BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Centre, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2026, Volume 17, Issue 2, pages: 226-239
Submitted: December 28th, 2025 | *Accepted for publication:* April 23rd, 2026

Agent-Oriented Control of Robotic Systems Based on Artificial Intelligence: Utilising Cutting-Edge Neuroscientific Insights for Verification, Validation, and Optimisation

Oksana Yashyna
PhD of Technical Sciences, Associate Professor, Associate Professor of the Department Software Engineering, Faculty of Information Technologies, Khmelnytskyi National University, Khmelnytskyi, Ukraine, yashynao@khmnu.edu.ua, https://orcid.org/0000-0001-7816-1662

Denys Makaryshkin
PhD of Technical Sciences, Associate Professor, Associate Professor of the Department of Automation, Computer-integrated Technologies and Robotics, Faculty Of Information Technologies, Khmelnytskyi National University, Khmelnytskyi, Ukraine, makaryshkin@ukr.net, https://orcid.org/0000-0003-3447-811X

Yurii Forkun
PhD of Technical Sciences, Associate Professor, Associate Professor Of the Department of Automation, Computer-integrated Technologies and Robotics, Faculty Of Information Technologies, Khmelnytskyi National University, Khmelnytskyi, Ukraine, yvforkun@gmail.com, https://orcid.org/0000-0002-7906-4191

Roman Kustovskyi
PhD Student, Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine, roma.kust99@gmail.com, https://orcid.org/0000-0003-0785-7202

Vitalii Sorokolit
PhD Student, Department of Computer Science, Khmelnytskyi National University, Khmelnytskyi, Ukraine, sorvitali@gmail.com, https://orcid.org/0009-0002-0182-221X

Abstract: *Using artificial intelligence and the latest developments in neuroscience, the authors investigate the challenges in improving the efficiency of agent-oriented control in robotic systems. The goal is to develop a theoretical concept taking into account the existential-objective, simulation-cognitive, and neurobehavioural levels of control in agent-oriented systems. This will lay the groundwork for improving the hardware and software of such agents. The article employs methods of system-analytical and comparative analysis, neuro-oriented modelling of control concepts, and formal verification approaches. The research results in the creation of an original theoretical foundation for Neuro-Agentic Verification Control (NAVC), which combines digital twin simulation, synaptic neural networks, and cognitive-ontological principles. The authors paid significant attention to the integration of formal interfaces and natural language, and other cognitive systems. Finally, the authors provide formal proofs to demonstrate the validity of the key components of the proposed architecture. The article's international significance is determined by the relevance of addressing challenges related to transparency, safety, and reliability of autonomous robotic systems in critical domains, including transportation, industry, and defence.*

Keywords: *robotics; robotic system; Artificial Intelligence; neural network; neuro-control; computer vision; machine learning; intelligent control methods; software; NAVC model.*

How to cite: Yashyna, O., Makaryshkin, D., Forkun, Y., Kustovskyi, R., & Sorokolit, V. (2026). Agent-oriented control of robotic systems based on artificial intelligence: Utilising cutting-edge neuroscientific insights for verification, validation, and optimisation. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *17*(2), 226-239. https://doi.org/10.70594/brain/17.2/13

1.Introduction

Currently, the capabilities of artificial intelligence (AI) and robotic systems are evolving "at a geometric rate," creating new and even unique opportunities for building autonomous, adaptive, and self-learning systems capable of flexible interaction with dynamic environments. At the core of this technological progress lies the agent-oriented approach, which enables a new paradigm for the design and control of robotic systems based on distributed intelligent behaviour (Wan et al., 2022). We also consider neuroscience developments, including neural network models, neural control, and cognitive software. These tools provide new means for the verification, enhancement, and testing of robotic systems.

2. Relevance of the Study

The rapid development of AI technologies is opening broad horizons for the advancement of robotics. Robotics is now acquiring the capacity for efficient distributed collective control of simpler entities – drones, logistics units, and other autonomous systems. Although agent-based control methods are widely used and effective in manufacturing (Li et al., 2021), their functionality in these domains remains relatively well-defined and finite. There is currently a lack of more effective and adaptive learning agents for robotics, which is especially critical in conditions of high uncertainty and limited data.

Unfortunately, we have observed that when adaptive behaviour systems possess a significant degree of freedom, they often make mistakes. Consequently, the user cannot fully benefit either from their implementation or from standard software verification and validation (V&V) methods.

