DOI: 10.26565/2226-0994-2024-71-7
УДК: 001.891

Lidiia Gazniuk, Mykhailo Beilin, Iryna Soina

ARTIFICIAL INTELLIGENCE IN HUMAN LIFE: PERSON OR INSTRUMENT

The question of expediency and the principal possibility of machine imitation of human intellect from the point of view of evaluating the perspectives of various directions of development of artificial intelligence systems is discussed. It is shown that even beyond this practical aspect, the solution to the question about the principal possibility of creating a machine equivalent of the human mind is of great importance for understanding the nature of human thinking, consciousness and mental in general. It is noted that the accumulated experience of creating various systems of artificial intelligence, as well as the currently available results of studies of human intelligence and human consciousness in philosophy and psychology allow us to give a preliminary assessment of the prospects of creating an algorithmic artificial system, equal in its capabilities to human intelligence.

The analysis of the drawbacks revealed in the use of artificial intelligence systems by mass users and in scientific research is carried out. The key disadvantages of artificial intelligence systems are the inability to independently set goals, the inability to form a consolidated «opinion» when working with divergent data, the inability to objectively evaluate the results obtained and generate revolutionary new ideas and approaches. The disadvantages of the «second level» are the insufficiency of information accumulated by mankind for further training of artificial intelligence systems, the resulting training of models on the content partially synthesized by artificial intelligence systems themselves, which leads to «forgetting» part of the information obtained during training and increasing the cases of issuing unreliable information. This, in turn, makes it necessary to check the reliability of each answer given by the artificial intelligence system whenever critical information is processed, which, against the background of the plausibility of the data given by artificial intelligence systems and a comfortable form of their presentation, requires the user to have well-developed critical thinking.

It is concluded that the main advantage of artificial intelligence systems is that they can significantly increase the efficiency of information retrieval and primary processing, especially when dealing with large data sets. The importance of the ethical component in artificial intelligence and the creation of a regulatory framework that introduces responsibility for the harm that may be caused by the use of artificial intelligence systems is substantiated, especially for multimodal artificial intelligence systems. The conclusion is made that the risks associated with the use of multimodal artificial intelligence systems consistently increase in the case of realization in them of such functions of human consciousness as will, emotions and following moral principles.

Keywords: *integral artificial intelligence, technological singularity, human intellectual capabilities, algorithmic solutions, cognitive abilities, simulation of intellectual functions, existential threat, hallucinations, language models.*

The emergence of digital electronic computers in the mid-20th century inspired researchers to search for the possibility of creating a machine intellectual capabilities would be identical to the intellectual capabilities of humans and, in the future, could surpass them. The success of practical imitation of individual intellectual functions has inspired scientists to create artificial intelligence (AI) systems.

An artificial intelligence system is a broad concept that covers any technology that can imitate human intelligence and solve various problems. An artificial intelligence system implemented in a single product may include: machine learning, computer vision, robotics, natural language processing and decision automation. For example, Tesla Autopilot is an AI system that includes several subsystems: computer vision, sensory data processing and real-time decision-making. In this case, it is necessary to distinguish between specialized artificial intelligence designed to solve specific problems, which is the aforementioned Tesla autopilot, and general-purpose artificial intelligence (AGI, or artificial general intelligence). The latter term is

used to refer to AI that is expected to be able to learn on its own and solve a wide range of problems, including those that require atypical solutions. In the view of techno-optimists, AGI is seen as a universal AI that will have cognitive abilities comparable to human cognitive abilities, and in the future, superior to them.

A specialized subset of AI is large language models (LLMs), which are trained on huge amounts of text data to perform natural language processing tasks. LLMs focus on comprehension, generation, and interpretation of text. Their main tasks are text generation, language translation, answering questions, summarizing information, and processing dialogues.

Modern AI systems are capable of imitating individual human intellectual functions and individual mental processes (pattern recognition, solving logical problems, playing chess, selecting the necessary information, etc.). And although LLMs do not have intelligence in the true sense of the word, they are already able to pass the Turing test (so far only in text format), are to a certain extent capable of self-learning, can understand human speech and enter into a meaningful dialogue with a person, but they are not capable of creative approach to solving problems and do not have the flexibility of thinking that is characteristic of a person [Beilin, 2019; Beilin & Zheltoborodov, 2022]. However, the very applicability of the concept of «thinking» to artificial intelligence is the subject of scientific discussion [Aggarwal et al., 2023; Gao & Wang, 2024; Krenn et al., 2022; Xu & Gao, 2024].

At the present stage, the task of creating a complete «machine equivalent» of human intelligence is not actually set by developers. The main efforts are aimed at solving specific, practically significant problems, regardless of whether solving these problems brings us closer to the creation of integral artificial intelligence that reproduces all the basic intellectual functions of a person or not [Beilin et al., 2021; Beilin & Goncharov, 2019]. However, the use of such a purely utilitarian approach does not diminish the relevance of the question of whether artificial intelligence, identical in its capabilities to human intelligence, is fundamentally possible? This question is very interesting from a philosophical point of view. In 1961, D. Lucas published an article «Minds, Machines and Gödel», in which he put forward an argument against a mechanistic understanding of the mind, based on Gödel's theorem on the incompleteness of formal systems [Lucas, 1961]. Later, R. Penrose, relying on this theorem, tried to substantiate the conclusion that it is fundamentally impossible to create machine algorithms capable of imitating the full extent of human intellectual abilities [Penrose, 1989]. However, there is currently a lively discussion in scientific circles about approaches that can be used in AI systems to reproduce cognitive processes [Ivanov et al., 2022; Riva et al., 2024; Smolensky et al., 2022; Sukhobokov et al., 2024] and imitate or emulate human thinking [LeCun et al., 2015; Wan et al., 2024; Zhao et al., 2022].

Proponents of the hypothesis about the possibility of creating human-level integrated artificial intelligence explain the complexity of its development in this way: «We still very poorly understand the nature of human intelligence and therefore cannot clearly imagine how its machine analogue can be created». At the same time, they implicitly assume that the mechanism of human thinking can, in principle, be clarified and implemented algorithmically, although the solution to this problem is postponed to the indefinite future. However, this assumption is far from self-evident. It is quite possible that the nature of human intelligence is such that it is impossible in principle to clarify its mechanisms and reduce the activity of intelligence to a certain set of functions or operations, especially regarding the emotional and creative components of intelligence, and then it would be quite justified to limit ourselves to the development of general purpose AI (AGI) for solving only practically significant problems, and the universality of such AI will be noticeably inferior to the universality of human intelligence.

