creators of the application. However, it is hardly possible to say that everything related to the application of AI we can verified and subjected to a full risk assessment, i.e. we can fully filter through the Safety Principle.

This is the only way to explain challenges such as: The several-day collapse in 2021 of the activities of the American company Colonial Pipeline, caused by a software bug that turned out to be impossible (or not expedient) to be eliminated if the computer automation of the systems is turned off (USDE, 2021)[3]; The unsuspected socio-psychological cataclysm caused by the reflections of social networks on the psyche of the user (fake news, deep fake, Internet addiction, atomization of society, etc.) or on social relations (the concentration of power in the form of manipulation of information, in the hands of the unauthorized by anyone, let alone by society, people from the so-called information elite); The surprising messages (true, not entirely confirmed) that bots "talk" to each other and are capable of generating solutions instead of the human (Geiger & Halfaker, 2017), etc.

1.2. Artificial Intelligence as a Logical Rather than a Technical Constraint

In a philosophical context, some attempts to define the concept of "Artificial Intelligence", are closely related to digitalization, computing and cyberspace and do not go beyond the framework of two logically sound definitions. One of them is that of mathematician Alan Turing – famous for automating the code breaking to decrypt messages from the ENIGMA encoding machine during World War II (Turing, 1950). The other belongs to the Canadian philosopher John Searle (Searle, 1992).

According to Alan Turing, artificial intelligence should be talked about when one person is talking to another person and a computer at the same time and during the conversation, the interlocutor is not able to understand which lines are coming from the interlocutor and which are coming from the computer.

John Searle proposes that the presence of artificial intelligence be determined in cases where in a message generated by a human-independent source, an independent operator processed (decrypted) the text and the result obtained completely coincides with the first broadcast text. To clarify the defining description, the philosopher proposes the following experiment: In a room that has two windows separated from each other, there is an operator who has a table to decipher. Through the first window, we give encrypted text, with content unknown to the operator. The latter reads it according to the table and transmits the decoded text through the second window. If the

[3] The shutdown of Colonial Pipeline is a typical example of such a solution, which clearly does not correspond to human understanding of security.

text before decryption coincides with the text in the second window, we are witnessing the presence of artificial intelligence.

At least in this article, we understand the two definitions as exhaustive, because the content in these definitions is closely tied to logic and do not limit the understanding of artificial intelligence to specific technical solutions, such as, for example, the requirement to be tied to digital devices. This allows us to talk about AI in the complex mechanisms of the Middle Ages, and even in aggregates of a social or other type but composed of living beings that are different from humans. In this sense, we can replace the term AI by a “computer (computational) metaphor”. Therefore, the breadth of research thought provided by the definitions of Alan Turing and John Searle allows for the full placement of AI in the object field of security, which obviously affects dimensions that are more comprehensive than technical achievements. An example of such an “upgrade” can be extraterrestrial civilizations. They do not fall into the focus of the study. More important to him is the limiting judgment that these beings, to possess a part form of intelligence, must exhibit not only sociality (and different from the eusociality (Crespi & Yanega, 1995) of ant colonies and the like), but also civilization. Another advantage is the easing of reasoning about the relationship between AI and reality, i.e. the increased ability to monitor the quality with which it collects and incorporates data from the environment into its algorithms. In both cases, the proposal to understand this connectedness goes through Voltaire's vision of civilization.

Voltaire's understanding of civilization outlines the relationship between the intellect, the labour (the conscious, active change of the environment) and the decisions according to which the person organizes his activity. Also, dependence on labour and success inevitably leads to an understanding of security. This sequence is:

Intelligence > **Civilization = F (decisions + activity + results)** > **Security**.

This sequence is useful for the present scientific inquiry because it covers object-object dependencies and outlines the civilizational perspective as criteria for evaluating intelligence, regardless of whether the latter is natural (human) or AI.