To guarantee the safety and efficiency of such systems, it is becoming increasingly evident that it is necessary to incorporate advances in neuroscience, particularly models of cognitive processes. This is especially crucial for V&V, specifically the integration of computer vision, intelligent self-learning methods, and neuro-control. In our view, this also creates both new potential and new challenges, as these algorithms self-modify through use and self-learning.

We observe that rigorous formal methods for evaluating the accuracy of adaptive software that utilise neuro-behavioural modelling concepts are not yet sufficiently developed (Sreekanth et al., 2023). To ensure a comprehensive design, verification, and operational safety lifecycle, it is necessary to develop novel approaches that integrate agent-oriented control with principles from neuroscience.

Furthermore, an analysis of the existing body of work indicates that the potential of neuroscience for such control is significant. Neuroscience already provides biomorphic data for creating hybrid cognitive agents, refining artificial neural network topologies, and developing biologically-inspired decision-making models. However, the methodology and means of implementing these achievements for dynamic, self-learning robotic systems remain a fundamental challenge.

The scientific community has long been actively exploring agent-oriented systems through interdisciplinary approaches. For instance, the works of Jennings (2001) laid the foundations of agent modelling for complex distributed systems, but they did not take into account cognitive processes. At the turn of the millennium, researchers proposed formal methods for describing agent behaviour (Weiß & Ciancarini, 2002), yet their approaches left open the question of effective verification of self-learning agents.

Studies from various years, particularly Arkin's (1998) work on behavioural models in robotics, provided a foundation for the development of bio-inspired systems, but these still require further refinement in the context of adaptive neuro-control. Once-innovative approaches, such as the multi-layered agent architecture with elements of machine learning by Rosenblatt and Payton (1989, June), demonstrated limited scalability in real-world scenarios. Around the same time, the integration of neuro-cognitive mechanisms proposed by Franklin and Graesser (1997) within the LIDA project pointed to the potential of modelling agent systems analogous to human thinking that did not offer ready-made solutions for formal verification of dynamic behaviours.

A distinct niche and subsequent stage is occupied by research on deep neural networks and their capacity for self-learning (Schmidhuber, 2015), although the issue of methodological transparency of such models for V&V remains largely unresolved.

3. Purpose of the Study

The aim of this article is derived from the outlined challenges, previous research experience, and current technological trends. It first focuses on reviewing neuroscientific advances that could serve as a foundation for optimising agent-oriented control of robotic systems. Following this, the article seeks to develop an original model that could become a basis for improving the verification, validation, and optimisation of agent-oriented robotic control grounded in neural network architectures.

This objective was pursued through two principal methodological vectors: first, by applying system-analytical methods, generalisations, and logical inferences within the frontier of valid neuroscience-related theories and practices; and second, by employing modelling and prognostic approaches alongside extrapolations of theoretical insights to the core transformation target — verification, validation, and optimisation of agent-oriented control systems based on neural networks. To propose our own, currently theoretical-methodological model, we conducted a revision and projection of existing methods: probabilistic, neurocognitive, and others, and applied them systematically to outline future directions requiring empirical validation.

The design and theoretical-methodological foundation of the article are informed by recent practice-oriented studies that reflect a marked increase in interest towards the formal verification of neuro-controlled robotic systems. Of particular relevance is the work of a European research group (Araujo et al., 2023), which conducted a comprehensive systematic review of modern V&V approaches for autonomous agent systems, highlighting the limitations of traditional techniques in rapidly changing environments. Equally valuable are the contributions of a UK-based team led by Fisher et al. (2021), who identified practical challenges in verifying inspection robots, where error-free operation under high-risk conditions is critical. In this regard, the VALU3S project (Verification and Validation of Automated Systems' Safety and Security, 2020–2023) serves as an illustrative case of comprehensive integration of computer vision and generative models into formal frameworks to ensure the reliability of autonomous systems in industrial settings, providing over 40 specific tools (VALU3S Consortium, 2021).

Attention was also drawn to a recent study in which the authors proposed a Bayesian approach to real-time verification of autonomous robots (Zhao et al., 2024). This demonstrated the possibility of using probabilistic models to adaptively assess the behaviour of agents during operation.