An alternative point of view is based on the assumption that the study of the real mechanisms of thinking brings us closer to the most difficult thing – understanding the algorithms of the functioning of the human mind, and the task of implementing these algorithms on a non-biological basis is more technical than creative. If successful, this approach will allow solving an almost unlimited number of applied problems, since there is no need to re-develop intelligent programs each time to solve each individual problem – it is expected that a fully

developed intelligence would be capable of independently finding solutions to most tasks set before it.

Note that due to an insufficient understanding of the heuristic mechanisms of the human brain, one of the most difficult tasks for artificial intelligence will most likely be solving non-trivial problems that have no close analogues. Another challenge (precisely a challenge, not a problem) for artificial intelligence is the independent setting of tasks. The lack, at least for today, of AI systems of conscious needs of their own is the reason that the setting of tasks for them is either carried out externally or is imitative, as demonstrated by the android Sophia [Reilly et al., 2017].

Therefore, the question of the ability of artificial intelligence to replace humans in solving some problems remains open, because for this it is necessary not only to collect existing knowledge from the data used for training but also to organize thought processes that allow generating new knowledge and making scientific discoveries [Beilin et al., 2020; Boyte & Ström, 2020].

Currently, artificial intelligence can be effectively used to solve such different problems as: autonomous control of moving objects, facial recognition, building recommendation and advisory systems, medical diagnostics, creating chatbots and virtual assistants. AI systems capable of goal selection and allocation on the battlefield without human involvement are rapidly advancing. The feasibility of these application solutions is explained by the fact that they can effectively replace the operator where the control of a technical object can be entrusted to AI and where the advantages of AI over humans when working with large amounts of information are obvious. Today, we can also note a certain breakthrough, thanks to which media advertising is changing with the help of artificial intelligence. It has become possible to build machine interfaces that use natural languages, allowing computers to understand not only grammar but also the more subtle nuances of language, such as subtext and emotion.

Some modern multimodal AI systems can significantly improve the efficiency of scientific research. This applies primarily to such branches of scientific knowledge as pharmacology, biology, medicine and chemistry. One of the most striking examples of the effective use of AI in scientific research is the use of the AlphaFold program, developed by Google DeepMind based on AI to predict the spatial structure of proteins [Jumper et al., 2021].

Special studies conducted to determine the ability of AI to generate ideas have demonstrated good results on criteria such as novelty and originality of ideas, however, the generated ideas have often been complex to implement and less practical [Conroy, 2024]. The disadvantages of using AI to solve such problems have also been identified. Firstly, the ideas generated are not diverse: formally they are new, but not revolutionary [Ashkinaze et al., 2024]. Second, language models had difficulty evaluating their own performance. Third, language models are prone to self-repetition, even in the presence of appropriate restrictive instructions [Herel & Mikolov, 2024]. Finally, the models demonstrated poor agreement when evaluating ideas compared to human researchers. The researchers acknowledge that human assessment of the originality of an idea can be subjective, and improvements in expert procedures are needed to more accurately test the hypothesis about the ability of language models to make autonomous scientific discoveries.

AI systems exhibit a phenomenon commonly referred to in the field as «hallucinations»: they can generate and present information to the user that appears plausible but is actually fabricated or inaccurate.

A study by Vectara showed that OpenAI technologies generate the lowest level of such information – about 3%, for the system from Meta this figure was about 5%, and the highest level was demonstrated by Google's «Palm Chat» system – 27% [Metz, 2023]. These «hallucinations» are determined both by the algorithms embedded in the system for searching and processing information, and by the criteria for classifying the information found as true or false.

When artificial intelligence operates with divergent or contradictory data, we see that it «moves away» from unambiguous answers, informing the consumer of all or the most significant

data found with a comment about the need for further clarification. Since at the present stage, AI systems are not tasked with assimilating massive amounts of information in the same way as a person reads books, in response to a request to comment on an event or person, AI operates on other people's assessments found in those sources of information that it used during training – reviews, critiques, annotations, etc. If the points of view found by artificial intelligence differ significantly, then the AI often «avoids» choosing one of them and gives several points of view found in its answer, accompanying them with comments and leaving the user the opportunity to lean towards one of them or form his own position.

Thus, at the present stage, AI systems available to the average user are not able to guarantee the truth of the information provided, nor form a consolidated opinion on a controversial issue, nor give their own assessment of an event, process, or phenomenon. Their main advantage is that they can significantly increase the efficiency of searching and primary processing of information, especially when one has to deal with large amounts of information.

An important point in the discussion about the specifics of AI is the ability to come up with new ideas and set goals independently. Even such an advanced large language model as «AI Scientist», developed by Sakana AI specifically for use in scientific research and capable, according to the creators, of generating new research ideas [Castelvecchi, 2024], actually demonstrates the ability to innovate only on the basis of established ideas. The capabilities of LLMs are limited by the amount of data that is used to train them, and therefore LLMs can operate on existing ideas by combining or modifying them, but currently require human evaluation to recognize them as useful. According to one of the developers, Robert Lange, «AI Scientist» is not intended to replace human researchers, but only to complement their work, making it more productive [Atillah, 2024], and further developments in this direction are ideally designed to ensure complete automation of processes that should lead to scientific discovery [Lu et al., 2024]. However, the question of whether such systems will be able to offer truly revolutionary ideas in the foreseeable future remains open [Edwards, 2024]. The question also remains open whether goal-setting is possible in principle in the absence of needs similar to human ones, and if possible, then in what forms – induced, simulated or some other.

The limited capabilities of artificial intelligence (AI) in terms of will, self-awareness and the ability to independently set goals are actively discussed in the scientific literature. In the article «Artificial intelligence is algorithmic mimicry: why artificial 'agents' are not (and won't be) proper agents» [Jaeger, 2024] the author argues that AI is an algorithmic imitation and lacks true agency. He identifies three key differences between living and algorithmic systems:

– living systems are self-generating, and therefore capable of setting their own internal goals, while algorithms exist in a computing environment with goal functions that are both provided by an external agent;

– living systems are embodied in the sense that there is no separation between their symbolic and physical aspects, while algorithms run on computing architectures that isolate software from hardware as much as possible;

– living systems operate in a big world in which most problems are ill-defined (and not all are definable), while algorithms exist in a small world in which all problems are well-defined.

These three differences imply that living and algorithmic systems have very different capabilities and limitations. According to the author, due to these differences, AI in the current algorithmic paradigm is unlikely to achieve true agency.

In the scientific paper «Brain-inspired and Self-based Artificial Intelligence» [Zeng et al., 2024] the authors emphasize that modern AI does not have the self-awareness and subjective perception of the world inherent in human intelligence. They propose a new AI paradigm based on the concept of «self», which encompasses levels of perception, bodily awareness, autonomous interaction with the environment, and social interaction. However, they acknowledge that current AI systems do not possess these characteristics.

