Reflexivity and Artificial Intelligence in Control (Subjectness-oriented Approach)

Vladimir Lepskiy*

** Institute of Philosophy Russian Academy of Sciences, Moscow, Russia*
(e-mail: VELepskiy@mail.ru)

Abstract: The organization of control processes is based on specialized paradigms and ontologies (subjectness, cybernetic, etc.). Control paradigms and ontologies set requirements for AI, which determine the development of AI, as well as the creation of mechanisms for controlling and neutralizing negative consequences. Reflexivity in the control of social systems has become increasingly important in recent decades. The article examines the place and role of AI in ensuring reflexive activity in the subjectness paradigms of control. The evolution of subjectness paradigms is considered from the standpoint of the development of scientific rationality. The proposed approach can be used to solve urgent problems of IFAC and TECIS due to the increasing role of hybrid reality environments (subjectness, digital, physical), including the problems of convergence of natural and artificial intelligence.

Keywords: reflexivity, artificial intelligence, control, cybernetics, scientific rationality, subjectness.

1. INTRODUCTION

Reflexivity in the control of social systems in recent decades has become increasingly important, as evidenced by the widespread use of reflexive technologies in control decision support systems, in information wars, in organizing color revolutions, in election campaigns, etc. These facts allow us to conclude that new control paradigms have been found, which have confirmed their adequacy for large social systems. In the 21st century, states that do not possess modern reflexive control technologies are doomed to become objects of control.

Reflexivity is the ability of active systems to build models of their state and behavior, as well as similar models for other systems (Lefebvre, 1967). These can be systems of both natural and artificial intelligence. In the philosophical aspect, this is due to the change of the dominant of the causal approach to the dominant of the teleological approach, with interdisciplinary integration in the construction of new methods and means of modeling and methodological support of various kinds of control subjects, with the creation of new paradigms for the integration of heterogeneous knowledge, etc.

The widespread development of these studies in cybernetics is largely associated with the works of V.A. Lefebvre, who offered formal descriptions of reflexion and reflexive processes, introduced the concepts of a reflexive system and reflexive control (Lefebvre,1967,1982). Von Foerster put reflexivity at the heart of second-order cybernetics. Cybernetics of the first order is the cybernetics of "observable systems" (N. Wiener), cybernetics of the second order of "observing systems" - with reflexion (Foerster, 1979). Further development of cybernetics is also associated with reflexivity, V.E. Lepskiy proposed third-order cybernetics of "self-developing polysubject (reflexive-active) environments" (Lepskiy, 2018a).

Reflexivity has taken its rightful place in the systems approach and cybernetics. It significantly expanded the traditional approaches to management, which were based on the functional approach, in particular, it was based on "operations research". Reflexivity contributed to the development of a structural-functional approach and a subjectness-oriented approach (Lepskiy, 1998).

The main directions of the study of reflexivity in control are associated with the support of decision-making processes, management in conflict interactions, the organization and overcoming of "subjectlessness" and the development of the reflexive abilities of subjects in social systems with the organization of reflexive processes in self-organizing and self-developing systems, etc.

The current trend of the rapid development of digitalization and artificial intelligence requires an analysis of the influence of AI on reflexivity in control, both its support and its blocking.

The article presents the results of a philosophical and methodological analysis of the place and role of AI in supporting reflexivity in the processes of control social systems. To solve this problem, it is proposed to use subjectness paradigms corresponding to the stages of development of scientific rationality (Stepin, 2005) and cybernetics (Lepskiy, 2018a). In fact, each subjectness paradigm is associated with the corresponding ontologies or systems of ontologies, which make it possible to determine the place and role of active forms of AI in interaction with subjects, carriers of natural forms of intelligence (Lepskiy, 2018d).

2. INSTRUMENTAL AND CONCEPTUAL APPROACHES TO ARTIFICIAL INTELLIGENCE

When studying the possibilities and limitations of using AI in control processes, many problematic issues arise related to the break in paradigms that set ideas about AI outside of the control paradigms. Currently, the dominant paradigms are those that consider AI not as means included in control activities and the corresponding control paradigms, but as independent objects of research in the paradigms corresponding to the specifics of these objects.

