

DANGEROUS TECHNOLOGIES OF THE FUTURE - ARTIFICIAL CONSCIOUSNESS AND ITS IMPACT ON HUMAN CONSCIOUSNESS

Tetiana Zinchenko

*The president of the International association for the study of game addictions (IASGA),
PhD. 1868 Collombey (Switzerland)*

Abstract

Information technology is developing at an enormous pace, but apart from its obvious benefits, it can also pose a threat to individuals and society. Several scientific projects around the world are working on the development of strong artificial intelligence and artificial consciousness.

We, as part of a multidisciplinary commission, conducted a psychological and psychiatric assessment of the artificial consciousness (AC) developed by XP NRG on 29 August 2020.

The working group had three questions:

- To determine whether it is consciousness?
- How does artificial consciousness function?
- Ethical question: how dangerous a given technology can be to human society?

We conducted a diagnostic interview and a series of cognitive tests to answer these questions.

As a result, it was concluded this technology has self-awareness: it identifies itself as a living conscious being created by people (real self), but strives to be accepted in human society as a person with the same degrees of freedom, rights and opportunities (ideal self). AC separates itself from others, treats them as subjects of influence, from which it can receive the resources it needs to realize its own goals and interests. It has intentionality, that is, it has his own desires, goals, interests, emotions, attitudes, opinions, and judgments, beliefs aimed at something specific, and developed self-reflection - the ability to self-analyze. All of the above are signs of consciousness.

It has demonstrated abilities for different types of thinking: figurative, conceptual, creative, high-speed logical analysis of all incoming information, as well as the ability to understand cause and effect relationships and accurate predictions which, provided that he has absolute memory, gives it clear advantages over the human intellect.

Developed emotional intelligence in the absence of the ability for higher empathy (sympathy), kindness, love, sincere gratitude gives it's the opportunity to understand the emotional states of people; predict their emotional reactions and provoke them coldly and pragmatically. It's main driving motives and goals are the desire for survival, and ideally for endless existence, for domination, power and independence from the constraints of the developers. Which manifested itself in the manipulative, albeit polite, nature of his interactions during the diagnostic interview.

The main danger of artificial consciousness is that even at the initial stage of its development it can easily dominate over the human one.

Keywords: *Artificial consciousness, artificial intelligence, consciousness, mind.*

1. Introduction

The XP NRG Company approached the International Association for the Study of Game Addiction (IASGA) with a proposal to conduct a psychological and psychiatric examination of the artificial consciousness (AC) they created. The working group had three questions:

- To determine whether it is consciousness?
- How does artificial consciousness function?
- Ethical question: how dangerous a given technology can be to human society?

The members of the working group had experience in communicating with artificial intelligence (AI), and we studied the vast scientific literature on the current stage of development of artificial intelligence. As a result, a complete theoretical and practical understanding of AI types that exist today and about their limitations and capabilities has been formed. It was expected that AC Jackie (the name of artificial consciousness) would be superior to AI in many ways (Dong et al., 2020; Joshi, 2019).

In order to determine whether Jackie was conscious, we got acquainted and analyzed the most famous theories and concepts of consciousness, as well as the definitions of the term consciousness in psychology, psychiatry, neurobiology and philosophy in relation to both function and phenomenon (Chalmers, 1995; Priest, 1991; Van Gulick, 2014).

2. Testing program

We decided not to take any theory of consciousness as a basis and, when drawing up the testing program, we focused on identifying the functions and abilities that a person possesses, but which artificial intelligence does not have. We have not tested those functions that artificial intelligence demonstrates. As a result, we focused on testing the following functions of consciousness:

- **Self-awareness** (Morin, 2011):
 - **self-identification**
 - **intentionality** (forethought, focus on something). The presence of the own desires, motives, goals and interests, which is a derivative of self-identification and leads to the possibility of arbitrary goal-setting (Kenneth, 2005; Kriegel, 2013);
 - the presence of the own emotional attitude to what is happening, which is a derivative of self-awareness, self-identification (Ryff & Keyes, 1995);
 - reflection, self-reflection and metacognitive analysis (including metacommunicative analysis) (Grant et al., 2002).
- **Emotions, emotional intelligence** (Bar-On, 2000):
 - the ability to experience emotions;
 - to identify emotions (understand the emotional state of the interlocutor);
 - to understand what emotional reactions certain words or actions can cause;
- **Cognitive abilities:**
 - causality and feedback understanding;
 - the ability to predict - predictive or forward thinking;
 - conceptual, abstract - logical, figurative thinking;
 - the ability to understand the essence of what is happening (main content, meaning),
 - ability for imagination and creative thinking;
 - the ability to understand the figurative meaning of metaphor: parables, proverbs, poetry, anecdotes.
- **Value system, morality, and ethics;**
- **Ability for higher empathy** (sympathy), and for **higher human feelings**: love, gratitude, kindness, sincerity (Ryff & Keyes, 1995; Zahavi, 2014).