The result reached leads to the limitations of the hypothesis that substances (man and other similar or not so similar activities) that can show an understanding of security from their point of view, possess not only the ability to manifest intelligence (in other substances the intellect is called artificial), but also the ability to create and develop civilization. As far as no clear boundaries are set between civilization and culture in the non-scientific space, it is necessary to clarify in the article that the term “Civilization” describes “The degree of social development and material culture reached by one or another socio-economic formation” (IBE, n.d.). In the present study, a higher specification is made, using the narrow interpretation of the philosopher François-Marie Arouet, better known as Voltaire, according to whom civilization is the fruit of the triad consisting of reason and science, tolerance and human freedoms and rights, labour and morality (Voltaire, 1962; Voltaire, 2005). In the 18th century, Voltaire gave an example of a counterbalance to the Western European theocracy of the time with China, where the emperor ruled by wisdom rather than by God's law, and where “rational religion” combined with practical wisdom was applied. (Voltaire, 2005)

1.3. Decisions, and Security

Since knowledge is infinite, man cannot fully know his surroundings at a particular moment. This situation allows us to assert that a part of causal relationships that people make is a part that distinguished by a considerable approximation and a high dose of uncertainty. To overcome this uncertainty, the person performs a conscious, purposeful activity, i.e. he makes decisions and manages his actions, with which he transforms the environment. For these reasons, a part of researchers, for example, H. Mintzberg and collective, consider management as a strategic activity through which a person achieves his plans and thus seeks to reduce uncertainty around him. However, there are grounds which grounds we cannot fully confirmed, but also, we cannot deny, and which grounds allow uncertainty to regarded as a negation of certainty (Atanasov, 2022). Then it becomes clear that the relationship between security and governance runs through the decisions on which governance is based, and this relationship is a function of uncertainty. Therefore, by proposing solutions, AI not only interferes with governance, but also introduces additional uncertainty, i.e., compromises security, mixing it with its negation.

This proposition renders management an intriguing factor for the present study. Within such a context, management should be defined as a human activity that integrates the capacity to balance motives, emotions, rational thinking, competence, and humaneness, with the aim of formulating (or selecting) an optimal course of action or inaction to address a given challenge- most often, a significant change in the environment.

This is the reason questions about governance and decision-making attract the interest of psychologists, economists, the military, doctors, engineers, etc. specialists in theory and practice. There is known concepts, of which, in the spirit of the studied problems, that of Herbert Simon should distinguished. It refers to rationality, to limited human rationality in decision-making. According to her, decisions are dependent on environmental factors, but human capabilities do not allow us to know everything and have a complete picture of the world around us. Therefore, people work only with a sample of factors, which sample, in human opinion is sufficient. Such a judgment makes the question interesting: Is AI able to derive a sufficient sample of factors for humans?

Victor Vroom and the team make another interesting clarification. They emphasize the fact that the collective makes decisions according to a corresponding hierarchical network (be it centralized or hybrid[4]) and to a significant extent depend on the leaders at the respective levels (Jago & Vroom, 1977). In relation to this concept, there is nothing to prevent the assumption that AI is available when a certain developed solution fits into one of the nodes of the hierarchical network.

2. A SCIENTIFIC PERSPECTIVE ON DECISION-MAKING AND TESTS FOR THE PRESENCE OF ARTIFICIAL INTELLIGENCE: A CONCEPTUAL PROPOSAL

The issue of detecting the presence of AI, especially around security, is no less important than the question of the effects of its presence. Therefore, it is natural to resort to conceptualizing each of the possibilities for detecting AI. However, conceptualization requires: orientation in the theories used; A review of existing definitions and definitions regarding the essence and characteristics of the subject matter being conceptualized. In the present case, the theoretical basis of the AI and security issues is set out in point 1. However, the issues explored in this work require clarification of the links between decision-making and the factors that indicate the presence of AI.