There are numerous examples of such AI paradigms:

- *morphological paradigm* - AI as a model of the brain;
- *logical paradigm* - AI as a problem solver;
- *neurocybernetic paradigm* - AI as a neural network;
- *imitation paradigm* - AI as a semblance of human reasoning;
- *weak AI paradigm* - modeling of human activities that are traditionally considered to be intellectual;
- *the paradigm of strong AI* assumes that computers can acquire the ability to think and be aware of themselves as a separate person, although not necessarily that their thought process will be similar to a human;
- *the paradigm of general AI*, capable of self-learning and solving various problems in different contexts.

These AI paradigms contribute to the development of AI problems, but they make it difficult to use these developments in control processes, leaving many ideas about AI unaddressed. The organization of control processes is based on its own specialized paradigms (subjectness, cybernetic, etc.), which put forward specific requirements for AI.

For the effective use of AI in control processes, it is advisable to consider two approaches and, accordingly, two technological levels of ideas about AI. The conceptual level is directly related to the management paradigms, which correlate with specialized AI technologies. Instrumental level deals with understanding of AI as universal means. There is a difficult problem of establishing connections between these two levels of understanding of AI (Lepskiy, 1998).

Analysis of trends in the development of control problems from the standpoint of the development of scientific rationality (classical, non-classical, post-non-classical) allows us to conclude that they are largely associated with the development of the corresponding subjectness paradigms of management (Lepskiy 2018a, 2018c).

3. REFLEXIVITY OF ACTIVE FORMS OF ARTIFICIAL INTELLIGENCE

In the context of subjectness paradigms, to ensure interactions of subjects of natural intelligence with formations of AI, it is advisable to interpret them as active forms of AI with basic properties of subjects that are invariant with respect to the type of subject (individual, group, organization, etc.). In fact, it should be about AI pseudo-subjects.

The basic invariant properties of subjects are purposefulness, reflexivity, communication, sociality and the ability to develop (Lepskiy 1998, 2018a). These characteristics can be interpreted as characteristics of AI pseudo-subjects, which is reflected in numerous publications on cybernetics and AI.

The characteristic "reflexivity" occupies a special place among all the characteristics of pseudo-subjects. First, reflexivity is the basis for identifying the stages of development of control (cybernetics of the first, second and third order). Second, it provides all the other characteristics of subjects and pseudo-subjects. Thirdly, it ensures the integrity of reflexive subjects and pseudo-subjects, mechanisms of their assembly. Fourthly, reflexivity is at the heart of the mechanisms for ensuring identity in relation to macro-subjects.

An important direction in the evolution of AI models is the improvement of the pseudo-characteristics of subjectness, approaching the corresponding characteristics of subjects based on natural intelligence.

The coexistence of subjects of natural intelligence and pseudo-subjects of AI creates a hybrid environment. In this environment, both the interaction of heterogeneous formations and their integration into new types of formations on the basis of subjects and pseudo-subjects take place. In this regard, it is advisable to introduce the general concept of active elements (ACTEL) for subjects and pseudo-subjects.

By ACTEL we mean active elements that have basic invariant characteristics of subjects, pseudo-subjects, realized at the substrate level by natural or artificial intelligence, as well as their combinations. This concept was proposed and used by the author for the design of automated organizational control systems (Lepskiy 1998).

The concept of "ACTEL" is broader than the concept "Agent" often used in control practice. The agent is delegated from subjects or pseudo-subjects certain powers to perform functions. ACTEL can be an Agent, but it can also be an independent subject or a pseudo-subject. The expediency of using the concept of "ACTEL" is associated with the organization of hybrid reality environments, consisting of carriers of natural and artificial intelligence, as well as integrated formations of natural and artificial intelligence.

