Modeling reflection in artificial intelligence systems: state of art and prospects

Sergey Listopad[1*], *Vladimir* Matsoula[1], and *Alexander* Luchko[2]

[1]Kaliningrad Branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 5 Gostinaya Str, Kaliningrad 236022, Russian Federation
[2]Avtotor Information Technologies LLC, 4a Magnitogorskaya Str, Kaliningrad 236013, Russian Federation

Abstract. The paper is devoted to analyzing the current state of research and assessing the prospects for modeling reflexive processes using artificial intelligence systems, in particular, multi-agent systems for relevant simulation of collective problem solving. Reflexive modeling of each other by agents will ensure the development of a coherent model of the control object, the purpose of collective work and norms of interaction, as well as the developing of effective interaction between agents. This will allow the system, by self-organizing, to adapt to the characteristics of arising problems, in particular, to take into account their complex structure, the network nature of conditions and goals, opacity, subjectivity and dynamism.

1 Introduction

Solving practical problems arising in production [1], economic [2] and social systems [3] by teams of specialists under the leadership of a decision maker can potentially provide their comprehensive analysis, taking into account different points of view and interests [4]. Such teams are capable of solving weakly formalized and informal problems that are initially undefined as well as the knowledge necessary to solve them. Long-existing teams of specialists demonstrate self-organization and adapt to changes in the problem situation, developing anew a solution method relevant to the next problem. A significant disadvantage of collective problem solving is the long duration of discussions required to develop an agreed solution, which makes it impossible to use this approach for highly dynamic problems. In this regard, to increase the speed of developing solutions, computer modeling of collective problem solving is relevant, as well as the processes and effects that arise during it.

One of the key phenomena that ensures the development of a team as a self-organizing environment in which subjects (specialists and their groups) harmoniously coexist and interact to achieve a common goal is reflection. Reflexive control processes and mechanisms for simulating reasoning can significantly reduce the time spent on excessive communication between team members. The reflection of subjects ensures the stability of collective reasoning in conditions of inaccessibility of one or more team members. In this case, reflective subjects reason "for the absent members", develop solutions that would be

* Corresponding author: ser-list-post@yandex.ru

proposed by inaccessible specialists, imitate the transmission of such solutions to themselves and build their further reasoning in accordance with them [5]. This paper discusses the concepts of reflection and reflexive control, as well as issues of their computer modeling.

2 Concepts of reflection and reflexive control

Reflection is a concept of philosophical discourse that characterizes the form of human theoretical activity in understanding one's actions, culture and its foundations; the activity of self-knowledge, revealing the specifics of the mental and spiritual world of man [6]. Problems associated with reflection are interdisciplinary in nature, studied by philosophy, psychology, pedagogy, and depending on the field of study, the corresponding aspect of reflection is highlighted [7]. Reflection is both a unique property inherent only to a person, a state of awareness of something, and the process of representing one's own content to the psyche [8]. According to the "temporal" principle, situational, retrospective and perspective reflection are distinguished, and according to direction inter- and intrapsychic are, which can additionally be divided into cooperative, communicative, personal and intellectual [7]. The depth of understanding of oneself and other people varies, described by the "rank of reflection", i.e. the number of successive inclusions of other characters into the consciousness of a certain character [9].

The concept of reflexive control originated in the early 60s of the last century in the context of applied research of the Moscow Methodological Group [10] in the works of V.A. Lefebvre [9, 11], who defined it as the process of transferring the grounds for decision-making by one of the opponents to another. Developing these works, V.E. Lepsky proposes to expand the interpretation of the reflexive aspects of control by including in this concept the control of the structures of awareness processes, i.e. reflexive structures [12].

Nowadays reflexive technologies in the activities of an enterprise are used to position an organization in the market, increase their level of competitiveness, coordinate goals in joint-stock companies between the board of directors and the meeting of shareholders, expand the space of analyzed alternatives and increase the efficiency of decision-making. Considering reflexive control of an organization in a broad sense, A.V. Karpov defines it as management, built taking into account and on the basis of psychological patterns of reflection, a synonym for psychologically grounded, humanistically oriented management [7].