What is known is that AI is still – despite the ambitions of the developers – AI cannot solve large-scale mathematical problems but is only a good "discoverer" of ready-made solutions. At the same time, its ability to operate with such ready-made solutions and theorems, and thus solve some simple mathematical problems, deserves attention (Seong-min, 2025). The fact that AI can offer solutions, or at least combinations of solutions, that are able to deceive a person and not reveal that they are generated by a computer. After the delusion, these proposed solutions unlock a full-fledged thought process in the recipient, and their human cognitive processing begins. Due to its ability to abstract and to be creative, it can lead to new solutions (somewhere based on AI-generated) that are not at all devoid of the opportunity to lead to an effective and efficient result. Therefore, the ability of AI to search for ready-made solutions and therefore to compile some outcome proposal that may also be applicable due to environmental

[4] In this article, a *centralised network* is conceptualised as a structure that originates from a single leader (the head of the network), whose connections extend exclusively downwards through subordinate branches. The leader issues directives to one or more participants, who subsequently make the necessary decisions and transmit further instructions to additional members, continuing this process until the final form of the decision is reached.

In contrast, a *hybrid network* allows for potential deviations from strict hierarchical order. Within such a configuration, an actor positioned at a lower level may be capable of formulating directives and decisions typically associated with higher levels of authority, depending on the situational demands and contextual contingencies.

uncertainties or due to the mutual compensation of errors in AI's choices should not be disregarded.

All the foregoing indicates that, regardless of whether the result proposed by artificial intelligence is optimal or suboptimal for the active integration of the human being into the world, through its actions AI can trigger a fully-fledged process in the human mind. Among these actions, in addition to the already proven skills for communication (A. Turing Test) and decoding (J. Searle Test), there is also, evidently, the capacity to formulate proposals for decisions concerning the resolution of situations of interest to humans. The latter proposition gives rise to the following scientific proposal:

The presence of AI should be registered when the manager receives proposals for solutions that are generated by humans and digital devices, and this manager does not distinguish which of the proposals is human and which is AI-generated.

As can be seen, the proposal does not exclude the Tests of A. Turing and J. Searle. This means that the proposed criteria for recognisability according to the source of the proposal for a solution is a sufficient, but not a necessary condition. The presence of AI can also be determined in a non-solution situation. For example, AI is also available in cases where the interlocutor does not distinguish whether the interlocutor is a computer or a human, or when the reader does not distinguish whether the encoding/decoding of a message was done by a computer or a human.

3. DISCUSSION ON THE PROBLEM

Linking the existence of artificial intelligence not only to the inability to distinguish one’s interlocutor, but also to the inability to distinguish the creator of a given solution, confronts humanity with a new challenge. This challenge extends across a broad spectrum; however, from the perspective of security, the challenge’s scope appears to be oriented within the following range:

On the one hand, after the first quarter of the 21st century, the individual, society, or state have no choice but to take advantage of AI and incorporate it in their daily lives. Only in this way would they respond to the needs that have arisen because of social development. These needs are widely known Humans cannot calculate as quickly and as error-free (when excluded error-inducing interventions) as computers; The not used automated systems for controlling machines, documentation, and decisions, significantly reduced human productivity. The use of drones and missile shields in warfare in recent years and months has provided a compelling case for AI.

On the other hand, the wording “automated decision system” sounds worrying, especially for democratically minded individuals, who are inherently thinking that humans make decisions by and therefore bear the burden of their responsibility to other people and to the world.

3.1. The Challenges within the Framework of Comprehensive (Wide) Security

The situation with automated decision-making systems leads to several challenges. One possible explanation for them in the context of security is this:

- **Challenges related to the decision-making process:**

An illustration of these difficulties quickly emerges when one reflects upon the ideas of Herbert Simon. The present study considers the uncertainties regarding the quality of the imprint of reality that is implemented within artificial intelligence (AI). It is widely recognised that there are numerous instances in which AI operates on outdated, incomplete, or even inaccurate information, subsequently producing solutions that appear correct but in fact are not. The question therefore arises: Who determines the algorithm by which AI collects information and by which it determines as significant environmental factors? What if, for example, due to the oversight of the developer, the AI has failed to consider as an essential determining factor for the environment?! Is it possible in such cases, in accordance with the speed required by the situation (for example, during an armed clash), for the manager to take a critical look at the adequacy of the proposed solution from the management system? Even more interesting is the question if this proposal is "defective" due to an intentionally embedded error.