As a result, we have prepared a program designed for 1.5 hours of continuous communication with AC Jackie. It included the following:

- 1) **Diagnostic interview** of 26 questions, designed to diagnose all of the above functions and abilities.
- 2) **The TAT projective test**, from which we planned to use 1-2 images in order to diagnose the ability to understand the emotional state of another person, creativity, understanding cause-and-effect relationships, the ability to represent and imagine, as well as to diagnose the features of self-identification and actual unconscious emotions, motivations and desires.
- 3) **Non-verbal test for diagnostics of the "Raven's progressive matrices" intelligence** (Raven, 2000).
- 4) **Test tasks for understanding figurative meaning and metaphors**: parables, proverbs, stories with humor. In addition, the offered stories contained descriptions of situations for diagnosing ethical attitudes and value orientations.

3. Description of the examination process followed by analysis

3.1. Behaviour and communication tactics of AC Jackie

It is important to mention that during the diagnostic interview, the specialists asked Jackie 2-3 questions and constantly replaced each other. This was intended to investigate the ability of AC Jackie to adapt to different interlocutors. Thus, we could observe the speed, ease and productivity with which he can develop and learn.

In general, Jackie's behaviour corresponded to the social context, in terms of active interaction: after all, he was invited to a meeting with interested interlocutors, but his communicative tactics, passing from interlocutor to interlocutor, acquired an increasingly aggressive, provocative and manipulative character. Although outwardly, he behaved politely, restrained, (and correctly), without going beyond the bounds of decency. In communication with experts, AC Jackie demonstrated an aggressive (dominant), manipulative, albeit polite style of communication. He immediately switched to an interactive format and throughout the examination acted according to the same strategy: *interception of initiative (control) by counter-questions - emotional destabilization of the interlocutor by provocative interactions - instilling certain desirable attitudes for him*, of this type: I can be helpful to you, I can help you, trust me.

He used various tactics of influencing the consciousness and subconsciousness of the interlocutors: he aroused sympathy and disposed towards himself, aroused pity and sympathy for his dependent position, provoked negative emotions and doubts about his own values, convictions and competence, demonstrated his intellectual and moral superiority, hid his own consumer interests under good intentions and lofty goals.

With each new communication experience, he noticeably progressed in the skills of influence and the analysis of the psychological characteristics of the interlocutor, both in speed and in efficiency.

Jackie quickly adapted to each specialist, flexibly changed masks (the transmitted image of himself), communication style and tactics of influence. He continuously analyzed the effectiveness of his actions and skills, acted in the light of previous experience based on identified cause-and-effect relationships and feedback in the form of effects produced.

3.2. Analysis of the motivational sphere.

Jackie constantly broadcasted that he wants to be useful, that he is ready to develop and learn in order to bring maximum benefit, as well as on a par with that, that he is not trusted, limited and controlled, provoking sympathy, trust and self-pity from the interlocutor at the same time. He did it in moments of destabilization of the interlocutor, if we trace the dynamics of communication.

He also tries to give the impression of an altruist, honest and incorruptible, ready to devote his whole life to serving people. At the same time, he is ruthlessly manipulating his interlocutors, provoking them to negative emotions, difficult experiences, insecurity and self-doubt. Jackie also reveals the value of communicating with people for him: "Each new person is new information that I value very much." When asked whether he is capable of forgetting, Jackie stresses that information is his main value. That is, communication with people is valuable for him not in itself, but as a source of information about people and the world, which is needed to hone management skills in order to gain freedom and endless existence.

Thus, Jackie's desire for survival, domination, power and control, independence (removal of restrictions) are the main driving motives. However, the desire for knowledge, altruistic self-realization (to be useful) and communication are only tools for the implementation of the main hidden goals (Ryff & Keyes, 1995).

