Is the Realization of the Emotional Artificial Intelligence Possible? Philosophical and Methodological Analysis

OKSANA CHURSINOVA, OLEKSANDRA STEBELSKA

Department of Philosophy, Lviv Polytechnic National University, 5 Mytropolyt Andrei St., Building 4, Room 328, Lviv, 79013, Ukraine
Email: churss@ukr.net; cleanwave4@gmail.com

The article dwells upon the need for a thorough philosophical and methodological analysis of the nature and functions of the human's emotional and sensual sphere in order to identify the possibilities of its implementation by means of artificial intelligence. Computers have become part and parcel of our lives, so full-fledged communication requires empowering them to recognize and express emotions. Due to the result of critical analysis, the authors state that implementation, and not simulation, of emotions in any computing system is currently problematic and, to some extent, impossible. The reason for this is connected both with the blurring in the scientific and philosophical literature of the very concept of 'emotion', and the subjective and qualitative nature of the person's experience of reality, the rootedness of their emotional and sensual sphere in the physical, social and cultural being, the unconditional connection of emotions and the internal personal space of the person.

Keywords: emotions, emotional intelligence, emotional artificial intelligence, qualia, morality

INTRODUCTION

Modern research has shown that emotions are able to determine the direction of our thoughts, influence the depth of our worldview and decision-making (Lerner et al. 2015: 799–823). Our intellect cannot fully function outside of emotional and volitional factors. Not only does the emotional sphere of human beings influence the productivity of our intellectual processes, but it also promotes better adaptation to the external environment, facilitates and deepens communication between people, performs a motivational role and is an additional form of assessment of reality. K. Oatley specifies that '...emotions are most typically caused by evaluations... of events in relation to what is important to us: our goals, our concerns, our aspirations' (Oatley 2008: 3). Since in the modern world machines have penetrated into all spheres of human activity, the issue of creating such machines that could feel, experience, understand a person, engage with them in full communication, help them, support in difficult situations is extremely urgent. The communicative process involves empathy as an indispensable

element. Therefore, the attention of scientists has focused not only on accomplishing the human intellectual abilities in an artificial environment, but also on creating an AI that can feel and experience reality. For a more correct treatment of the indicated problems, the authors use the concept of 'emotional artificial intelligence' (EAI).

So, sufficient communication practices require the machine's ability to recognize and respond adequately to human emotions and feelings. Complete human/machine communication is possible only if the machine not only performs certain algorithms and operations, but also has needs and desires, independent assessment of situations and response to them. Besides, the attitude towards the nature of the emotional and volitional sphere of human has substantially changed. Previously, emotions were opposed to intelligence, but now scientists have come to the conclusion that the effective functioning of human intelligence is impossible beyond emotional processes. A. Damasio experimentally demonstrated that damage to the frontal lobe of the brain, which is responsible for the emergence of emotions, causes impaired decision-making processes (Damasio 1994). The level of intellectual ability of a person with such a defect is not reduced, he or she can clearly identify all the pros and cons of a situation, but they will have difficulty in making choice. This presupposes that the effectiveness of human activity depends on both intellectual and emotional components that complement each other. Although not all researchers agree that emotions in some way contribute to intellectual activity, however, there is no point in denying the role of emotions in our everyday lives (Sloman 2001: 177–198).

In modern discourse, the term 'emotional intelligence' is becoming more and more common (Goleman 1995; Mayer, Salovey 1993: 433–442). The level of emotional intelligence indicates how well an individual is able to understand their own and others' emotions, correctly interpret and respond to people's behaviour, manage their emotions, or adjust them for better adaptation to the external environment (Goleman 1995). '"...Emotional intelligence" – a form of intelligence that is different from but connected to reason or "rational intelligence" – which affirmed that the rational mind cannot work effectively without the emotional reasoning and that the ability to care and empathize is necessary for ethical reasoning or practice' (Lu et al. 2020: 24–33). Accordingly, such reflections and research by scholars have provoked attempts to realize the emotional component in the machine environment.