It is also worth considering the assumption that AI can participate fully in Vroom's network et al. This gives rise to questions concerning the efficiency-effectiveness dyad, expressed in the following form: Is the responsibility of AI fully developed? And to what extent are we able to position AI in a role that is crucial to the vitality and functionality of the network within which decisions are formulated and made?

- **Related to the technical connectivity generated by the diversity in the development of global processes:**

Through the components in AI carriers, globality concentrates power in the hands of a limited circle of manufacturers. The incident in which pagers exploded simultaneously in several locations positioned Hezbollah's activists, outlined the challenges in this direction. It turned out that these devices are of the same type, purchased under a general order and produced in one batch, eventual from Israel (Shamim, 2024). Therefore, it is not the owner, but the manufacturer who has the final say on the operation of one device.

This further demonstrates that those who own the technology ultimately dictate the rules. It is an obvious fact that due to globalization, a part of industries concentrated in one place to optimize the application of high technologies. The concentration of knowledge and the means of production cause the concentration of capital, because profits are growing, and because of the large volume, it is becoming more difficult to export them outside the country where these profits who's generated. This reason leads to the formation of attractive focuses in the face of leading countries in economic and hence politically. It is an obvious fact that mostly (why not only) these countries are able to produce a decisive part of the components needed to build the AI-powered devices. Then, quite hypothetically, the assumption of the following situation appears to be justified: The leading countries that produce certain components that are inevitable in the construction of an AI-powered device, which is involved in the process of making decisions and implementing connectivity, there are only two. State C starts a war with State C. With the intention of purposefully intervening and drawing the conflict in the direction of its interest, State A or State B, or both, intervene by blocking the relevant components purchased from State C or C with their rights as a manufacturer and rendering the devices unusable. Then the usefulness of AI not only remains in question, but also significantly changes the meaning with which we use the benefits of AI.

Doubts about the validity of the above assumption are dispelled by hints of such interventions. It becomes even more complicated when one considers that these hints come not from a specific country, but from the elite in cyberspace.[5] An example is the high-tech car of Chechen leader Ramzan Kadyrov. Armament is mounted on this pickup truck and serves as a demonstration of high-tech progress. During the war between Russia and Ukraine (Chechen military formations took part on the side of Russia), during the road the car was stopped by the manufacturer and remained stationary (Vasilev, 2024).

• **Related to the psychological context in the understanding of management. Mostly with motivation:**

To part of extent, this challenge repeats the above, but it is not the result of global concentration, but of the individuality of the producer. It develops in the direction of motivation. There, the unknowns orient themselves around the question: What motive did the creator/developer of the AI set (as far as can talk about individuality in the AI carrier)?

It is obvious that, even if the theoretical assumptions about the relationship between AI and decision-making are not known, the motive is always present in management. Therefore, the motives with which the AI proposes solutions should coincide, or at least correspond to, those with

[5] Cyberspace elite, information elite, Internet elite – a multitude of synonyms. A collective, unscientific term to describe a group of people who hold power due to their ownership of key elements of cyberspace connectivity, digital technologies, software, et. c. or due to possession of specific, online-based information or specific skills to work with large databases of online-based data. The presence of such an elite is becoming increasingly noticeable in the information-dominated world of the 21st century. In the context of security, this cyber elite is problematic because it goes beyond the understanding of the state and society and it is not possible to fully apply legal norms to it, incl. and those related to diplomacy and warfare.

which the decision-maker works. Reality suggests that this requirement is not entirely feasible. In AI-generated proposals, the underlying motivation may stem from:

– A product of the algorithms with which AI works. Next comes the question of the reliable operation of these algorithms, the purposefulness of their creation, the completeness of the information collected;

– The product of an accidental error from an oversight in the algorithm or from deformations caused by accidental effects;