The article is particularly relevant for methodologically addressing the problem and risks of fuzzy neuro-oriented decisions in urgent situations, namely those decisions that mimic natural cognition and behaviour. Of particular relevance is the understanding and conceptual modelling of Spiking Neural Networks (SNNs).

We also hope that our findings will refine and concretise current approaches to safety assurance in multi-agent systems, especially in the context of dynamically changing environments and evolving inter-agent interaction policies.

4. Review of Neuroscientific Advances Valid for use in Agent-Oriented Neural Network Control of Robotic Systems

Recent advances in neurobiology and engineering regarding the development of automated decision-making and control systems have unveiled new possibilities for the design, programming, and self-learning of agent-based robotic systems. Among the most significant are the dynamics of signal processing in spiking neural networks (Niu et al., 2023). The practical "human-like" qualities of these systems, such as sensitivity, resolution, and cognitive flexibility, have been enhanced through the incorporation of neuroplasticity functions. This has significantly accelerated signal

processing. Thanks to neuroplasticity, highly adaptable models have become possible; models that can be formalised, programmed, and are expected to provide responsive feedback.

On the other hand, neurophysiological research on brain functions related to memory needs to be projected onto multi-agent systems for autonomous navigation in environments with obstacles and solving complex tasks (Patai & Spiers, 2021). The development of agent-driven and neuro-controlled systems with efficient decision-making processes, transparent for validation and safe deployment, should be facilitated by progress in explanatory neuroscience.

We are encouraged by the fact that important works in this area have recently been published, the most notable of which is the book "Integrating Artificial and Human Intelligence through Agent-Oriented Systems Design" (Miller & Rusnock, 2024). It demonstrates the feasibility of creating hybrid agent architectures that combine the capabilities of artificial intelligence with adaptive cognitive models unique to humans. According to the authors, fruitful interaction between a human and an autonomous agent is only possible under specific anthropomorphic conditions. Thus, an agent-oriented system should, and already can, imitate and adapt to human cognitive methods, emotional responses, and decision-making processes in conditions of high informational entropy and multitasking.

It is clear that traditional linear algorithms are insufficient for these tasks. Therefore, the development of next-generation agent-oriented systems is based on a neuropsychological, multi-level, and not yet fully defined interaction, akin to human sensitivity, responsiveness, and insightfulness. The development of new management policies and protocols, both prospectively and in early implementations , must be based on bio-inspiration, which yields high efficiency and resilience to obstacles and excessive entropy (Abdalla & Mishra, 2021).

Currently, we have several modalities (profiles) for applying agent-oriented neuro-control to robotic modules. These modalities are represented by three approaches. The first is valid for controlling systems in open spaces (e.g., in agriculture or forestry). An agent-oriented system of this type must be built as a multi-level organisation with a defined protocol of possible situational scenarios, behavioural patterns, and intervention roles (Andy et al., 2025). This model retains many traditional and well-proven techniques. While neuro-intervention here is only possible fragmentarily, the system is noted for its high reliability and serves as an example of a clear, extended architecture.

A more promising approach is a system for simultaneous mass control of simpler intelligent mechanisms (e.g., robots, drones, etc.) (Ganesan et al., 2021). It can, via satellite communication, manage elements across a large area using deep learning. Here, the agent-oriented approach is complemented by flexible neural network policies, Sequence-to-Sequence (Seq2Seq) models with attention mechanisms, Pointer Networks, and Transformer architectures for the relative autonomy of controlled elements (e.g., a drone selecting its personal route within the overall mission). The intelligent collaboration of drones with a "queen" agent brings the system closer to genuinely bio-inspired models. Risk prediction and adaptability to changes in the operational context are the main advantages of such agent-oriented control. It is clear that current software is not yet mature or proven enough to apply neuro-adaptive practices fully.

An example of the third type of system is the achievement of Finnish engineers, who integrated agent-oriented algorithms with the neuro-symbolism of artificial intelligence (Pakkala et al., 2025). Such systems are capable of learning as a situation unfolds and can improvise or reproduce valid behaviour. In our view, this realises the aspiration to formalise knowledge through neuro-symbolic representations and to provide verification procedures through ontologically-controlled agent behaviour.

We can draw conclusions regarding these types. Pakkala et al. (2025) investigate how to create cognitive and ontological support for safe situational control, while Ganesan et al. (2021) investigate how to increase the flexibility of neural networks to enable collective dynamic control. Andy et al. (2025) explore how to organize agent systems for spatial management. Thus, with each proposed model we have illustrated, the quality of explainability, formalised verification, and

neuro-integration increases. This enables the development of next-generation hybrid systems in complex robotic environments with the highest degree of transparency and trust.