An experimental study of the specifics of the reflexive organization of thinking during joint solving creative problems was carried out in [13]. When jointly solving a creative problem, a specific problem-oriented conflict situation arises due to the discrepancy between the intellectual capabilities of the subjects and the requirements of the problem, along with the need to jointly solve it, i.e. developing a unified strategy and problem solving method. The reflection is considered as a mechanism for resolving such a situation, which has a significant impact on the effectiveness of joint creative work, indicating the relevance of its modeling in collective artificial intelligence systems and the study of its influence on the quality of solutions offered by such systems.

Western analogues of Soviet and Russian studies of reflexive control are the works, appeared since the 1970s, of researchers in the field of second and higher order cybernetics, as well as meta-cybernetics, who noted the importance of the positive feedback in the control and self-organization of systems of various kinds [10]: H. von Foerster [14], E. Von Glasersfeld [15], L. Lofgren [16], N. Luhmann [17], U. Maturana [18], G. Pask [19], G. Roth [20], S.J. Schmidt [21], E. Schwarz [22], R. Uribe [18], F. Varela [18], P. Watzlawick [23], M. Zeleny [24]. Comparing the works of Soviet and Western scientists in this area, V.A. Lefebvre noted a more developed epistemology in Western works and, at the same time, a focus on solving specific problems and clarity in the formulation of Soviet research [25].

Unlike first-order cybernetics, which distinguish subject and object, points to a supposed independent world outside of us, second-order cybernetics is itself cyclical: a person views himself as part of the world that he observes [26]. If first-order cybernetics is based on "hard control" (subject-object context), then second-order cybernetics uses "softer", subject-subject forms of control, focused on self-organization processes [27]. The evolution of cybernetics occurs in the direction of expanding ideas about an object and a control system: first-order cybernetics considers objects as "observable systems", and second-order cybernetics as active "observing systems". The transition from the "subject–object" paradigm to the "subject–subject" paradigm of control has led to new ideas about it types: reflexive control [11], information control [28], control of active systems [29], etc. [27].

The next step in the evolution of cybernetics is the paradigm of third-order cybernetics, based on the increasing role of subjectivity and reflexivity in control [27]. E. Schwartz, considering the inadequacy of autopoiesis in living systems, pointed out that this problem can be solved with an additional network of learning processes called autogenesis, and living systems represent a triad "system-metasystem-metametasystem", the elements of which are connected by autopoiesis and autogenesis [30]. R.G. Mancilla notes that power, culture and institutions are important for social cybernetic systems of the third order [31]. According to V.E. Lepsky, in third-order cybernetics, the object and the control system merge into a single whole in a self-developing reflexive-active environment (SDRAE), i.e. there is a transition from the "subject–subject" paradigm of second-order cybernetics to the "subject–metasubject (SDRAE)" paradigm [27]. SDRAE is a holistic formation, a metasubject having invariant properties for different types of subjects: activity, reflexivity, communication skills, sociality, ability to develop. The active elements that make up such an environment can be formed on the basis of natural and/or artificial intelligence. The organization of their interaction with each other and with the environment is determined by their systems of values or principles, ontologies and specialized subject-oriented information platforms.

Currently, work is appearing in the field of fourth-order cybernetics, but the concept cannot be called fully formed and generally accepted. In [32], the concept of metacybernetics is proposed as a general theory of cybernetics of the second and higher orders, which is considered in the terminology of agency theory.

3 Modern approaches to modeling reflexive processes and control

The foundations of mathematical modeling of reflexive processes and control are laid in the works of V.A. Lefebvre [11]. D.A. Novikov and A.G. Chkhartishvili, within mathematical modeling of reflexive processes, attempted to construct a unified concept of equilibrium for the class of reflexive games [33]. They study various situations of collective decision-making, the dependence of agents' gains on the ratio of their reflection ranks, the conditions for the existence and feasibility of equilibria, and show the limitations of the maximum expedient reflection rank. In [34], theoretical and methodological foundations, conceptual provisions and practical recommendations for modeling reflexive management of herd-like behavior in enterprises are developed in order to increase the efficiency of their functioning. In [35] B.A. Kobrinsky proposes using fuzzy representations in medical artificial intelligence systems to formalize the reflexive processes that arise in the mind of a doctor when making a diagnosis. Paper [36] discusses the issues of creating intelligent automated teaching systems with a reflection function that provide an individual approach to the formation of a training program based on interaction between the student, teacher, tutor and methodologist models.