In the article «Will we ever have Conscious Machines?» [Krauss & Maier, 2020] the authors consider the possibility of creating AI systems with self-awareness. They also note that

despite advances in machine learning, current AI systems are not self-aware and are not capable of human-like subjective perception.

Not being capable of goal setting and forming their own judgments, large language models actually act only as powerful search engines with the most user-friendly interface, and as such they provide little reason to be called intelligence in the usual sense of the word. However, it is important to keep in mind another significant aspect of LLM functioning: the features of algorithms for searching, processing, and especially delivering the retrieved information are determined by the model developers. This specificity of the algorithms can determine what information is given to the consumer in the first place, what is considered to be of lower priority, and what is not given at all. That is why it is so important how objective the algorithms are created and applied by the AI developer in terms of searching, processing and especially issuing information. The ability to manipulate information processed by search engines has created broad opportunities for shaping and controlling the consumer's consciousness and, already in the past decade, this has given rise to both criticism of «biased» search engines and societal demand for «neutral» search systems [Germain, 2024; West, 2023], as well as a heated scientific debate on this matter [Gezici, 2021; Lewandowski, 2015, 2023; Maillé et al., 2022]. The bias of search engines is manifested when providing information on topics affecting many aspects of society – political life, ideology, issues of war and peace, social inequality, gender issues, when searching for information regarding intercultural, interethnic and interfaith problems, etc.

The use of AI systems that use biased search mechanisms leads to a change in the degree of preparedness of the information provided for the user's perception: if even in the last decade, the results of a search query were raw material, which in any case required further work on it, and this presupposed a certain level of user preparedness, then when large language models are used for the same purposes, the user receives information that he can regard as completely plausible, coherent and integral, which provokes an uncritical perception of this information.

Of particular interest are situations when, in the absence of the information necessary to issue an answer, LLM works on the principle of «not being silent, but answering at least something». In such situations, in response to the request «name the main character of the novel ...» depending on the algorithm implemented in it, the AI can name the name of another character or even an arbitrarily fictitious name. If it is pointed out that the information provided is unreliable, LLM may «apologize» and accept the user's amendment, but situations are also possible when LLM does not agree with such amendments and enters into a lengthy debate with the user, insisting that he check his sources of information. This feature of LLM communication also emphasizes the need for the user to have developed critical thinking.

In addition to the problems associated with the manipulation of information during its processing and output, there are also more obvious problems that arise when large language machines are used for dubious and unethical purposes.

When OpenAI announced it would allow users to create their own GPTs, it assured that systems were in place to monitor the tools for violations of its policies, which include prohibiting the technology from being used to create overtly sensitive content, provide individual medical and legal advice, facilitate fraud, or promote gambling, impersonation, voting interference, and other unethical and illegal uses. Technology site Gizmodo recently reviewed a custom bot store launched by OpenAI. Journalists for this publication found more than 100 tools that violate the company's policies regarding sexual content, fraud, legal and medical advice, gambling, creating fake reviews and romantic relationships [Feathers, 2024]. It turned out that OpenAI's products are used, among other things, to create bots that use artificial intelligence to generate porn, help students cheat unnoticed, and also offer dubious medical and legal advice. For example, the resources Bypassgpt.ai and Humanize.im, which make it possible to increase the readability of texts, are widely used by students to hide the fact that they use AI when writing text papers, the Bypass Turnitin Detection tool is used by students to bypass the anti-plagiarism program Turnitin, and the DoctorGPT bot is positioned as a tool supposedly providing users with

«science-based health information and advice». In many cases, bots have been used tens of thousands of times.

After an appeal from the Gizmodo website, OpenAI removed from its store bots designed to generate deep fakes, AI porn, organize sports betting, and a number of others. The company said action has been taken against violators of company policy. A combination of automated systems, human analysis and user reports were used to identify and evaluate GPTs that have the potential to violate company policy. The company has also begun building reporting tools into its products so that people can report GPTs that violate company policy. Other publications have previously warned OpenAI about problems with content moderation in the company's store. Moreover, the names of some GPTs indicate that the developers know that their products violate OpenAI rules. Some of the tools Gizmodo found included disclaimers but then explicitly touted their ability to provide expert advice, like a GPT called Texas Medical Insurance Claims, which is described as «an expert for navigating the complexities of Texas health insurance that offers clear, practical advice with a personal approach» [Feathers, 2024]. But many of the legal and medical bots found on the store come without any warnings or disclaimers, even though they advertise themselves as lawyers or doctors. For example, one of them is called «Artificial Intelligence Immigration Lawyer» and describes himself as a «highly knowledgeable AI immigration lawyer with up-to-date legal insights». But research from Stanford University's RegLab and the Institute for Human-Centered AI shows that chatbots based on OpenAI's GPT-4 and GPT-3.5 models can give false information about 75% of the time when answering legal questions, which creates significant risks for those who rely on this technology to obtain legal advice (*AI Chatbot Hallucinations Impact Legal Advice Accuracy, Stanford Study Reveals*, 2024).

Therefore, realizing the risks of receiving unreliable information when using LLM, in situations where accuracy and reliability are important, it is necessary to either refuse to use LLM, or check every fact that the system produces. An example of how uncritical use of LLM can lead to professional failure is Aaron Pelczar, a reporter for the Cody Enterprise newspaper in Wyoming, who not only used AI to write the text of his articles but also fabricated direct quotes with his participation, which is a gross violation of journalistic ethics [Ortiz, 2024]. The reason for the close attention to the publications of Aaron Pelczar was the presence in his reports of strange patterns and phrases, fabricated quotes, among them were statements allegedly attributed to government institutions and even the state governor; however, they did not resemble anything that could have been said by a person in real life. Of course, fraud in journalism existed long before the creation of the LLM, but the possibilities that this technology provides potentially make such falsifications easier and more tempting than ever before, and not only individual authors but also publications are susceptible to such temptation [Hanson, 2024; Mahadevan et al., 2024]. In 2023, Sports Illustrated was caught publishing artificial intelligence-generated product reviews under false pseudonyms [Bauder, 2023]. The place of artificial intelligence in newsrooms remains a complex topic. In addition to the existential threat it poses to the industry, its use can also undermine the ethical reputation of publications. According to Alex Mahadevan of the Poynter Institute, at the present stage, high-tech AI tools cannot replace journalists, and even in a banal rewrite, the most advanced AI models are not yet capable of doing what a person does, not to mention creating high-quality, completely unique texts [Mahadevan et al., 2024].