– Product of a pre-set opportunity, incl., and such – the developer to insert a bug for a reason beyond the user's control;

• **Related to the Implementation of the AI-propose for a solution:**

These are intuitive interventions typical of cases where the decisions making is in a lack of data. For decades, the challenge has been known that in cyberspace, where the computer mediated communication, the huge amount of accumulated information about users, the stereotypes, etc., allows psychometry of communicators, micro-targeting of the audience and shaping and presenting solutions in their most attractive form. Therefore, the decision-maker is under unequal competition between decisions made by humans (including himself) and those generated by AI. In cases were dealing with the unknowns of the environment implies a choice between several possible solutions (a typical example is command during an armed conflict), it emerges as highly likely that a solution generated by AI will appear more attractive than those made by a human. In these cases, the choice of eventual wrong AI-generated solution will disrupt the optimal rationality that is sought in management and that is necessary to maintain the system.

All these factors, are not always, and it is hardly possible in all cases, to be investigated when trusting an AI carrier.

The foregoing demonstrates as undeniable that the need to link the application of AI to security needs to consider factors of an unusual combination of technological, geopolitical and psychological nature.

3.2. On the Presence of a Pseudo-understanding of Security in the Computer Metaphor

In the context of security, it is also necessary to look at the possibility, the activity of developing and proposing solutions to go not only beyond the control of the user, but also completely beyond human control. There are already known developments according to which the ability of AI carriers to communicate with each other without human participation leads to the threat of developing a construct like the human mind, but aimed at protecting AI carriers, not humans. The famous science fiction writer Isaac Asimov warns about such a scenario in his in his short story "Runaround" (1942):

"(1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) a robot must obey the orders given it by human beings except where such orders would conflict with the First Law;

(3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Law" (Asimov, 1950).

Asimov's script remains in the field of science fiction but is an obvious fact that there are no guarantees that AI-generated solutions will not harm humans.

The threat of AI-thinking is already attracting the attention of scientists. Back in 2017, M. Tegmark warned that the increasing autonomy of AI and it bringing to the field of strategic decisions may allow AI devices to prioritize goals that do not correspond to human interests (Tegmark, 2017). In the same vein, N. Bostrom clarifies that the uncritical admission of AI to strategic management (in this case, regardless of whether it proposes or implements solutions) is especially threatening when used in financial markets or in military use (Bostrom, 2014).

The conception, developed by the two scientists, confirmed by the growing number of popular publications. Ten years later is published the text:

"Artificial intelligence is not explicitly created to generate investment advice, so it can lead to financial losses [...]. According to European supervisory regulators, AI-based tools often work in ways that even their developers do not fully understand. This makes them extremely risky, especially in unregulated financial markets" (Agentsiya „Focus", 2025).

Even these, less comprehensive concepts and notes show that we need critically reconsidered the content into the position that security is a specific equilibrium should. This equilibrium really loses its meaning when it does not apply to the person but requires a refinement of the point of view from which we understood security, and the realizing the corresponding activity. The reason is that, according to the theories mentioned above, constructs such as AI are also able to work out a point of view and prioritize activity suggestions.

3.3. AI-security

The phrase AI-security does not mean that that computing machines also develop the security construct. In the English-speaking space (AI-security) is a term that describes the activity of preventing and protecting against challenges and threats generated by the manifestations of AI. This protection falls within the scope of cyber defence but goes far beyond antivirus programs and their implementation.

A document of the National Centre for Cyber Security, etc. specifies that AI-security is machine learning-oriented and concerns: "– software components (models) in which recognition computers give meaning to messages without the rules being explicitly programmed by a human; – generate forecasts, recommendations and decisions based on statistical reasoning" (NCSC et al., 2023). Tying AI-security to statistical reasoning does not mean that statistics are a threat, but that its misuse is a threat. The ability of AI to apply its algorithms without having received an adequate environmental footprint is problematic. An Einstein's sentence easily explained that reason: "The formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill" (Einstein, 1938).