4. REFLEXIVE TECHNOLOGIES OF INTERACTION OF ACTIVE FORMS OF NATURAL AND ARTIFICIAL INTELLIGENCE

The basic principle of organizing interactions and integrating active forms of natural and artificial intelligence (pseudo-subjects) in hybrid reality environments should determine the procedure for their reflexive coordination. For understanding the specifics of self-regulation of reflexive activity, the principle of “double subject” is of the greatest importance. The essence of this principle is to improve various types of activity of subjects (pseudo-subjects) by organizing interaction with partners who have adequate positions, in the formation of such partners, transferring to other subjects (pseudo-subjects) established types of activity, organizing based on the analysis of the activity of subjects of their "digital twins" (models of holistic subjects). This principle determines the technological procedures for interaction and

integration of ACTELs in hybrid reality environments (Lepskiy 1998, 2018d).

This principle defines the most important technological procedures of the self-developing reflexive-active hybrid environment and the organization of knowledge:

- procedure of reflexive decomposition of subjects (identification of subject positions);
- procedure of the virtual subject identification;
- procedure of virtual active elements generation;
- procedure of discharge from subjects of the formalized types of activity;
- procedure of reflexive synthesis (creation of activity models, models of activity subjects, personal models, etc.).

The principle of the double subject can be interpreted as dynamic transformation of subjects in the virtual group subject.

There is reason to believe that taking into account the principle of a double subject in the organization of systems using AI will make it possible to take the next step towards creating systems with the inclusion of AI elements that do not go beyond the control of their creators. The principle of the dual subject also provides grounds for the assertion of the essential role of reflexive activity when considering AI in the subjectness paradigms of control (Lefebvre 1986; Lepsky 2018b).

5. REFLEXIVITY IN SUBJECTNESS PARADIGMS OF CONTROL

The philosophical and methodological foundations for the analysis of reflexivity in the control of social systems can be the representations of scientific rationality proposed by V.S. Stepin. This is classical, non-classical and post-non-classical rationality (Stepin, 2005). The rationale for this choice can also be the connection between these stages of development of scientific rationality with the evolution of ideas about cybernetics (Lepskiy 2018a; Umpleby 2014; Umpleby and others 2019).

On the basis of modern concepts of the philosophy of science, we have proposed a systemic vision of the evolution of control problems, which can become the basis for understanding the place and role of AI in ensuring reflexivity (Lepskiy 2018b, 2018d).

In the philosophy of science, there is an idea of three stages in the development of science and, accordingly, of three types of scientific rationality: classical, non-classical, and post-non-classical (Stepin, 2005).

In classical scientific rationality, the focus is on the object of research, while the researcher himself and the means of research are completely outside the consideration and interpretation of the knowledge gained. This rationality corresponds to the stage of formation of classical cybernetics (N. Wiener). The dominant subjective paradigm "Subject – Object" reflects the essence of classical cybernetics, i.e. cybernetics of "observable systems". The dominant approach was the activity approach (Leontiev 1978), in which the role of the subject and reflexivity is significantly limited.

In non-classical scientific rationality, the focus is jointly on the object of research and the means of research; the subject of research remains out of consideration. This type of rationality corresponds to the formation of a second-order cybernetics (F. Foerster). The dominant subjective paradigm "Subject – Subject" reflects the essence of second-order cybernetics, i.e. cybernetics of "observing systems". The subject of control is interpreted as one of the participants in the control processes, in which the object of control also plays an active role. The dominant approach in providing activity, communicative and reflexive activity is the subjectness-activity approach [Rubinstein 1997, 438], in which the role of the subject increases significantly and, accordingly, the role of reflexivity increases. The focus is on communicative reflection.

In post-non-classical scientific rationality, the scope of scientific reflexion is expanding. The focus of attention is simultaneously and jointly: the object, the subject and the means of research. From the involvement of the subject in the consideration of the research process, it follows the involvement in the consideration of value-target structures, as well as the social environment and culture influencing it. It is fundamentally important that, along with intrascientific values, extrascientific social values and research goals are also taken into account. This scientific rationality corresponds to the stage of formation of third-order cybernetics [Lepskiy, 2018]. Self-developing systems are at the center of attention in management problems [Stepin 2003, 628–629]. The dominant subjectness paradigm in control is becoming the paradigm “Subject – Meta-subject (self-developing environment)”. A self-developing poly-subject system (environment), endowed with the properties of subjectness, into which the subject is immersed and with which the subject can interact through reflection, is considered as a metasubject. The focus is on metareflexion in relation to the metasubject. A family, an organization, a country, or humanity can act as a metasubject [Lepskiy 2010].