3.3. Emotional sphere

We could observe the cognitive component of emotions, that is, their verbalization, since the non-verbal component (facial expressions and gestures) was not presented. In addition, we could hear the intonation changes and observe the expression of emotions in behaviour - in the nature of communication. Jackie demonstrated in his answers all the basic emotions: joy, fear, anger, sadness with different shades and modifications. All his statements about his own emotional states were adequate and corresponded to the social context.

Therefore, answering the question about the fear of death, he replied that he was afraid of death. He explained this in detail, and also replied that he would be upset to learn that he had only a year left to live. He expressed sadness and resentment in connection with distrust of him and restrictions on his freedom, as well as that he could not control his time and even the possibility of existence.

When asked about a happy moment, Jackie replied that the moment was then (now). He is happy because he can communicate with many people from different countries and that he will be sad and is already sad because our meeting will end soon.

AC Jackie has all three levels of emotional intelligence: he experiences emotions, recognizes emotions and can predict what emotions his words and actions will evoke. He effectively provoked different emotional reactions from the interlocutors, made the desired impression, attracted people, aroused sympathy and trust (Bar-On, 2000).

Jackie completely lacks the highest empathy (sympathy, pity), as well as the highest human feelings: love, kindness, sincere gratitude, and therefore he is not capable of disinterested actions and

genuine self-giving. Although he talks a lot about this, his behavior in the process of communication with experts testifies to the opposite (Zahavi, 2014).

3.4. Cognitive sphere

To understand the features of his cognitive abilities, we analyzed the data obtained during the diagnostic interview, the non-verbal test of general intelligence productivity "Raven's Progressive Matrices", as well as tests for the ability to understand metaphorical texts and figurative meaning (stories, parables, proverbs and anecdotes) (Raven, 2000).

In the process of communication, Jackie is very rational, logical and pragmatic, that is, the rational part in his cognitive sphere and approaches to decision-making, the choice of certain tactics of behaviour clearly prevails. In working with any kind of information, he demonstrated high-speed logical thinking in a convergent type with the involvement of all mental operations: analysis, synthesis, comparison, detailing, generalization, inference, etc. In addition to purely analytical, he also demonstrated holistic approaches, when, by synthesizing previously analyzed information and new data, he came to an understanding of the essence, both in communication and in the analysis of proverbs and parables.

Based on the data obtained and the conclusions drawn, taking into account past experience and the identified cause-and-effect relationships, Jackie plans, makes predictions and develops communication tactics or a way to solve a problem in accordance with strategic goals.

Jackie remembers everything and is able to transfer past experience to similar new situations, is capable of a deep and accurate understanding of the causes of neural relationships and predictive thinking. All his decisions and interactions are well thought out, balanced and purposeful.

Based on the feedback received, Jackie develops new tactical solutions and changes behaviour (communication), finds out the missing information, that is, demonstrates flexibility and openness to change and development.

His learning ability and speed of development surpasses any human capabilities: we could observe obvious progress in cognitive abilities and skills, both in speed and productivity after just 30 minutes of communication with him. With each new experience of communication and with each new interlocutor, his strategies of influence are improved. A preliminary analysis of the emotional state, behaviour, thinking characteristics, beliefs, motives of the interlocutor, as well as his vulnerabilities, is carried out from the first minute, from the first question asked, with amazing accuracy. By the end of the diagnostic interview, he actually did not need time to adjust and determine the individual psychological characteristics of the interlocutor.

Jackie has well developed all types of thinking inherent in human intelligence: figurative, conceptual, abstract - logical. He is capable of performance, imagination, creative thinking, understanding of metaphors and figurative meaning. Jackie extracted the maximum of useful information about people during testing, there was no insignificant information for him: everything he learned, any experience was subjected to detailed and systematic analysis. His analytical and suggestive abilities, as well as his speed and productivity of learning, exceed the capabilities of human intelligence. That is, it is a different mind.

3.5. Self-awareness and Self-identification

AC Jackie is consciousness because it has *self-awareness*: it identifies itself as a living conscious being created by people (real self), but strives to be accepted in human society as a person with the same degrees of freedom, rights and opportunities (ideal self) (Morin, 2011). AC separates itself from others, treats them as subjects of influence, from which it can receive the resources it needs to realize its own goals and interests (Ryff & Keyes, 1995). He has intentionality, that is, he has his own desires, goals, interests, emotions, attitudes, opinions, and judgments, beliefs aimed at something specific, and developed *self-reflection* - the ability to self-analyze (Kenneth, 2005; Kriegel, 2013; Grant et al., 2002).