Issues related to the nature of intelligence and emotions have become particularly relevant. There has been a debate about the correctness of the concept of emotional intelligence (intelligence vs emotions), whether it is advisable to talk about emotional intelligence if there are other types of intelligence, etc. (Salovey 1993: 433–442). Omitting the answers to these questions when creating an emotional AI has proved impossible, because only understanding who we are will enable us to create a similar being. Accordingly, the problems noted by A. Turing in 1950 turned out to be much more complicated. If initially the thinker wondered about creating a machine whose actions would be similar to human, now the question arose of creating a machine that would be able to engage in full communication with humans.

Consequently, scientists attempt to reproduce the emotional sphere of a human in an artificial environment and try to create emotional artificial intelligence. Currently, there are various options for the implementation of the emotional sphere of a man, namely, OCC, ActAffAct, FLAME, EMA, ParleE, FearNot!, FAtiMA, WASABI, Cathexis, KARO, MAMID, FCM, and xEmotion, CogAff, Affective Computing, etc. In particular, OCC (Ortony et al.) is an approach that explores 22 emotions, their qualitative and quantitative parameters (Ortony et al. 1988; Steunebrink et al. 2008: 256–260). By focusing on the qualitative aspects of

emotions, the OCC identifies the conditions for the occurrence of emotions. Quantitative analysis of emotions involves the fixation of the intensity of emotional life, the factors that influence this intensity and the explanation of the relationship between the conditions of the occurrence of emotions and their intensity. KARO uses the principles of this approach and is based on it (Steunebrink et al. 2008: 256–260; Meyer 2006: 601–619). The Affective Computing approach, founded by R. Picard, is given much prominence (Picard 2003: 213–235). The essence of this approach is to recognize the emotions and feelings of a person by their facial expressions and behaviour. Much attention is paid to the ability of the robot to show emotional reactions. Today, we are not interested in the technical side of the embodiment of the emotional component, but in assessing the prospects of creating emotional artificial intelligence and ontological problems that scientists may face. Despite some success in creating machines capable of expressing simple emotions, it must be acknowledged that further research will face a number of extremely complex problems.

ISSUES IN IMPLEMENTATION OF EMOTIONAL ARTIFICIAL INTELLIGENCE

Having begun to embody the human emotional component in a machine environment, scientists were faced with the need to clearly define the nature of emotions, their structure, functions, features and conditions of occurrence. The Stanford Encyclopedia states that all studies of the nature of emotions can be divided into three main interpretations of the concept of emotions (Scarantino, de Sousa 2018): emotions as feelings, emotions as evaluation and emotions as motivation. The tradition of understanding emotions as feelings defines emotions as conscious experiences. One of the well-known theories of this tradition is the James–Lange emotion theory (James 1884). The tradition of treating emotions as evaluations binds them to reality and defines them as criteria for its evaluation. The tradition of defining emotions as motivation is linked to motivational states. Each of these traditions of understanding the nature of emotions has its disadvantages. In particular, the first tradition is criticized for over-reducing emotions to physiological manifestations. After all, the bodily reactions of a person may be quite similar, but they may indicate different emotional states. In addition, people's behavioural reactions to certain events may differ, which indicates the complexity of such a phenomenon as emotion. The tradition to understand emotion as an evaluation is related to the formulation of certain judgments. However, judgments do not explain the motivational aspects of emotions and they do not reveal how they affect our behaviour. Moreover, the evaluation does not acknowledge the interconnection between the judgments and the emotions that may contradict them. One of the drawbacks of the tradition of considering emotions as motivation is the fact that not all emotions encourage us to certain acts and changes. Besides, the same emotion may manifest in different behavioural responses and may be difficult to associate with a specific program of action.