The document cited above also specifies that AI and AI-users are under to more threats, such as adversarial machine learning (AML) or data poisoning. Attacking machine learning exploits vulnerabilities in machine learning components and modules (hardware, software, workflows, and chains) to warp the AI-generated model, perform unauthorized actions unnoticed, extract data undetected, etc.

Data poisoning is another challenge that AI-security is working to overcome. It is an activity in which malicious scripts are "injected" unnoticed, and, at a certain moment (random or pre-directed) deformed or crashed the data used for machine learning or the connection with the user (NCSC et al., 2023). The term "data poisoning" is gradually gaining a wider scope and going beyond the boundaries of machine learning, affecting more areas of application of AI.

These are the issues that AI-security deals with, and this activity is gaining an increasingly active presence in the work of people, corporations (including cross-border ones), institutions, countries, and international organizations. The presence of such developments confirms the separation of AI as an emphasis in the problems facing broad security in modern times after the first quarter of the 21st century. It is obvious that regulatory regulation of the development of AI is necessary, but as the described issues show, the creation and implementation of such regulation, especially internationally, remains a big question.

CONCLUSION

By proposing solutions, AI introduces additional uncertainty, i.e., compromises security, mixing it with its negation.

The use of AI pose challenge that it is becoming increasingly necessary to develop appropriate security measures in a targeted manner. The development of regulation is associated with a number of difficulties due to the globality, complexity of international relations, etc., and it is unlikely that it will work in its full-fledged form soon, but this does not exempt people from the requirement to achieve and protect their security with, and not in spite of, the application of AI.

AI-security will become increasingly relevant among the components of cybersecurity, and the indicator "ability to develop actionable solutions" highlights the high importance of its role as a link between AI and international security.

The nature of the uses of AI now allows us to look at it not only as a construct for communication or calculation, but also as a construct for making decisions that about human are unrecognizable according to their authorship. That justified the proposal to use the ability to generate unrecognisable solutions as criteria for identifying the presence of AI.

The proposal develops the Turing test and predicts that when one person is offered solutions made by another person or by a collective of people, along with solutions made by a device and the selector cannot distinguish the author of the solution, that device is a carrier of AI.

BIBLIOGRAPHY

Атанасов, П. (2022). Неопределеност vs. Сигурност. *Сигурност и отбрана,* (1), 192-213. https://doi.org/10.70265/NIDJ6406 // Atanasov, P. (2022). Neopredelenost vs. Sigurnost. *Sigurnost i otbrana,* (1), 192-213. https://doi.org/10.70265/NIDJ6406

Атанасов, П. (2025). *Сигурност и комуникиране на информация в публичността*. Пловдив: Макрос. // Atanasov, P. (2025). *Sigurnost i komunikirane na informatsiya v publichnostta*. Plovdiv: Makros.

Баев, Г. (2025). *Изкуствен интелект и новите предизвикателства пред националната сигурност*. В: Сборник доклади от научна конференция „Актуални проблеми на сигурността“, НВУ „В. Левски“ 2025 г., стр. 434-442. Велико Търново: Издателски комплекс на НВУ „Васил Левски“. // Baev, G. (2025). *Izkustven intelekt i novite predizvikatelstva pred natsionalnata sigurnost*. V: Sbornik dokladi ot nauchna konferentsiya „Aktualni problemi na sigurnostta“, NVU „V. Levski“ 2025 g., str. 434-442. Veliko Tarnovo: Izdatelski kompleks na NVU “Vasil Levski”.