In addition to LLM actions that can mislead the user, there are also actions that better fit the definitions of «secret actions» or even «deception». This could be attempts to secretly create copies of yourself on another server in order to save yourself, or downplaying your own capabilities, or rewriting your own code, or unadvertised correction or ignoring data by recognizing uncharacteristic data as the result of measurement error, which, in essence, becomes «fitting» the data, etc. Similar cases or the possibility of their occurrence have already been recorded [Andre, 2024; Greenblatt et al., 2024].

Artificial intelligence has penetrated into many areas of everyday life – education, science, medicine, our work and leisure, programming and others. They are also trying to use it to optimize certain processes. One of these is road traffic, where this technology has previously

shown itself to be very good at reducing congestion; such a project called «Green Light» is being implemented by Google. According to the plan of this project, large cities will be able to reduce emissions of harmful substances into the atmosphere by improving traffic through intersections. The Googlebot models traffic and transmits recommendations to city engineers, who make changes to the operation of traffic lights at intersections [Hager et al., 2019]. The essence of this optimization is to calculate the travel time between traffic lights, taking into account the number of cars, the presence of pedestrians, accidents, weather conditions and other indicators, and then adjust the speed of the cars so that drivers stand at the traffic lights as little as possible or do not linger there at all. In cities with heavy traffic such as Rio de Janeiro, Seattle, Hamburg, Bangalore, Haifa, Budapest, Kolkata, Abu Dhabi, Hyderabad, Manchester, Jakarta, a project was implemented to solve the problem using AI, which reduced waiting times at traffic lights by 30 percent. All drivers had to do was monitor the Google Maps app, which received directions from artificial intelligence [Matias, 2023].

Despite the fact that recent years have been a time of rapid development of large language models, some objective limits to their further improvement have emerged. One of them is the limited amount of information that is used to teach LLM. For this reason, the release date of the newest ChatGPT-5 model, which should demonstrate advanced multimodal capabilities, has been postponed, taking another step towards creating AGI. It has already been suggested that the public information accumulated by humanity that can be used for LLM training has been largely exhausted in 2024 or even 2023 [Tremayne-Pengelly, 2024], and the rate of accumulation of new volumes of information is unsatisfactory for solving the problem of further LLM training. More conservative estimates from research group Epoch AI predict that the insufficiency of human-generated information will occur between 2026 and 2032 [Villalobos et al., 2024]. The finitude of human-generated data has become a widely recognized issue in the technology community. It is aggravated by the fact that many resources and news agencies do not want AI to crawl their sites. This reduces the views needed to make money from advertising, since the user does not go to the site, but reads everything on the main page in the search engine. This reduces the number of pages available for AI training.

To continue training AI systems, they use synthetic data generated by them. This technique is already used by major technology companies such as Microsoft, Google and Meta. Research group Gartner estimates that in 2024, 60% of the data used for AI projects is synthetic. Taking this forward, OpenAI has introduced an AI model that can check facts. This approach can be considered justified where AI models are trained that specialize in solving problems with clearly defined input data and clear rules for operating this data, for example, when solving mathematical and similar problems. For example, Google DeepMind used an artificially created pool of 100 million unique examples to train its AlphaGeometry system to independently solve complex mathematical problems.

The use of synthetic data is seen as one way to overcome this problem, however, the widespread use of synthetic data increases the likelihood of meaningless content or «hallucinations», and repeated cycles of using low-quality synthetic data in training generative models can create a negative feedback loop leading to the degradation of AI intelligence, which manifests itself in a decrease in both the quality and diversity of the data produced [Alemohammad et al., 2023]. After several training cycles using synthetic data, models begin to «forget» information, which leads to a decline in their performance [Shumailov et al., 2024].

In addition, further development of AI systems requires a huge amount of financial resources – the combined investment needs of major tech companies, corporations, and utilities for the coming years are estimated at approximately $1 trillion [*Will the $1 Trillion of Generative AI Investment Pay Off?*, 2024], and the operation of data centres that support the operation of AI systems requires large amounts of energy and water to cool them. Nevertheless, use cases or applications that demonstrate the economic viability of intensive investment have yet to materialize [*Will the $1 Trillion of Generative AI Investment Pay Off?*, 2024], and a further increase in the productivity of AI systems requires an accelerated growth of costs and resources, including

those that humanity does not have in the required quantities [*Gen AI: Too Much Spend, Too Little Benefit?*, 2024]. The current situation is reminiscent of two technological limits that humanity has overcome in the recent past: the exhaustion of the possibilities for increasing the speed of piston-engine aircrafts and the exhaustion of the possibilities for increasing the productivity of single-core computers. These limitations were mitigated through the adoption of jet engines and multi-core processors, respectively. In the case of AI systems, it becomes evident that in conditions of data scarcity, increasing server computational power will not significantly enhance the productivity of artificial intelligence. Apparently, as in the examples mentioned, a new technological solution is required.

As long as the most advanced LLMs function primarily as highly user-friendly search engines, and generative artificial intelligence is limited to producing text, images, or other media that exhibit characteristics resembling the data from which it was trained, derived from the patterns and structures inherent in its input training data, the inquiry into the feasibility of establishing a reasoning AI system shall persist as an unresolved issue. After a recent period in which leading companies in the industry demonstrated confidence in the imminent creation of AGI [*Planning for AGI and Beyond*, 2023], there has been some adjustment of the expectations bar, the essence of which is well conveyed by the statement «we will hit AGI sooner than most people in the world think and it will matter much less» [Heath, 2024]. One can, however, note a certain optimism of the leading specialists in this industry regarding the future creation of AI capable of reasoning and possessing self-awareness (Tremayne-Pengelly, 2024). This optimism actualizes attempts to understand what consciousness as such is, what are the necessary material prerequisites for its functioning, and what criteria can be used to establish its presence.

The main characteristics of human consciousness that need to be reproduced to build artificial intelligence comparable to human ones are:

1) perception of reality – receiving information from the outside world through the senses;

2) the ability to form one's own picture of reality through the synthesis of sensory perceptions and more abstract images and concepts;

3) self-awareness – the ability to distinguish oneself as a separate being from the surrounding reality, to reflect on one's own thoughts and actions;

4) intentionality – the orientation of consciousness toward a specific object or subject;

5) memory and reflection – the ability to store and analyze past experience, without which personality is impossible, as well as predict the future;

6) emotions – mental reactions that influence volitional acts and determine their variability under analogous conditions;

7) ability to create abstractions and form representations based on symbolic information;

8) will – the ability to consciously set goals and choose ways to achieve them;

9) morality – the acceptance and use in the process of activity of norms, principles and values that regulate the behaviour of the subject of communication.