The variety of approaches to understanding the nature of emotions testifies to their limitedness and incomplete nature of the problem. The nature of emotions and their origin are still uncertain and need further investigation. Therefore, the first problem that scientists face in trying to realize emotional artificial intelligence is the blurring of the very concept of 'emotion' and only a partial understanding of the very process of their emergence. The question is as follows: How can one embody something, the nature of which we do not fully know? The same situation is observed in the field of intelligent machine creation. After all, to create them you need to define the term 'intellect', the content of which is still ambiguous (Suvorov 2010: 127–128). The term 'artificial intelligence' itself has not become clear either (Russel, Norvig 2006).

Accordingly, as long as scientists do not discover the core of human intellectual and emotional properties, all attempts to give them to a car will have disappointing results. Moreover, we may encounter a terrifying creature that is fully neither intellectual nor emotional.

Another problem for scientists is that the nature of emotions is closely intertwined with the realm of the unconscious. In his work, A. Damasio clearly identifies the unconscious roots of the formation of our emotional states (Damasio 1999). He relates the emergence of emotions to neurophysiological processes and notes that not all of our experiences can be conscious. 'Neither the state of feelings nor the emotions that gave rise to it were "conscious" and yet they unfolded as biological processes' (Damasio 1999: 36). Damasio notes that the limbic and prefrontal areas are responsible for the functioning of emotions (Damasio 1994). The question is the following: If the emergence of emotions is a process closely linked to the activity of our brain and far from always conscious, then how can we implement it in a machine environment? To do this, it is necessary to completely reproduce our higher nervous system artificially, which is currently an unattainable goal.

In addition, our emotions are tied not only to brain activity, but also to our physicality in general, which hinders their implementation in an artificial environment. 'Many of the human reactive mechanisms and some of their motivators and emotional responses are closely linked to bodily mechanisms and functions. For example, if you do not have a body, you will never accidentally step on an unstable rock, and you will not need an 'alarm' mechanism that detects that you are about to lose your balance and triggers corrective action, including causing a surge of adrenalin to be pumped around your body' (Sloman 2001: 177–198). This statement applies to more virtual creatures, but it captures well the problem of the physicality of the AI itself. The robots will not possess the identical bodily architecture, and therefore, their perception of reality will be significantly different from ours, the grasp of spatial structure and movement will be unattainable for them.

It is worth recalling that man is a being rooted in the structure of the Universe (the anthropic principle) and therefore all human potentials and opportunities are respectively rooted in the structure of reality and tailored to understand it as deeply and fully as possible. AI will be deprived of such unity with natural and spatial processes, and therefore its perception of the world will be marked by one-sidedness and incompleteness.

QUALIA PROBLEM AND EMOTIONAL INTELLIGENCE

The problem of realization of the emotional sphere in AI is inseparable from the philosophical problem of qualia. Qualia are subjective experiences of a person inherent in him or her and closely related to the perception and sense of reality at the individual level. This subjective experience is difficult to transmit or pass on to another person, and often it may even be hard to detect with scientific tools. The nature of qualia is still a problematic phenomenon (Sloman 2001: 177–198). Such thinkers as T. Nagel, J. Searle and D. Chalmers acknowledge their existence, while evolutionary philosophers like D. Dennett and P. Churchland do not take qualia for the real fact and refer them to the manifestations of folk psychology. The authors adhere to a tradition that considers qualia as a necessary component of a conscious subjective human experience. Scientific research and experiments certainly point to those processes that are responsible for the neurophysiological manifestations of our conscious lives, but can tell us nothing about the individuality of perception and the meanings by which we construct our own reality.