The results of research on reflexive processes and control obtained for social, economic, organizational and other systems are currently being actively translated into the field of artificial intelligent systems, in particular multi-agent systems (MAS) [37], consisting of

many interacting software or hardware autonomous agents [33]. Due to the distributed nature of the formation of the MAS function in the interaction of many agents, such systems are well suited for modeling social effects of various kinds, in particular reflexive processes and control. At the same time, the structure of the MAS agent can vary significantly: from the simplest reactive agents to intelligent agents with a developed domain model, goal-setting mechanisms and a hierarchical structure containing operational, tactical, strategic and conceptual levels. Decision making, learning, adaptability and reflexivity of agent behavior are ensured at the strategic level of the hierarchy.

The research [38] is dedicated to a group behavior taking into account differences in the reflection ranks of MAS agents. It proposes a method of partitioning of a set of rational agents into subsets according to their reflection ranks. The issues of the influence of the distribution of agents by reflexion ranks (the shares of agents of one or another rank in the MAS) on the behavior of the MAS as a whole and the optimal distribution of the shares of agents of each rank from the point of view of a collective goal are considered. The work describes the general theory, performs mathematical modeling of individual cases, as well as simulation modeling using the example of the process of evacuating people from a building.

In [39], swarm intelligence systems are studied in which "intelligent" behavior arises as a result of the self-organization of a large number of primitive agents interacting with each other and the environment according to simple rules. The study of this system was carried out using the problem of maintaining the energy autonomy of a station in conditions of limited resources. Computer modeling of the system was carried out using Netlogo system. The paper shows that the developed algorithms for group search for resources provide targeted collective behavior of system agents and allow solving the problem of station survival with the location of resources unknown in advance.

The authors of [40] propose the method based on computer reflexive games for teaching and training both individual specialists and teams of information system operators. Computer reflexive games are a source of artificial influence on a person, while the analysis of his game decisions characterizes various aspects of his personality. During the learning process, such games promote safety, do not require the use of expensive equipment used by trained operators in the work process, and makes it possible to regulate the pace of learning and the activation of the certain individual skills of the student. In [41], reflexive models that take into account sudden and "shadow" events, for which it is impossible to assess their probabilistic nature, are proposed to be used to analyze the future in systems of distributed situational development centers for state strategic goal-setting and management.

An analysis of works in the field of modeling reflexive processes and control demonstrates that they are mostly devoted to the construction of mathematical models of these processes: the number of works in this direction is an order of magnitude greater than the number of works devoted to the development of automated systems that model these phenomena. Computer modeling of reflexive processes and control is carried out mainly on the basis of rather primitive agents that are not intended to solve practical problems.

4 Prospects for computer modeling of reflection in teams of specialists

To build intelligent systems comparable in level of development to teams of specialists, it is relevant to create reflexive-active systems of artificial heterogeneous intelligent agents (RASAHIA), which model each other's reflexive positions and have developed domain models and goal-setting mechanisms. The design features of RASAHIA are determined, first of all, by the class and characteristics of the problems for which they are intended to solve, such as weak formalization, complex structure, heterogeneity, network nature of conditions and goals, significant opacity (uncertainty) and dynamism:

- due to weak formalization, the problem solving process often begins with developing its formulation, because initially it may not be defined as well as the knowledge necessary to solve it;
- the significant uncertainty of the problem and the chaotic nature of the processes do not allow operating with traditional logical and probabilistic-causal methods. As a result, the system must operate in conditions that are unpredictable based on retrospective data [41];
- the heterogeneity, network nature of conditions and goals require that in order to solve the problem it is necessary to combine into a single system the methods of various artificial intelligence technologies and computer models of specialists with different goals and views on the problem. These features are especially relevant for large-scale problems that require the involvement of various independent development teams in the construction of agents. In this case, agents, using reflexive mechanisms, must independently agree on goals, domain models, and a problem solving protocol [42];
- dynamism at the environmental and problem levels. The dynamism of the environment in which the agents of the system exist implies limited time to find a solution to the problem in accordance with some method or protocol, i.e. their work in real time. The dynamism of the problem limits the time to build a method for solving it and the impossibility of using previously developed methods unchanged;
- complexity leads to the need to study the problem and design the system at different levels. A problem is an interconnected set of subproblems and elementary tasks that can form a hierarchical system, and the properties of the problem as a whole are determined not only by the properties of its components, but also by the connections between them.

RASAHIA is proposed to be created within the multi-agent approach [37] using the model of hybrid intelligent multi-agent systems [1]. The multi-agent approach will ensure greater autonomy of agents compared to objects in traditional object-oriented programming. Agents, unlike objects, have such integrated properties as reactivity, proactivity, sociality, etc. When constructing agents in their composition, a wide range of computational methods and artificial intelligence technologies required to create an integrated method for solving the problem posed can be used.

5 Conclusion

The paper examines the concepts of reflection and reflexive control, as well as the current state of research in the field of their computer modeling. The analysis of the reviewed publications showed that they are devoted primarily to mathematical modeling of reflexive processes, as well as computer modeling based on rather primitive agents that are not relevant to real specialists. As a promising direction for building intelligent systems that model reflexive processes and are comparable in level of development to teams of specialists, the development of RASAHIA is proposed and the main features of their design are considered.

The study was supported by the Russian Science Foundation grant No. 23-21-00218, https://rscf.ru/project/23-21-00218/.

References

1. A.V. Kolesnikov, I.A. Kirikov, S.V. Listopad, *Gibridnye intellektual'nye sistemy s samoorganizatsiey: koordinatsiya, soglasovannost', spor* [*Hybrid intelligent systems with self-organization: coordination, consistency, dispute*] (IPI RAN, Moscow, 2014)
2. A.V. Kolesnikov, S.V. Listopad, Systems and Means of Informatics **28(4)**, 31 (2018)
3. I.A. Kirikov, A.V. Kolesnikov, S.V. Listopad, S.B. Rumovskaya, Informatics and Applications **10(3)**, 81 (2016)
4. M.V. Samsonova, V.V. Efimov, *Technology and methods of collective problem solving: Textbook* (UlGTU, Ul'yanovsk, 2003)
5. V.A. Lefebvre, *Reflection* (Kogito-Center, Moscow, 2003)
6. A.P. Ogurtsov, Reflection (2018).
   https://iphlib.ru/library/collection/newphilenc/document/HASHc974bd5cb7cf4bf458b2c9
7. A.V. Karpov, *Psikhologiya menedzhmenta: Uchebnoe posobie* [*Psychology of management: Textbook*] (Gardariki, Moscow, 2005)
8. A.V. Karpov, Psikhologicheskii Zhurnal **24(5)**, 45 (2003)
9. V.A. Lefebvre, *Initial ideas of the logic of reflexive games*, in Proceedings of the conference “Problems in the study of systems and structures (Moscow, 1965)
10. V.N. Usov, *Reflexive management: philosophical and methodological aspect* (SUSU Publishing Center, Chelyabinsk, 2010)
11. V.A. Lefebvre, *Conflicting Structures* (Leaf & Oaks Publishers, Los Angeles, 2015)
12. V.E. Lepskiy, *Control technologies in information wars (from classics to post-non-classics)* (Kogito-Center, Moscow, 2016)
13. I.N. Semenov, S.Yu. Stepanov, M.I. Naydenov, L.A. Naydenova, New research in psychology and developmental physiology **1**, 4 (1989)
14. H. von Foerster, *Observing Systems* (Intersystems Publications, USA, 1981)
15. E. von. Glasersfeld, *Cybernetics, experience, and the concept of self*, in M. N. Ozer (Ed.) A Cybernetic Approach to the Assessment of Children: Towards a more Humane Use of Human Beings (Routledge, New York, 2018)
16. L. Lofgren, Systems Research **13(3)**, 329 (1996)
17. N. Luhmann, Soziale Systeme **24(1-2)**, 18 (2019)
18. F. Varela, H. Maturana, R. Uribe, Biosystems **5(4)**, 187 (1974)
19. G. Pask, *Conversation, Cognition and Learning: A Cybernetic Theory and Methodology* (Elsevier, Amsterdam, 1975)
20. G. Roth, H. Schwegler, *Self-Organization, Emergent Properties and the Unity of the World*, in W. Krohn, G. Kuppers, H. Nowotny (Eds.) Selforganization: Portrait of a Scientific Revolution (Kluwer, Dordrecht, 1990)
21. S.J. Schmidt, Constructivist Foundations **6(1)**, 6 (2010)
22. E. Schwarz, Cybernetics and Human Knowing **4(1)**, 17 (1997)
23. P. Watzlawick (ed.), *The invented reality: how do we know what we believe we know?: (contributions to constructivism)* (Norton, New York, 1984)
24. M. Zeleny (ed.), *Autopoiesis, Dissipative Structures, and Spontaneous Social Orders* (Western Press, Boulder, CO, 1980)