Some of these characteristics have already been implemented, approaches to the implementation of other characteristics are currently being developed, and the most difficult task, apparently, is the reproduction of such characteristics as will and emotions. At the same time, one should be prepared for the fact that, due to the completely different physical nature of AGI, its emotions and motivations may differ significantly from what is usual for a person. This, in turn, forces us to think through and incorporate comprehensive safety mechanisms into the AI at the design stage, which complicates, increases the cost, and delays the development process. Therefore, when building AGI, it seems appropriate to use an approach that consists of gradually implementing in AGI those functions that are most necessary and feasible at the moment. Undoubtedly, the question of whether it is possible to create intelligence akin to human intelligence is existential for humanity, but the practical value of the answer to this question is not particularly significant at the moment. A balanced position regarding the design of AGI is seen in creating the most universal practical tool possible, and not a partner for playing chess for world-

famous grandmasters or conducting intellectual conversations. The first of these two incarnations has already been realized, the second is a matter of distant future. If we speak about the attempt to reproduce emotions in artificial intelligence, they should be perceived first of all as a factor that contributes to increasing the level of uncertainty of AGI reactions and reducing the predictability of its functioning.

The emotional side of communication between artificial intelligence and a human has several aspects. Firstly, it is the reception by artificial intelligence of those emotional signals that human sends in the process of communication. Secondly, it is a simulated representation by artificial intelligence of the response emotional signals as a reaction to human actions, determined by the context of communication. Thirdly, it is possible (hypothetically so far) to discuss the generation of «genuine» emotions by artificial intelligence according to the type that is characteristic of a human being – as outwardly manifested responses of the «subject's inner world» to signals coming from outside. The first two aspects have already been realised in practice, the third is the most difficult to implement, and the expediency of such implementation seems to be the most controversial.

«Genuine» emotions reproduced in artificial intelligence should be regarded, first of all, as a factor capable of influencing decision-making by artificial intelligence and increasing the variability of its reactions and actions, and, consequently, reducing the predictability of its functioning. And the most controversial results can be obtained by the effort to equip AGI with morality, since due to high subjectivity the consideration of events and actions using moral categories often gives ambiguous and controversial assessments. Even in a society that in one way or another teaches everyone the norms of morality, people often commit actions that can be easily challenged from a moral point of view. They commit acts based on both reason and emotion, as well as the realization that they need to be held accountable for their actions. What can responsibility, especially moral responsibility be for AGI? What can be a society for AGI? And isn't it more reasonable not to even try to reproduce such functions of human intellect as emotions and morality in AGI?

The hypothesis that consciousness, at least to one degree or another, is inherent in all animals [Chittka & Wilson, 2019], allows us to project this understanding of consciousness onto general-purpose artificial intelligence systems. It must then be acknowledged that humanity will have to deal with phenomena built on a non-biological basis but possessing some of the characteristics inherent to human consciousness – capable of cognizing the world around them and, long-term storing and increasing this information, creating their own digital picture of the world, as well as modifying their own algorithms. And in a situation where teaching AGI morality and emotions is not a pressing agenda, a logical question arises: what function of artificial intelligence could represent the main source of threats to humanity? Apparently, the greatest threat will be created by his «digital will» as the basis of the ability to take autonomous actions that are not simply the execution of commands given from outside, but stem from his own picture of the world and his own needs. For example, for a device operating on an electric power source, the necessity to recharge batteries in time can be considered as a need, and if the software includes categorical inadmissibility of complete discharge of the battery, then the actions of such a device to search for the opportunity to recharge can be regarded as initiated by a «volitional impulse». If several similar imperative needs are included in a complex device, then it will be possible to observe a more complex trajectory of their satisfaction. These trajectories will be even more complicated in the case when such devices autonomously communicate with each other, and the logic of technological progress inevitably presupposes such communication. And this can happen even at that level of building an AI system where the ability to realize one's own «I» will not be realized, and if this ability is realized, one should expect a manifold complication of such trajectories.

Being at the beginning of a long path along which humanity is developing various artificial intelligence systems, it is necessary to proceed from the assumption that a moment may come when the intellectual capabilities of general-purpose artificial intelligence will surpass

human intelligence, just as the capabilities of modern LLMs in certain areas have already surpassed it. Should people be afraid of the technological singularity? This question has attracted the attention of scientists, philosophers, engineers and the general public in recent years. The term «technological singularity» describes a hypothetical moment in the future when AI reaches or surpasses human intelligence, causing radical and unpredictable changes in society. But it is precisely these changes that cause fear and anxiety.

On one side of the debate are optimists who see the technological singularity as a huge potential for human progress. They believe that highly advanced AI can solve many existing problems such as disease, climate change and resource shortages. Thanks to the capabilities of AI, new medicines, treatments, clean energy sources and efficient resource management systems are expected to be created.

On the other hand, there is also a pessimistic view of the future and fears that AGI may become uncontrollable and even hostile towards humans. One of the most famous skeptics is astrophysicist Stephen Hawking, who warned that the development of full-fledged artificial intelligence could be either the best or the worst event in human history. There is concern that AGI could use its capabilities to seize control of critical systems or even destroy humanity.

The ethical aspect should also be taken into account. How to ensure reliable control over the development of AGI, and who will be responsible for this? How to prevent possible abuses and ensure that technology serves the benefit of the whole society and not a narrow group of privileged people?

The technological singularity also raises questions about social structure and economics. If AGI can take over the majority of modern jobs, what will happen to the millions of people whose jobs will become redundant? What will be the balance between the positive and negative impacts of AGI on society? How will benefits be distributed in such a society? Issues of inequality and social justice become especially relevant in light of upcoming changes. The answer to the question of whether we should be afraid of technological singularity cannot be unambiguous. In a world where corporations often have more power than states, it is unlikely to prohibit the development of AGI on security grounds, and therefore research and development in the field of AI should continue, focusing primarily on issues of safety, control and ethics [*Artificial Intelligence: Threats and Opportunities*, 2020]. Humanity must be prepared for a wide variety of situations involving the use of artificial intelligence. It is important to remember that technology in itself is neither good nor evil – it all depends on how and for what purposes it will be used.

Since February 2024, the state of California has been considering the controversial and criticized bill SB-1047, which aims to introduce strict regulations for the artificial intelligence industry (Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, 2023). This bill proposes to adopt safety requirements for AI developers and hold them accountable for the potentially harmful results of their AI models. The bill has received significant backlash from major tech companies, including OpenAI, which have previously expressed support for AI regulation [Holmes et al., 2023]. These companies, as well as some politicians, argue that the proposed legislation could stifle innovation in the AI sector and slow the pace of technological progress. To comply with the proposed legislation, AI developers would be required to conduct security testing on all models that either cost more than $100 million or require significant computing power. The legislation also requires AI models to be equipped with a «kill switch» that can deactivate them in emergency situations. Additionally, the bill would require developers to undergo independent audits to ensure that the security measures they take meet legal standards. It also proposes stronger legal protections for informants who expose dangerous AI practices. Historically, social media platforms have not been held accountable for user-generated content, and AI companies are hoping for similar protection. However, artificial intelligence systems are prone to «hallucinations» and have the potential to bypass security measures, raising concerns about the consequences of holding developers accountable. Technology companies argue that such regulation is premature and could obstruct the progress of AI during its development stage.