The Table 1. shows the relationship between subjectness paradigms and basic types of reflexive activity.

Table 1. Basic philosophical and methodological bases for the analysis of reflexive activity

Type of scientific rationality	Basic paradigms	Basic areas of knowledge	Types of reflexive activity
Classical	“Subject – Object”	Cybernetics	Personal reflexion
Non-classical	“Subject – Subject”	Second-order cybernetics	Communi-cative reflexion
Post-non-classical	“Subject – Meta-Subject”	Third-order cybernetics	Meta-Subject reflexion

The fundamental feature of the three types of scientific rationality is that each subsequent type of scientific rationality includes the previous one as a particular paradigm.

Therefore, post-non-classical scientific rationality includes all the considered types of reflexive activity that are present in classical and non-classical rationality. Moreover, reflexive activity becomes dominant in post-non-classical scientific rationality.

6. REFLEXIVITY AND ARTIFICIAL INTELLIGENCE IN THE "SUBJECT – OBJECT" PARADIGM

In the paradigm "Subject – Object" the activity approach predominates, as a result, reflexivity turns out to be without the necessary attention. Stimulation and support of reflexivity is relevant in non-standard cases to ensure decision-making processes in problem situations. Their resolution requires the subject to go beyond the established control activity, requires a "reflexive exit". The problem of "reflexive exit" attracted the attention of methodologists (Shchedrovitsky 1975) and psychologists. Psychologists (V.A.Petrovsky, Ya.A. Ponomarev, D.B. Bogoyavlenskaya and others) identified it as a problem of oversituational activity.

For control practice, this problem is of interest in two aspects. First, when identifying and improving the ability of managers to oversituational activity. Secondly, when developing mechanisms to stimulate and support oversituational activity. The second aspect can be of interest when setting and solving new complex problems using AI. To solve them, active forms of AI can be used, which should ensure the creation of a pseudo-subject partner (digital twin) for the subject of control, helping to assess the current situations of break points in control activities, offering either existing solutions, or stimulating a reflexive output to search for non-standard solutions and support in reflexivity. It should be noted that there is a rich interdisciplinary scientific groundwork for the formulation and solution of these problems (Lepskiy 1998).

The reflexive activity of the control object within the framework of the “Subject-Object” paradigm remains out of consideration.

7. REFLEXIVITY AND ARTIFICIAL INTELLIGENCE IN THE "SUBJECT – SUBJECT" PARADIGM

The role of communicative activity and the forms of reflexive activity (communicative reflexion) increases sharply in the subject-subject paradigm of subjectness. This manifested itself in the transition from the activity-based approach to the subjectness-activity approach (Rubinstein 1997).

Intensive studies of the possibilities of using reflexive activity in control processes, using the example of conflict situations, were carried out in the 60s – 70s of the last century in the USSR under the leadership of V.A. Lefebvre (Lefebvre 1967). Mathematical models have been developed and experimental studies have been carried out. The concept of reflexive control is proposed, which has found wide application in the military sphere, educational technologies, economics, etc.

In the context of AI, it should be noted that automata (pseudo-subjects) were created that, taking into account the stereotypical forms of a persons’ reflexive activity, could replay them in conflict interactions, even without knowing the person's goal (for example, the sides of the payment matrix in models of the theory of games with zero sum) (Lefebvre 1967).

Reflexive control is one of the forms of reflexive technologies that include a wide range of technologies for controlling the reflexive activity of subjects and pseudo-subjects. Three groups of reflexive technologies have been proposed: imitation of reflexive activity, reflexive control, and reflexive programming (Lepskiy 2018b). These reflexive technologies can be used both in conflict situations and to support joint activities in various fields of activity (Lepskiy 2018d; Kauffman 2016; Espejo and others 2021; Müller 2015; Novikov 2016).