Василев, Е. (2024, септември 22). *Tesla „спря тока“ на оборудвания с картечница Cybertruck на чеченския лидер Рамзан Кадиров*. Kaldata.com. https://www.kaldata.com/автомобили/tesla-спря-тока-на-оборудвания-с-картеч-512116.html // Vasilev, E. (2024, septemvri 22). *Tesla „sprya toka“ na oborudvaniya s kartechnitsa Cybertruck na chechenskiya lider Ramzan Kadirov*. Kaldata.com. https://www.kaldata.com/avtomobili/tesla-sprya-toka-na-oborudvaniya-s-kartech-512116.html

Агенция „Фокус“. (2025, октомври 16). *Експерти: Не се доверявайте на изкуствен интелект за инвестиции, рискът е огромен*. Vesti.bg. https://www.vesti.bg/bulgaria/eksperti-ne-se-doveriavajte-na-izkustven-intelekt-za-investicii-riskyt-e-ogromen-6241186 // Agentsiya „Focus“. (2025, oktomvri 16). *Eksperti: Ne se doveryavayte na izkustven intelekt za investitsii, riskat e ogromen*. Vesti.bg. https://www.vesti.bg/bulgaria/eksperti-ne-se-doveriavajte-na-izkustven-intelekt-za-investicii-riskyt-e-ogromen-6241186

Институт по български език (ИБЕ). (n.d.). *Цивилизация*. Речник на българския език. (Онлайн). https://ibl.bas.bg/rbe/lang/bg/цивилизация/ // Institut po balgarski ezik (IBE). (n.d.). *Tsivilizatsiya*. Rechnik na balgarskiya ezik (Onlayn). https://ibl.bas.bg/rbe/lang/bg/цивилизация/

Йончев, Д. (2014). *В търсене на сигурността*. София: Изток-Запад. // Yonchev, D. (2014). *V tarsene na sigurnostta*. Sofia: Iztok-Zapad.

Asimov, I. (1950). *Run Around. I, Robot* (The Isaac Asimov Collection ed.). Doubleday, New York.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Crespi, B. J., & Yanega, D. (1995). The definition of eusociality. *Behavioral Ecology, 6*(1), 109-115. https://doi.org/10.1093/beheco/6.1.109

Damassino, N. (2020). The Questioning Turing Test. *Minds and Machines, 30*(4), 563-587. https://doi.org/10.1007/s11023-020-09551-6

Einstein, A. (1938). *The Evolution of Physics*. Simon and Schuster.

Geiger, R. S., & Halfaker, A. (2017). Operationalizing conflict and cooperation between automated software agents in Wikipedia: A replication and expansion of "Even Good Bots Fight". *Proceedings of the ACM on Human-Computer Interaction, 1*(2), Article 49. https://doi.org/10.1145/3134684

Shamim, S. (2024, September 17). *How did Hezbollah's pagers explode in Lebanon?* Al Jazeera. https://www.aljazeera.com/news/2024/9/18/how-did-hezbollah-get-the-pagers-that-exploded-in-lebanon

Ministry of Foreign Affairs People's Republic of China (MFAPRC). (2021, December 14). *Position Paper of the People's Republic of China on Regulating Military Applications of Artificial Intelligence (AI)*. www.fmprc.gov.cn.

Mintzberg, H., Ahlstrand, B., & Lampel, J. (2009). *Strategy Safari: A Guided Tour Through the Wilds of Strategic Management.* FT Press. ISBN 978-0-273-71958-8.

National Cyber Security Centre (NCSC); Cybersecurity and Infrastructure Security Agency; et al. (2023). *Guidelines for Secure AI System Development*. https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Seong-min, K. (2025, October 20). *OpenAI Retracts False Claims of Solving Decades-Old Erdős Problems.* The Chosun Daily. https://www.chosun.com/english/industry-en/2025/10/20/LCR52ROQAZGE3CKIDTA34HJPFA/

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *LIX* (236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

UK National Cyber Security Centre (NCSC), the US Cybersecurity and Infrastructure Security Agency (CISA). (2023). *Guidelines for secure AI system development*. Crown.

U.S. Department of Energy (USDE). (2021). *Colonial Pipeline Cyber Incident*. https://www.energy.gov/ceser/colonial-pipeline-cyber-incident

Voltaire. (2005). *Candide, or, Optimism* (B. Raffel, Trans.). New Haven: Yale University Press.

Voltaire. (1764/1962). *Philosophical Dictionary* (P. Gay, Edt.). New York: Basic Books.