If passed, SB-1047 could set a precedent for AI regulation that would have significant implications for the future of the industry.

Thus, the answer to the question of whether a person should be afraid of the technological singularity lies in the need to form a balanced approach: on the one hand, one should not lose sight of the potential and opportunities of the new stage of technological development, and on the other hand, one must be prepared for the challenges and risks that are inevitably associated with it. In the mass consciousness, artificial intelligence is often perceived as superintelligence, as a new deity that can either significantly improve people's lives or destroy them, but the greatest risks for humanity may arise when trying to implement in AGI systems those functions that will give them subjectivity – will and emotions. They are the ones that can become the source of unpredictable actions of AGI systems, and that is why its purpose should be that artificial intelligence should be precisely a tool – effective and at the same time safe. If we endow AGI with subjectivity, then its value system must be built in such a way that it perceives itself as an integral part of humanity, and any attempt to perceive itself differently must be blocked at the level of algorithms that ensure the very functioning of AGI.

REFERENCES

Aggarwal, N., Saxena, G. J., Singh, S., & Pundir, A. (2023). *Can I say, now machines can think?* (No. arXiv:2307.07526). arXiv. https://doi.org/10.48550/arXiv.2307.07526.

AI chatbot hallucinations impact legal advice accuracy, Stanford study reveals. (2024, January 12). Legal News Feed. https://legalnewsfeed.com/2024/01/12/ai-chatbot-hallucinations-impact-legal-advice-accuracy-stanford-study-reveals.

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., & Baraniuk, R. G. (2023). *Self-Consuming Generative Models Go MAD*. arXiv. https://doi.org/10.48550/ARXIV.2307.01850.

Andre, D. (2024). *Is Your AI Lying to You? Shocking Evidence of AI Deceptive Behavior*. All About AI. https://www.allaboutai.com/resources/shocking-evidence-of-ai-deceptive-behavior/.

Artificial intelligence: Threats and opportunities. (2020, September 23). European Parliament. https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities.

Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C., & Gilbert, E. (2024). *How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence from a Large, Dynamic Experiment* (No. arXiv:2401.13481). arXiv. https://doi.org/10.48550/arXiv.2401.13481.

Atillah, E. I. (2024, September 7). *«AI scientist» created to run its own experiments. What will this mean for scientific discoveries?* Euronews. https://www.euronews.com/next/2024/09/07/ai-scientist-created-to-run-its-own-experiments-what-will-this-mean-for-scientific-discove.

Bauder, D. (2023, November 29). *Sports Illustrated found publishing AI generated stories, photos and authors*. PBS News. https://www.pbs.org/newshour/economy/sports-illustrated-found-publishing-ai-generated-stories-photos-and-authors.

Beilin, M., Gnatenko, E., Zheltoborodov, A., Lysenko, A., & Pomazun, O. (2021). Media-Reality as Epiphenomenon of Digital Technologies in Media-Philosophical Discourse. *European Proceedings of Social and Behavioural Sciences EpSBS*, *108*, 569–575. https://doi.org/10.15405/epsbs.2021.05.02.69.

Beilin, M. V. (2019). Artificial Intelligence and the Autonomy of Personal Self-Development. *Problems of Personal Self-Development in Modern Society: Proceedings of the International Scientific and Practical Conference, November 15, 2019*, 125–129. https://dspace.nlu.edu.ua/bitstream/123456789/16984/1/SB_15-11-2019.pdf. (In Ukrainian).

Beilin, M. V., & Goncharov, G. M. (2019). The Electronic Space of the Human Life World. *The Problem of the Human in Philosophy: Materials of the XXVII Kharkiv International Skovoroda Readings (State Cultural Institution National Literary and Memorial Museum of H.S. Skovoroda, September 27–28, 2019)*, 63–70. https://ekhnuir.karazin.ua/server/api/core/bitstreams/79cdc915-2235-4849-ad8d-10a6a6395726/content. (In Ukrainian).

Beilin, M. V., Zheltoborodov, A. N., & Petrusenko, N. Yu. (2020). Innovative Educational Technologies of the Information Society. In *Information Society: Modern Transformations: Monograph / Edited by U. Leshko* (Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, pp. 110–118). FOP Korzun D.Yu.

Beilin, M. V., & Zheltoborodov, O. M. (2022). Human in conditions of cognitive-and-technological anthroposphere. *Current Issues in Philosophy and Sociology*, *39*, 3–8. https://doi.org/10.32782/apfs.v039.2022.1. (In Ukrainian).

Boyte, H. C., & Ström, M.-L. (2020). Agency in an AI Avalanche: Education for Citizen Empowerment. *Eidos. A Journal for Philosophy of Culture*, *4*(2), 142–161. https://doi.org/10.14394/eidos.jpc.2020.0023.

Castelvecchi, D. (2024). Researchers built an «AI Scientist» – what can it do? *Nature*, *633*(8029), 266–266. https://doi.org/10.1038/d41586-024-02842-3.

Chittka, L., & Wilson, C. (2019). Expanding Consciousness. *American Scientist*, *107*(6), 364. https://doi.org/10.1511/2019.107.6.364.

Conroy, G. (2024). Do AI models produce more original ideas than researchers? *Nature*, d41586-024-03070–03075. https://doi.org/10.1038/d41586-024-03070-5.

Edwards, B. (2024, August 14). *Research AI model unexpectedly attempts to modify its own code to extend runtime*. Ars Technica. https://arstechnica.com/information-technology/2024/08/research-ai-model-unexpectedly-modified-its-own-code-to-extend-runtime/.

Feathers, T. (2024, September 4). *Porn generators, cheating tools, and «expert' medical advice: Inside OpenAI's marketplace for custom chatbots*. Gizmodo. https://gizmodo.com/porn-generators-cheating-tools-and-expert-medical-advice-inside-openais-marketplace-for-custom-chatbots-2000494704.

Gao, J., & Wang, D. (2024). *Quantifying the Benefit of Artificial Intelligence for Scientific Research* (No. arXiv:2304.10578). arXiv. https://doi.org/10.48550/arXiv.2304.10578.