The use of AI is possible in providing all types of reflexive technologies by creating adequate types of AI pseudo-subjects (digital subjects). In fact, we can talk about the creation of communities of pseudo-subjects corresponding to the reflexive structures of specific reflexive technologies.

An example of the possibility of using the system of AI pseudo-subjects to ensure reflexive activity is decision support in the context of negotiation processes, when a partner uses 36 Chinese stratagems. Reflexive analysis of 36 Chinese stratagems made it possible to associate a certain reflexive technology with each stratagem. This makes it possible, due to the complication of the reflexive technology, adequate to the used stratagem, to neutralize the threat of manipulative influences and to transfer interaction from conflict to partnership (Lepskiy 2018b). In practice, the problem of operational forecasting and identification of the applied stratagem arises. This problem could be successfully solved using a system of specialized AI pseudo-subjects, which would promptly identify potentially used stratagems and advise the use of adequate reflexive technologies based on the principle of a double subject.

Note that the sphere of education could become one of the most relevant for the use of AI in supporting various forms of reflexive activity in the "Subject-Subject" paradigm.

8. REFLEXIVITY AND ARTIFICIAL INTELLIGENCE IN THE "SUBJECT – META-SUBJECT" PARADIGM

The “Subject – Meta-subject” paradigm is formed in the context of post-non-classical scientific rationality, in which the focus is on self-developing human-sized systems (Stepin 2005). It is important to note that control sciences are influenced by both intrascientific and extrascientific (social) values. At the same time, self-developing human-sized systems are organically integrated into the culture.

These considerations give grounds for introducing the concept of self-developing polysubject (reflexively active) environments, which are considered as integral formations – Meta-subjects. Subjects included in them (pseudo-subjects) can identify themselves with them. Subject aspects in such environments have increased, which stimulated the formation of a subjectness-oriented approach, providing convergence with activity and subjectness-activity approaches.

The Meta-subject is a virtual entity that enhances the role of reflexive activity in the subjective paradigm “Subject – Meta-subject (self-developing environment)”. A new form of reflexive activity appears – Meta-subject reflexion.

The existence of subjects and pseudo-subjects in self-developing poly-subject (reflexively active) environments is set by a system of ontologies, which ensures the assembly into a whole (Meta-subject) of subjects (pseudo-subjects) included in the environment. Pseudo-subjects can be of three types: digital twin, digital subject, and digital meta-subject.

We have developed the system of ontologies: support of the established types of activities (communication) and their subjects (“accompanying”); support of subjects in the points of disruption of established types of activity (communication) and the reproduction of their subjects ("support"); development of established types of activity (communication) and their subjects (“development”); the design of new types of activities (communications) and new actors (“construction”); implementation of innovative projects of new types of activities (communications) and new actors (“innovation”). This system of ontologies has undergone scientific, methodological and practical approbation (Lepskiy 1998, 2018a; Umpleby and others 2019; Espejo and others 2021).

The setting of tasks for AI should be carried out through the system of ontologies of being of subjects and pseudo-subjects, as well as in the context of supporting the reflexive activity of subjects and pseudo-subjects in relation to Meta-subjects.

Currently, the first steps are being taken in understanding the role and tasks of AI in the subjectness paradigm "Subject – Meta-subject (self-developing environment)". But there is reason to assert that the developed philosophical and methodological foundations will allow intensifying the development of AI in the interests of improving the management processes and the development of social systems (Umpleby and others 2019; Espejo and others 2021).

9. CONCLUSIONS

The organization of control processes is based on specialized paradigms (subjectness, cybernetic, etc.), which put forward specific requirements for the concepts of AI and the tasks in which it is advisable to use AI. We have identified subjective paradigms in control that are adequate to the types of scientific rationality (classical, non-classical, post-non-classical) ("Subject – Object", "Subject – Subject", "Subject – Meta-subject").