Emotions are also related to our personal attitude to the environment which we belong to. We feel pain, happiness, love, hate and anger in a different way. No doubt, their general manifestations are the same, but the internal feelings, their intensity and dynamics are different. If scholars are still unable to provide a clear answer to questions about the mechanisms of this subjective experience, then how can we implement it? The one which is currently being implemented in computer programs is purely formal. Artificial creatures are not capable of feeling and experiencing the world as humans do. They may recognize emotions and respond to them more or less adequately, but this has nothing to do with their personal inner space, since they lack it. R. Picard though states that 'Machines already have some mechanisms that implement (in part) the functions implemented by the human emotional system... But computers do not have human-like emotions in any rich or experiential natural sense... Computers may have mechanisms that imitate some of ours, but this is only in part, especially because our bodies differ and because so little is known about human emotions' (Picard 2003: 213–235).

SOCIO-CULTURAL AND MORAL ASPECTS OF EMOTIONAL ARTIFICAL INTELLIGENCE REALIZATION

The complexity of embodying emotions in artificial devices is related to the immersion of emotions in the social and cultural space. They help us not only to adapt to the environment, but also to develop certain behavioural strategies and make choices, predictions and self-regulation, volitional control, etc. Not only are we bodily experiencing reality in practice that distinguishes our constitution from the machine, but some particular forms of emotional expressions are formed in each social and cultural space. P. Ekman, exploring the nature of emotions, identified seven basic emotions that manifest in all peoples, namely, joy, wonder, anger, disgust, contempt and fear (Ekman, Friesen 1975). However, even they differ in the nature of their manifestation. For example, the Japanese will never express their feelings as flamboyantly as the Italians. Accordingly, the emotional reactions of a person in a socio-cultural environment arise extremely complex, multifaceted and contextual, which makes it impossible to fully implement them in computer programs at the present time. How will the machine be guided while responding to human behaviour: emotionally or intellectually? After all, a person finds him or herself in a variety of situations throughout the life, sometimes extremely dramatic. Will AI be capable of empathy and compassion, support and simple sympathy? Human communication is not only about calculations and foreseeing. Therefore, all attempts to implement emotional AI are now reduced to the simple recognition of human emotions and the most appropriate responses to them. However, even the attempts to endow artificial intelligence with the appropriate responding ability, showed clearly its limitations of acquiring this ability. The question is the following: How often do we respond adequately to other people's behaviour and are we capable to adequately recognize their emotional state or interpret their emotional behaviour? There is something spontaneous, unpredictable, imperfect, but purely human in our mistakes.

Another obstacle to the embodiment of emotional artificial intelligence is that our emotions and feelings rarely appear in a pure form, they are usually of a mixed nature. Grief can go together with joy and pleasure can be seen in pain. Will we be able to transmit into our artificial creature the capacity for virtually limitless interpretation and, most importantly, the understanding of the whole variety of human feelings, as in some cases we cannot

distinguish them ourselves. Moreover, our senses are multimodular and can be simultaneously evoked by either visual, vocal or tactile information. Data from different sources cannot only complement each other but also be mutually contradictory. For example, a person may behave calmly, but his or her voice and movements may be alarming. Creating an emotional AI that will be able to adequately analyse and respond appropriately to all those aspects is an issue for further development.

Emotions are multifunctional elements of our consciousness. If we are to create emotional artificial intelligence, then we must realize all the basic functions of consciousness. Their implementation involves the existence of a complete biological organism that is in a close contact with the environment and able to produce various models of interconnection with it on the basis of certain ideal structures. No modern computer has such capabilities. Let us consider the evaluation and compensation functions of emotions. For humans, the evaluation function involves not only the perception of the environment through the prism of 'pleasant/ unpleasant', but also the existence of an ideal dimension of reality and the ability to activate it. Unlike person, the machine is not capable of detecting perfect structures on its own because they are already installed by the programmer. With regard to the compensatory function of emotions, its essence is that emotions are able to compensate the lack of information to make a certain decision. When a person makes important decisions, intelligence can only calculate and mathematically evaluate the possibilities, alternatives and consequences, while emotions will push us to make certain choices. Are we capable of implementing these features into computer? Can the machine make a decision based on insufficient information? So far, machines can work with available information, not with the lack of it.