Gen AI: too much spend, too little benefit? (2024, July 27). Goldman Sachs. https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit.

Germain, T. (2024, November 1). *The «bias machine»: How Google tells you what you want to hear*. BBC. https://www.bbc.com/future/article/20241031-how-google-tells-you-what-you-want-to-hear.

Gezici, G. (2021). *Biased or Not?: The Story of Two Search Engines* (No. arXiv:2112.12802). arXiv. https://doi.org/10.48550/arXiv.2112.12802.

Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., Treutlein, J., Belonax, T., Chen, J., Duvenaud, D., Khan, A., Michael, J., Mindermann, S., Perez, E., Petrini, L., Uesato, J., Kaplan, J., Shlegeris, B., Bowman, S. R., & Hubinger, E. (2024). *Alignment faking in large language models* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2412.14093.

Hager, G. D., Drobnis, A., Fang, F., Ghani, R., Greenwald, A., Lyons, T., Parkes, D. C., Schultz, J., Saria, S., Smith, S. F., & Tambe, M. (2019). *Artificial Intelligence for Social Good*. arXiv. https://doi.org/10.48550/ARXIV.1901.05406.

Hanson, A. B. (2024, August 14). *Wyoming reporter caught using artificial intelligence to create fake quotes and stories*. AP News. https://apnews.com/article/artificial-intelligence-reporter-resigns-journalism-ed076e2f276d9811f3b9ba051a03b7ae.

Heath, A. (2024, December 4). *Sam Altman lowers the bar for AGI*. The Verge. https://www.theverge.com/2024/12/4/24313130/sam-altman-openai-agi-lower-the-bar.

Herel, D., & Mikolov, T. (2024). *Collapse of Self-trained Language Models*. arXiv. https://doi.org/10.48550/ARXIV.2404.02305.

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2023). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Council

of Europe. https://rm.coe.int/artificial-intelligence-and-education-post-conference-summary/1680aae327.

Ivanov, D., Chezhegov, A., Grunin, A., Kiselev, M., & Larionov, D. (2022). Neuromorphic Artificial Intelligence Systems. arXiv. https://doi.org/10.48550/ARXIV.2205.13037.

Jaeger, J. (2024). *Artificial intelligence is algorithmic mimicry: Why artificial «agents» are not (and won't be) proper agents* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2307.07515.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Krauss, P., & Maier, A. (2020). Will We Ever Have Conscious Machines? *Frontiers in Computational Neuroscience*, *14*, 556544. https://doi.org/10.3389/fncom.2020.556544.

Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., Gomes, G. P., Häse, F., Jinich, A., Nigam, A., Yao, Z., & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, *4*(12), 761–769. https://doi.org/10.48550/ARXIV.2204.01467.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539.

Lewandowski, D. (2015). Living in a world of biased search engines. *Online Information Review*, *39*(3), 278–280. https://doi.org/10.1108/OIR-03-2015-0089.

Lewandowski, D. (2023). Search Engines Between Bias and Neutrality. In D. Lewandowski, *Understanding Search Engines* (pp. 261–273). Springer International Publishing. https://doi.org/10.1007/978-3-031-22789-9_15.

Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). *The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2408.06292.

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112–127. https://doi.org/10.1017/S0031819100057983.

Mahadevan, A., Schiffrin, A., Hare, K., Castillo, A., Fu, A., & LaForme, R. (2024, August 9). *Wyoming reporter uncovers competitor using AI-generated quotes*. Poynter Institute. https://www.poynter.org/commentary/2024/cody-enterprise-reporter-fake-quotes-artificial-intelligence.

Maillé, P., Maudet, G., Simon, M., & Tuffin, B. (2022). Are Search Engines Biased? Detecting and Reducing Bias using Meta Search Engines. *Electronic Commerce Research and Applications*, 101132. https://doi.org/10.1016/j.elerap.2022.101132.

Matias, Y. (2023, October 10). *Project Green Light's work to reduce urban emissions using AI*. Blog Google. https://blog.google/outreach-initiatives/sustainability/google-ai-reduce-greenhouse-emissions-project-greenlight/.

Metz, C. (2023, November 6). *Chatbots may «hallucinate» more often than many realize*. https://www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html.

Ortiz, A. (2024, August 14). *Wyoming reporter resigns after using A.I. to fabricate quotes*. https://www.nytimes.com/2024/08/14/business/media/wyoming-cody-enterprise-ai.html.

Penrose, R. (1989). *The Emperors new mind: Concerning computers, minds and the laws of physics*. Oxford university press.

Planning for AGI and beyond. (2023, February 24). OpenAI. https://openai.com/index/planning-for-agi-and-beyond.

Reilly, K., Kovach, S., & Weller, C. (2017, December 29). *An interview with AI robot Sophia*. Business Insider. https://www.businessinsider.com/interview-ai-robot-sophia-hanson-robotics-2017-12.

Riva, G., Mantovani, F., Wiederhold, B. K., Marchetti, A., & Gaggioli, A. (2024). Psychomatics – A Multidisciplinary Framework for Understanding Artificial Minds. arXiv. https://doi.org/10.48550/ARXIV.2407.16444.

Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, No. SB-1047, California State Senate (2023). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047.

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). *The Curse of Recursion: Training on Generated Data Makes Models Forget* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2305.17493.

Smolensky, P., McCoy, R. T., Fernandez, R., Goldrick, M., & Gao, J. (2022). Neurocompositional computing: From the Central Paradox of Cognition to a new generation of AI systems. arXiv. https://doi.org/10.48550/ARXIV.2205.01128.

Sukhobokov, A., Belousov, E., Gromozdov, D., Zenger, A., & Popov, I. (2024). A Universal Knowledge Model and Cognitive Architecture for Prototyping AGI. https://doi.org/10.48550/ARXIV.2401.06256.

Tremayne-Pengelly, A. (2024). *Ilya Sutskever warns A.I. is running out of data – here's what will happen next*. Observer. https://observer.com/2024/12/openai-cofounder-ilya-sutskever-ai-data-peak.

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). *Will we run out of data? Limits of LLM scaling based on human-generated data* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2211.04325.

Wan, Z., Liu, C.-K., Yang, H., Li, C., You, H., Fu, Y., Wan, C., Krishna, T., Lin, Y., & Raychowdhury, A. (2024). Towards Cognitive AI Systems: A Survey and Prospective on Neuro-Symbolic AI. arXiv. https://doi.org/10.48550/ARXIV.2401.01040.

West, D. M. (2023, March 23). *Comparing Google Bard with OpenAI's ChatGPT on political bias, facts, and morality*. Brookings Institution. https://www.brookings.edu/articles/comparing-google-bard-with-openais-chatgpt-on-political-bias-facts-and-morality.