Based on the analysis of the research, we can summarise: the studied methods of agent-oriented control of robotic systems are multimodal, multifunctional, and have the potential for use in modern processes of verification, validation, and optimisation. This experience demonstrates how deep neural network regulation, cognitive-ontological procedures and structural modelling can work in concert and ensure high adaptability of agents. The ability of these systems to process several types of signals demonstrates multimodality, and their simultaneous processing of navigation, recognition, planning, and coordination of groups with simultaneous adaptation to new goals demonstrates multifunctionality. By combining the previously mentioned methods, it is possible to develop a new generation of agent-oriented, neural-controlled robotic systems with a high guarantee of reliability and validation efficiency.

5. Prospects for Optimising Agent-Oriented Control of Robotic Systems Based on Neural Networks

It is observed that increasing attention is being paid to the use of the principles of spike activity in neural networks (spiking neural networks - SNNs), which allow the achievement of greater biological plausibility in modelling agent behaviour and provide high temporal accuracy of decisions, which is critically important for autonomous robots. Experimental research in the field of neuro-behavioural adaptation is increasingly applying Bayesian and probabilistic models for practical neuro-application (Nicenboim et al., 2021) and for predicting changes in agent policies in real time. This enables dynamic assessment of reliability levels and adjustment of parameters during the operational phase.

However, for the modelling of an original neurocognitive-ontological agent-oriented system with multi-level verification and validation, multiscale cognitive models will be of great importance. These models combine neuro-symbolic representations with empirically studied patterns of sensorimotor learning (D'Angelo & Jirsa, 2022). Clearly, they can be used as a basis for standardised assessments of agent behaviour in uncertain situations. In addition, efforts are currently being made to develop explanatory artificial intelligence methods. These methods leverage known neurocognitive patterns to explain, discover, and verify solutions to the systems being analysed. This incorporates the neurodidactic principle: agent-based systems change because they learn. Therefore, flexible models based on cause-and-effect algorithms are appropriate.

Moreover, modern neuro-visualisation techniques, including the virtual reconstruction of cognitive maps, can be applied to create test environments for verifying the consistency of an agent's decision-making with biologically grounded cognitive strategies.

Recently, combined cognitive models have been published (Fortino et al., 2017). They integrate attention, memory, and planning systems, the effectiveness of which has been empirically proven. Therefore, there are opportunities to develop agent-oriented management to the level of systems that are understandable, formally verified, and flexible in practice.

These trends demonstrate that the future of verification and validation in neural agent systems lies in the close collaboration of neuroscience, cognitive modelling, and next-generation machine learning. We propose a generalised systematisation of neural network systems (Table 1).

Table 1. Neural Network Models for Agent-Oriented Control

Model/method name	Brief description	Prospects and limitations
Spiking Neural Networks (SNN)	Can reproduce/simulate neuroactivity in detail based on high-frequency biopulses.	The lack of testing of such programmes is compensated by their economy, resolution, and high accuracy.
Bayesian models	Predict, estimate, and modulate agent behaviour in real time	Can verify a significant amount of current data, but are resource-intensive
Multi-scenario models	They take into account several neural symbols at once, switch between possible scenarios and learn based on sensorimotor skills.	Although they are very explicit and easily verified, their flexibility is low.
Artificial intelligence based on neuroanalog models	Determines cause-and-effect relationships, patterns; effectively learns during control and explication of agent decisions	Effective, generative, but requires innovative programmes, which are lacking.
Neurovisualisation as the basis of machine vision	Integration of neuroimaging devices and special digital platforms is able to detect and test the problem and the decision made according to the algorithms of the human brain.	High resolution, evaluation ability, but the need for resources and hardware complexity
Explainable AI Inspired by Neuroscience	Focus on causal relationships to explain agent decisions and ensure transparent validation.	Promising for safe implementation, but current tools are not fully mature for complex systems.
Neurovisualisation-Based Testing Environments	Virtual spaces for validating the cognitive consistency of agent decisions with biologically grounded strategies.	Provide quality testing of cognitive patterns but are limited in modeling real multi-agent scenarios.
Hybrid (combined) models Objectify neurocognitive data for their further “understanding” and use by machine agents	Objectify neurocognitive data for their further “understanding” and use by machine agents	Despite high divergence and versatility, they are not provided with sufficiently effective software.