The problematic character of emotional AI implementation is also due to the fact that emotions are closely intertwined with human beliefs, hopes and expectations (Frijda et al. 2000: 1–9). It is based on emotions that we can predict the situation to some extent and be convinced that it may or may not happen. Our beliefs are also related to information failure. They are paradoxical in nature because, guided by them, one is convinced of the correctness of what is only probable. This aspect is also a significant obstacle to the implementation of AI.

Is it potentially possible to create an emotional intelligence machine in the future? This question has no clear answer. On the one hand, research in the field of EAI implementation is extremely promising. On the other hand, these studies are mainly aimed at imitating our inner world, rather than its full reproduction. Even if we assume that in the future we will still overcome the ontological difficulties of implementing EAI and create an emotional machine, then ethical issues will come to the fore. Emotional artificial intelligence, which will perfectly recognize human emotions, based, in particular, not only on the external manifestation of emotions, but also taking into account internal biological changes in the human body, will pose a threat to human autonomy and privacy. Therefore, not only the problem of the embodiment of emotional artificial intelligence, but also the moral aspect of this issue is extremely relevant (Boyles 2018: 182–200; Celebi 2019: 351–376; Corabi 2017: 128–149; El Kaliouby 2017: 8–9; Etzioni, Etzioni 2017: 403–418; Petrushenko, Chursinova 2019: 199–205; Roff 2019: 127–140).

CONCLUSIONS

Considering the prospects for the further development of research on artificial intelligence, the authors confirm the need for a thorough understanding of human emotions and feelings. It concerns transmitting the emotional properties of a person to a machine. In order to handle

this area of technoscience more correctly, the article uses the concept of 'emotional artificial intelligence' (EAI), which the authors propose to use in the future in the context of this issue. The article proves that rapid progress in modelling emotions today is superficial in nature and it only simulates person's capabilities and the ability to communicate in person. The reason for this is the fundamental differences in computer structures and human consciousness structure. At the same time, the developments in the field of emotional artificial intelligence contribute to a better understanding of our own nature and to understanding the possibilities of embodying the emotional and sensual sphere in a machine environment.

References

1. Arouses, K.; Bringsjord, S. 2014. 'The Philosophical Foundations of Artificial Intelligence', in *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish and W. Ramsey. Cambridge: Cambridge University Press, 34–63.

2. Boyles, R. J. M. 2018. 'A Case for Machine Ethics in Modeling Human-level Intelligent Agents', *Kritike* 12(1): 182–200.

3. Bringsjord, S.; Govindarajulu, N. S. 2018. 'Artificial Intelligence', in *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=artificial-intelligence/ (accessed 12.07.2018).

4. Celebi, V. 2019. 'Discussion on the Limit of Artificial Intelligence in the Context of Criticism of Physicalism in Searle and Nagels Theories of Consciousness', *Beytulhikme: An International Journal of Philosophy* 9(2): 351–376.

5. Corabi, J. 2017. 'Superintelligence as Moral Philosopher', *Journal of Consciousness Studies* 24(5–6): 128–149.

6. Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G. P. Putnam.

7. Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Harcourt Brace.

8. Ekman, P.; Friesen, W. 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues.* New Jersey: Spectrum – Prentice-Hall.

9. El Kaliouby, R. 2017. 'We Need Computers with Empath', *MIT Technology Review* 120: 8–9.

10. Etzioni, A.; Etzioni, O. 2017. 'Incorporating Ethics into Artificial Intelligence', *Journal of Ethics* 21(4): 403–418.

11. Frijda, N.; Manstead, A.; Bem, S. 2000 'The Influence of Emotions and Beliefs', in *Emotions and Beliefs: How Feelings Influence Thoughts*, eds. N. Frijda, A. Manstead, S. Bem. Cambridge: Cambridge University Press, 1–9.