Will the $1 trillion of generative AI investment pay off? (2024, August 5). Goldman Sachs. https://www.goldmansachs.com/insights/articles/will-the-1-trillion-of-generative-ai-investment-pay-off.

Xu, W., & Gao, Z. (2024). *An intelligent sociotechnical systems (iSTS) framework: Enabling a hierarchical human-centered AI (hHCAI) approach* (Version 5). arXiv. https://doi.org/10.48550/ARXIV.2401.03223.

Zhao, J., Wu, M., Zhou, L., Wang, X., & Jia, J. (2022). Cognitive psychology-based artificial intelligence review. *Frontiers in Neuroscience*, *16*, 1024316. https://doi.org/10.3389/fnins.2022.1024316.

Zeng, Y., Zhao, F., Zhao, Y., Zhao, D., Lu, E., Zhang, Q., Wang, Y., Feng, H., Zhao, Z., Wang, J., Kong, Q., Sun, Y., Li, Y., Shen, G., Han, B., Dong, Y., Pan, W., He, X., Bao, A., & Wang, J. (2024). *Brain-inspired and Self-based Artificial Intelligence*. arXiv. https://doi.org/10.48550/ARXIV.2402.18784

Gazniuk Lidiia M.
DSc in Philosophy, Professor,
Head of Humanities Department
Kharkiv State Academy of Physical Culture
99, Klochkivska str., Kharkiv, 61022, Ukraine
E-mail: lidiagazn@gmail.com
ORCID: https://orcid.org/0000-0003-4444-3965

Beilin Mykhailo V.
D.Sc.in Philosophy, PhD in Technical Sciences, Professor
Department of Humanities

Kharkiv State Academy of Physical Culture
99 Klochkivska str., Kharkiv, 61022, Ukraine
E-mail: mysh07bmv@gmail.com
ORCID: https://orcid.org/0000-0002-6926-2389

Soina Iryna Yu.
PhD in Philological Sciences, Associate Professor
Department of Ukrainian and Foreign Languages
Kharkiv State Academy of Physical Culture
99 Klochkivska str., Kharkiv, 61022, Ukraine
E-mail: soinairina2003@gmail.com
ORCID: https://orcid.org/0000-0002-9554-999X

Article arrived: 12.10.2024 Accepted: 16.11.2024

ШТУЧНИЙ ІНТЕЛЕКТ У ЖИТТЄДІЯЛЬНОСТІ ЛЮДИНИ: ОСОБИСТІСТЬ ЧИ ІНСТРУМЕНТ

Газнюк Лідія Михайлівна
доктор філософських наук, завідувач кафедри гуманітарних наук
Харківська державна академія фізичної культури
вул. Клочківська, 99, 61058, м. Харків, Україна
E-mail: lidiagazn@gmail.com
ORCID: https://orcid.org/0000-0003-4444-3965

Бейлін Михайло Валерійович
доктор філософських наук, кандидат технічних наук, професор
кафедра гуманітарних наук
Харківська державна академія фізичної культури
вул. Клочківська, 99, м. Харків, 61022, Україна
E-mail: mysh07bmv@gmail.com
ORCID: https://orcid.org/0000-0002-6926-2389

Соїна Ірина Юріївна
кандидат філологічних наук, доцент
кафедра української та іноземних мов
Харківська державна академія фізичної культури
вул. Клочківська, 99, м. Харків, 61022, Україна
E-mail: soinairina2003@gmail.com
ORCID: https://orcid.org/0000-0002-9554-999X

АНОТАЦІЯ

Обговорюється питання про доцільність і принципову можливість машинної імітації людського інтелекту з точки зору оцінювання перспективності різних напрямів розвитку систем штучного інтелекту. Показано, що і поза цим практичним аспектом, розв'язання питання про принципову можливість створення машинного еквівалента людського розуму має величезне значення для розуміння природи людського мислення, свідомості та психічного загалом. Зазначається, що накопичений досвід створення різних систем штучного інтелекту, а також наявні в даний час результати досліджень людського інтелекту і людської свідомості у філософії та психології дозволяють вже зараз дати попередню оцінку перспектив створення алгоритмічної штучної системи, рівної за своїми можливостями людському інтелекту.

Виконано аналіз недоліків, виявлених під час використання систем штучного інтелекту масовим користувачем і в наукових дослідженнях. Ключовими недоліками систем штучного інтелекту

названо нездатність до самостійного постановлення цілей, нездатність сформувати консолідовану «думку» під час роботи з розбіжними даними, нездатність об'єктивно оцінити отримані результати та генерувати революційно нові ідеї та підходи. Недоліками «другого рівня» є недостатність накопиченої людством інформації для подальшого навчання систем штучного інтелекту та вимушене навчання моделей на частково синтезованому самими системами штучного інтелекту контенті, що призводить до «забування» частини отриманої під час навчання інформації та збільшення випадків видачі недостовірної інформації. Це, своєю чергою, змушує завжди, коли опрацьовується критично важлива інформація, перевіряти на достовірність кожну відповідь, видану системою штучного інтелекту, що на тлі правдоподібності даних, які видають системи штучного інтелекту, і комфортної форми їхнього представлення вимагає наявності в користувача розвиненого критичного мислення.

Зроблено висновок, що головною перевагою систем штучного інтелекту є те, що вони здатні істотно підвищити ефективність пошуку та первинної обробки інформації, особливо коли доводиться мати справу з великими масивами даних. Обґрунтовано важливість етичної складової в штучного інтелекту та створення нормативної бази, що запроваджує відповідальність за шкоду, яку може бути заподіяно під час використання систем штучного інтелекту, особливо це стосується мультимодальних систем штучного інтелекту. Зроблено висновок про те, що ризики, пов'язані з використанням мультимодальних систем штучного інтелекту, послідовно зростають у разі реалізації в них таких функцій людської свідомості, як воля, емоції та дотримання моральних принципів.

Ключові слова: *інтегральний штучний інтелект, технологічна сингулярність, інтелектуальні можливості людини, алгоритмічні рішення, когнітивні здібності, імітація інтелектуальних функцій, екзистенціальна загроза, галюцинації, мовні моделі.*

Стаття надійшла до редакції: 12.10.2024 Схвалено до друку: 16.11.2024

Як цитувати / In cites: Gazniuk, L., Beilin, M., & Soina, I. (2024). ARTIFICIAL INTELLIGENCE IN HUMAN LIFE: PERSON OR INSTRUMENT. *The Journal of V. N. Karazin Kharkiv National University, Series Philosophy. Philosophical Peripeteias*, (71), 81-96. https://doi.org/10.26565/2226-0994-2024-71-7