We have taken into account the existing achievements and identified patterns. Now we present a new concept of a neurocognitive-ontological agent-oriented multi-level verification and validation system. We will call it Neuro-Agent Verified Controller (NAVC), based on our own experience and knowledge, as well as on the study of the above-mentioned neuroscientific methodologies.

In our opinion, the basis of the descriptive structure of such a model should be three main components:

1. The neurobehavioural block will include adaptability based on the Bayesian model and spike neural networks (Ganesan et al., 2021).

2. The control block is neurosymbolic scenarios that should correlate with highly contextual capabilities and which can already be programmed into formal algorithms (Pakkala et al., 2025).

3. The digital duplication block will be implemented through a complementary distribution of virtual model functions (Kovtunenko et al., 2025), which includes computation and intelligent response of an agent-oriented system.

Our NAVC model offers multi-level verification:

a) at the cognitive level via ontological validation;
b) at the neuro-adaptive level through continuous Bayesian reliability assessment;
c) at the operational level through simulation comparison with digital twins.

It is important to note that the proposed system actively employs the Ann-Arbor architecture (Dong, 2025), integrating natural language and formal queries into agent logic. The system requires a human operator, but with minimal involvement: the multifunctional interface requires only the most general task direction and adapts and deploys the activity scenario itself.

A distinctive feature of NAVC is its built-in mechanism for cognitively grounded optimisation, whereby the agent not only performs actions but also analyses their effectiveness from the standpoint of cognitive planning and decision-making patterns, verified within the digital twin before real-world deployment.

Although the model has not yet undergone empirical testing, it is considered it promising and capable of ensuring:

- High transparency and safety in critical domains such as transportation, energy, and defence.
- Flexibility and scalability in multi-agent distributed systems.
- Interpretable adaptability – agents can learn in real time while maintaining transparency in their decision logic.

It is anticipated that with a successful hardware and software implementation, NAVC will become a useful practically oriented concept, and its general characteristics are presented (Table 2).

Table 2. General characteristics of the author's model NAVC (Neuro-Agent Verifiable Control)

Model component	Essence	Functional advantages
Neurobehavioural module	The model adapts during use and the neural basis grounded in impulses (spikes) is dynamic, self-generating, and self-improving.	Adapts quickly, identifies opportunities, and is resilient in changing conditions.
Ontologically-Controlled Layer	Formal cognitive model based on neuro-symbolic representations and situational context ontologies.	Ensures decision explainability, formal verification, compliance with high-safety scenarios.
Digital Twin Layer	Simulation of agent actions in a distributed environment based on digital twins, relying on agent-oriented ABSynth platforms.	Enables pre-deployment verification and optimisation, fault tolerance, scalability.
Agent-Language Ann-Arbor Architecture	Context-sensitive integration of natural language and formal queries into agent logic using the Postline approach.	Universal human-operator interface, enhanced explainability, flexible programming via multilingualism.
Cognitively Grounded Optimisation	Self-analysis of agent decisions based on cognitive patterns and evaluation of their effectiveness in the digital twin before real-world application.	Optimises behavioural strategies, increases reliability, minimises errors in complex and unpredictable environments.

We predict and expect that our NAVC model (subject to valid experimental verification) will have high practical potential for implementation in complex robotic systems, particularly in

autonomous transportation, manufacturing automation, and defence technologies. It is forecasted that in the coming years, the development of hardware support for spiking neural networks and multi-agent platforms will significantly simplify the practical implementation of NAVC in industrial and mission-critical applications.

While the author's concept is a proposal and has not been tested in practical application, we offer formal evidence to demonstrate the feasibility of the key components of the architecture of the presented NAVC model.

Thus, if an agent operates in a dynamic environment with state uncertainty, incomplete sensory data, and variability of its own policies, then no single control level can fully guarantee correct behaviour. Therefore, a compositional multi-level verification mechanism is needed. Let π be the agent's policy, and X(t) represent the evolution of its states. According to the study, bioinspired spike dynamics in the SNN module gives an approximation of X(t) with low temporal error. However, this approximation does not provide cognitive consistency or compliance with the security scenario. Therefore, we introduce an ontological level that performs the mapping $\pi \mapsto \pi\Omega$, where Ω is a set of formal cognitive-security constraints specified in the article as necessary for "explanation" and formal validation of the agent's choice (see the sections on ontological representations and neurosymbolic models).