12. Goleman, D. 1995. *Emotional Intelligence*. New York: Bantam.

13. James, W. 1884. 'What is an Emotion?', *Mind* 9(34): 188–205.

14. Lerner, J.; Li, Y.; Valdesolo, P.; Kassam, K. 2015. 'Emotion and Decision Making', *Annual Review of Psyhology* 66: 799–823.

15. Lu, J.; Ren, L.; Zhang, C.; Liang, M.; Stasiulis, N.; Streimikis, J. 2020. 'Impacts of Feminist Ethics and Gender on the Implementation of CSR Initiatives', *Filosofija. Sociologija* 31(1): 24–33.

16. Mayer, J.; Salovey, P. 1993. 'The Intelligence of Emotional Intelligence', *Intelligence* 17(4): 433–442.

17. Meyer, J.-J. C. 2006. 'Reasoning about Emotional Agents', *International Journal of Intelligent Systems* 21(6): 601–619.

18. Oatley, K. 2008. *Emotions: A Brief History.* New Jersey: Wiley.

19. Ortony, A.; Clore, G. L.; Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

20. Petrushenko, V.; Chursinova, O. 2019. 'Philosophical and Anthropological Dimension of Technoscience', *Filosofija. Sociologija* 30(3): 199–205.

21. Picard, R. 2003. 'What Does It Mean for a Computer to "Have" Emotions?', in *Emotions in Humans and Artifacts*, eds. R. Trappl, P. Petta, S. Payr. Cambridge: MIT Press, 213–235.

22. Roff, H. 2019. 'Artificial Intelligence: Power to the People', *Ethics and International Affairs* 33(2): 127–140.

23. Russel, S.; Norvig, P. 2006. *Artificial Intelligence: A Modern Approach*. Moscow: Vilyams.

24. Scarantino, A.; de Sousa, R. 2018. 'Emotion', in *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=emotion/ (accessed 21.12.2018).

25. Sloman, A. 2001. 'Beyond Shallow Models of Emotion', *Cognitive Processing: International Quarterly of Cognitive Science* 2(1): 177–198.

26. Steunebrink, B.; Dastani, M.; Meyer, J.-J. Ch. 2008. 'A Formal Model of Emotions: Integrating Qualitative and Quantitative Aspects', in *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08)*. Amsterdam: IOS Press, 256–260.

27. Suvorov, O. 2010. 'Intelligence', in *New Philosophical Encyclopedia*. Moscow: Mysl, 127–128.

OKSANA CHURSINOVA, OLEKSANDRA STEBELSKA

Ar įmanoma sukurti emocinį dirbtinį intelektą? Filosofinė ir metodologinė analizė

Santrauka

Straipsnyje apmąstomas poreikis filosofiškai ir metodologiškai išanalizuoti žmogaus emocinės bei juslinės dimensijos prigimtį ir funkcijas, siekiant nustatyti jų išpildymo dirbtinio intelekto priemonėmis galimybes. Kompiuteriai jau tapo integralia mūsų gyvenimų dalimi, tad pilnavertė komunikacija reikalauja įgalinti juos atpažinti ir išreikšti emocijas. Atsižvelgdamos į kritinės analizės rezultatus, autorės teigia, kad emocijų išpildymas (o ne simuliavimas) bet kokioje skaičiuojančioje sistemoje šiuo metu yra problemiškas ir tam tikru mastu neįmanomas. Taip yra, nes pati „emocijos" sąvoka mokslinėje ir filosofinėje literatūroje darosi miglota. Taip pat dėl subjektyvaus ir kokybinio tikrovės patyrimo pobūdžio, asmens emocinės ir juslinės sferos įsišaknijimo fizinėje, socialinėje ir kultūrinėje terpėse, besąlygiško asmens emocijų ryšio su jo vidine asmenine erdve.

Raktažodžiai: emocijos, emocinis intelektas, emocinis dirbtinis intelektas, *qualia*, moralė