25. V.A. Lefebvre, *Second Order Cybernetics in the Soviet Union and the West*, in R. Trappl (ed.) Power, Autonomy, Utopia (Springer, Boston, MA., 1986)
26. K.N. Knyazeva, Filosofskie nauki **11**, 88 (2010)
27. V.E. Lepskiy, *Methodological and philosophical analysis of the development of management issues* (Kogito-Center, Moscow, 2019)
28. D.A. Kononov, V.V. Kulba, A.N. Shubin, *Information management: principles of modeling and areas of use*, in Trudy IPU RAN [Proceedings of the IPU RAS] (IPU RAN, Moscow, 2004)
29. V.N. Burkov, V.V. Kondratyev, *Mechanisms of functioning of organizational systems* (Nauka, Moscow, 1981)
30. E. Schwarz, Autogenesis, 32 (2021). https://ssrn.com/abstract=3826203
31. R.G. Mancilla, J. Sociocybern **9(1/2)**, 35 (2011)
32. M. Yolles, Systems **9(2)**, 34 (2021)
33. D.A. Novikov, A.G. Chkhartishvili, Reflection and control: mathematical models (Fizmatlit, Moscow, 2012)
34. S.S. Turlakova, *Reflexive management of herd behavior in enterprises: concept, models and methods* (IEP NAN Ukrainy, Kiev, 2020)
35. B.A. Kobrinskiy, *Reflection and fuzzy representations in medical artificial intelligence systems*, in Reflexive control. Abstracts of the international symposium (Publishing house "Institute of Psychology RAS", Moscow, 2000)
36. V.A. Uglev, Mathematical Structures and Modeling **1(45)**, 111 (2018)
37. M. Wooldridge, *An Introduction to Multiagent Systems* (Wiley, New York, 2009)
38. V.O. Korepanov, *Models of reflexive group behavior and management* (IPU RAN, Moscow, 2011)
39. S.N. Sirenko, A.V. Kolesnikov, G.G. Malinetskiy, O.S. Sirenko, *Multi-agent modeling of the dynamics of a bio-social system in the presence of social parasitism*, in Proceedings of the XII International Scientific and Practical Interdisciplinary Symposium “Reflexive Processes and Management” (Kogito-Center, Moscow, 2019)
40. E.P. Popechitelev, K.N. Bolsunov, Izvestiâ SPbGÈTU "LÈTI" **6**, 110 (2013)
41. Z.K. Avdeeva, A.N. Raykov, V.P. Bauer, S.N. Silvestrov, B.B. Slavin, A.A. Zatsarinnyy, K.K. Kolin, V.E. Lepskiy, G.G. Malinetskiy, *A system of distributed situational development centers for sustainable strategic management*, in Proceedings of the XII International Scientific and Practical Interdisciplinary Symposium “Reflexive Processes and Management” (Kogito-Center, Moscow, 2019)
42. S.V. Listopad, Systems and Means of Informatics **30**, 78 (2020)