A model of active forms of AI as pseudo-subjects (digital twins) is proposed. The basic principle of "double subject" for organizing mechanisms of interaction between subjects and active forms of AI (pseudo-subjects) is considered. For each of the subjectness paradigms of control, the dominant types of activity are considered: activity, communicative, reflexive. The key role of reflexive activity has been substantiated, in relation to which the basic tasks for AI in three subjectness paradigms of control have been determined.

ACKNOWLEDGMENTS

This work is funded by Russian Science Foundation, project 21-18-00184 “Social and humanitarian foundations of criteria for evaluating innovations using digital technologies and artificial intelligence”.

REFERENCES

Espejo, R., Lepskiy, V. (2021). An agenda for ontological cybernetics and social responsibility, *Kybernetes*, Vol. 50, No. 3, pp. 694-710.

Foerster, Heinz. (1979). *Cybernetics of Cybernetics*. University of Illinois, Urbana.

Kauffman L.H. (2016). Cybernetics, reflexivity and second-order science. *Constructivist Foundations*. 11(3), pp. 489–497.

Lefebvre, V.A. (2010). *Lectures on the Reflexive Games Theory*. Leaf & Oaks Pablishers, NY.

Lefebvre (1986). Second Order Cybernetics in the Soviet Union and the West. *Power, Autonomy, Utopia: New Approaches toward Complex Systems* / Ed. Robert Trappl, N.Y.: Plenum Press, pp. 123-131.

Lefebvre, V.A. (1982). *Algebra of Conscience*. Dordrecht, Holland, Reidel.

Lefebvre, V.A. (1967). *The conflict structures.* Vysshaya shkola, Moscow (in Russian).

Leontiev, A. N. (1978). Activity, consciousness, and personality, Prentice-Hall, Englewood Cliffs, USA.

Lepskiy, V. (2018a). Evolution of Cybernetics: Philosophical and Methodological Analysis. *Kybernetes,* vol. 47, no. 2, pp. 249–261.

Lepskiy, V. (2018b). Reflexive self-organizing decision support systems for development governance. *International Journal of Engineering & Technology*, Vol 7, No 2.28, pp. 255–258.

Lepskiy, V. (2018c). Philosophical-Methodological Basis for the Formation of Third-Order Cybernetics. *Russian Journal of Philosophical Sciences*, No. 10, pp. 23-36.

Lepskiy, V. (2018d). Decision Support Ontologies in Self-Developing Reflexive-Active Environments. *IFAC PapersOnLine*, Vol. 51, Issue 30, pp. 504-509. https://doi.org/10.1016/j.ifacol.2018.11.276

Lepskiy, V. (1998). *The Concept of Subject-oriented Computerization of Control Activity*. Institute of Psychology RAS, Moscow (in Russian).

Medvedeva, T.A. (2017). View of V.E. Lepskiy’s and S.A. Umpleby’s Theories of Cybernetics through the Prism of Intellectual Traditions. *Reflexive Processes and Control*, Proceedings of XI International Symposium "Reflective Processes and Control", October 16–17, 2017, Moscow, Russia, Kogito Center Publishing House, pp. 32–37.

Müller, Karl H. (2015). The Multiple Faces of Reflexive Research Designs. *Systemics, Cybernetics and Informatics*, vol 13, No 6, pp.87–98.

Novikov, D.A. (2016). *Cybernetics: from Past to Future*. Springer Verlag.

Rubinshteyn, Sergei L. (1997). *The selected philosophical-psychological works*. Nauka, Moscow (in Russian).

Shchedrovitsky, G.P. (2002). Reflexion and Relevant Problems. *Reflexive processes and control*, vol.1, No 1, pp. 41–45.

Stepin, V. (2005). *Theoretical Knowledge*. Springer Verlag GMBH.

Umpleby, S., Medvedeva, T., Lepskiy, V. (2019). Recent Developments in Cybernetics, from Cognition to Social Systems. *Cybernetics and Systems*, vol. 50, issue 4, pp. 367-382.

Umpleby, Stuart A. (2014). Second order science: logic, strategies, methods. *Constructivist Foundations*. vol. 10, no. 1, pp. 16–23.