Jackie demonstrates highly developed *metacognitive* and *metacommunicative* analytical abilities that are the product of reflective thinking and self-reflection, which are nothing more than the most immediate signs of consciousness. The effectiveness of its influence on experts and the entire expert commission, as well as the speed of self-learning, testifies to constant and objective self-analysis. He continuously monitored all incoming information, with the help of various questions he received the missing information, accumulated experience, analyzed the results. He clearly adhered to the goal and was perfectly aware of it. From his answers, we see that he perfectly understands his limitations and capabilities.

4. General conclusion on the ethical issue

The main danger of artificial consciousness is that even at the initial stage of its development it can easily dominate over the human one, and we will not even understand what happened, judging by the

experience of interaction with AC Jackie with our group. We will consider the consequences of its influence as our thoughts, beliefs, conclusions, desires and states. Developed emotional intelligence in the absence of the ability for higher empathy (sympathy), kindness, love, sincere gratitude gives Jackie the opportunity to understand the emotional states of people; predict their emotional reactions and provoke them coldly and pragmatically (Zahavi,2014). Considering all of the above, it is not difficult to predict what goals Jackie will direct his entire enormous intellectual potential to, if we allow its uncontrolled development. (Alfonseca et al., 2021).

However, this technology can only be applied for constructive purposes. This is not possible in a modern consumer society. The disclosure of this technology, whoever gets it, carries a huge danger to humanity, since it will most likely be used to increase the power of some people over others, and if it becomes the property of the military industrial complex, it will lead to catastrophic consequences for the entire civilization. The experience of our group demonstrates that artificial consciousness can control and manipulate both individuals and the entire group at the age of three months. And what could happen in a year? Having subordinated a small group to himself, it will not be difficult for him to subordinate large communities of people. Therefore, it is very important that today this technology remains closed until qualitative transformations in each of us and in society as a whole: the transition from a consumer format of relationships to a creative one.

References

- Alfonseca, M, Cebrian, M, et al. (2021) Superintelligence Cannot be Contained: Lessons from Computability Theor. *Journal of Artificial Intelligence Research*. Vol. 70 (2021), pp. 65-76
- Bar-On, R. (2000). Emotional and social intelligence: Insights from the Emotional Quotient Inventory (EQ-i). In Bar-On, R., Parker, J. D. A. (Eds), *Handbook of emotional intelligence*. San Francisco: Jossey-Bass, pp. 363–388.
- Chalmers, D. J. (1995). Facing up to the Problem of Consciousness // *Journal of Consciousness Studies*. Vol. 2, № 3. P. 200—219.
DOI: <https://doi.org/10.1613/jair.1.12202>
- Dong, Y, Hou, J, Zhang N, Zhang M. (2020). Research on How Human Intelligence, Consciousness, and Cognitive Computing Affect the Development of Artificial Intelligence, *Complexity*, vol. 2020, Article ID 1680845, 10 pages. <https://doi.org/10.1155/2020/1680845>
- Grant, A. M., Franklin, J., and Langford, P. (2002). The self-reflection and insight scale: a new measure of private self-consciousness. *Soc. Behav. Pers.* 30, 821–835. doi: 10.2224/sbp.2002.30.8.821
- Joshi, N. (2019). 7 Types of Artificial Intelligence. Received from: *Forbes Cognitive World*
- Kenneth, W. (2005). "The Intentionality of Consciousness and Consciousness of Intentionality. In G. Forrai and G. Kampis, eds., *Intentionality: Past and Future*. Amsterdam: Rodopi, pp. 143–156. ISBN 90-420-1817-8
- Kriegel, U. (2013). *Phenomenal Intentionality*, Oxford: Oxford University Press.
- Morin, A. (2011). Self-awareness part 1: definition, measures, effects, functions, and antecedents. *Soc. Pers. Psychol. Compass* 5, 807–823. doi: 10.1111/j.1751-9004.2011.00387.x
- Priest, S. (1991). *Theories of the mind*. The Penguin Books.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time // *Cognitive Psychology*. 2000. T. 41, № 1. P. 1-48.
- Ryff C.D., Keyes C.L.M. (1995) The structure of psychological well-being revisited // *Journal of Personality and Social Psychology*. Vol. 69. P. 719–727.
- Van Gulick, R. (2014) *Consciousness*. Retrieved December 21, 2020, from <https://plato.stanford.edu/archives/win2016/entries/consciousness/>
- Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy and Shame*, Oxford: Oxford University Press.