The use of neural networks in critical areas such as autonomous driving, tumour detection, and power plant control is becoming increasingly common, and errors in their operation have increasingly dangerous and costly consequences. This raises the problem of confirming the reliability of the obtained results – the problem of formal verification of controllers based on neural networks.

Traditionally, verification of neural networks has focused on testing – evaluating the network on a large set of points in the input space and determining whether its outputs correspond to the desired values. However, it is impossible to test all possible inputs, since the input space is often effectively infinite in size.

Formal verification provides an orthogonal alternative to testing; it offers mathematical guarantees that the artificial neural network will work correctly for any input from the space of possible inputs.

This article presents a study of methods for formal verification of artificial neural networks. During the study, the necessary mathematical apparatus for the formal description of the topic was formulated, and the most effective and stable algorithms for formal verification of neural networks were also identified. A more detailed analysis is provided of the mathematical methods underlying the defined formal verification algorithms, based on the reachability property of each neuron, and the selected methods are systematised.

There are four main approaches to the verification of artificial neural networks:

A. SAT/SMT solution

The neural network and the system of imposed constraints are presented in the form of a Boolean formula and can be expressed in conjunctive normal form. Next, the generated formula is checked by the automatic SAT solver. SMT is a variant of SAT for k-valued logic. The presence of a SAT output means that a negation of the property has been found – a counterexample. UNSAT means that there is no counterexample; in other words, the property is satisfied.

B. Linear Programming

When using linear programming (LP), the system of imposed constraints is defined as a set of linear constraints (a system of linear equations and inequalities), and the property is defined as an objective function that needs to be maximised or minimised depending on the conditions of the problem. Verification is performed automatically using linear programmers.

C. Proving theorems

The system of imposed constraints is presented as a model governed by mathematical principles, the property of which is the goal of the proof. The principle of the approach is to use axioms and rules to check whether the properties of the created model correspond to the original system.

D. Incomplete verification

This method is based on creating a simplified model of a given system with the same properties. The resulting model is not an exact reflection of the real system, but is an over-approximation. The created model is verified using other approaches.

Applied Mathematical Methods

The most common methods for verifying neural networks are based on the use of SAT/SMT and LP solvers. Therefore, the main task when creating a verification algorithm is to transform the given constraints into a form accessible to the solver. More precisely, for SAT/SMT solvers, this transformation converts the given constraints in the selected problem, usually represented by a segment, into symbolic representations connected by conjunctions or disjunctions, and for LP– into a system of linear inequalities. The neural network itself is then represented as a model that preserves all the properties of the neural network, but operates with specially introduced symbolic representations that differ for each approach.

The most common approach to verification is a combination of incomplete verification and the use of SAT/SMT solvers, or incomplete verification and LP.

The purpose of neural network verification is to check its performance on the entire possible set of input data. Input data are vectors, points in a multidimensional space. However, in this problem, we need to think differently - to "expand" the points to the entire possible input space, which represents a certain area, and estimate the size of the output data that the neural network can produce after applying them to any element in this area. In other words, the task is to transform the problem from discrete to continuous, preserving all the properties of the neural network and the one-to-one correspondence between the input and output sets. However, when expanding the input data area and further applying the neural network to it, the boundaries of the output set also expand, and the task arises in developing methods for narrowing the boundaries of the output set, preserving all the connections and traceability of transitions between neurons. The key idea of most of the applied methods is to go from vectors to functions, and to do this at each step: to evaluate the input and output sets for each neuron and to measure the results of applying the activation function. One of the most effective methods in this field is interval analysis and its analogues.

A. Symbolic interval analysis

Symbolic interval analysis is based on the following idea: to replace each input interval with a variable (symbol), then to perform one step through the neural network to obtain symbolic intervals (intervals containing symbolic variables), and then to repeat the iterative process. This idea is used in all available algorithms and is the basis of the ReLUVal algorithm.

Let us consider different approaches to interval estimation: suppose we have two airplanes and a neural network that estimates the behaviour of one of them using the distance to the other and the possible angle of rotation.

Input data: distance from the second airplane $\in [4; 6]$, the possible angle of rotation of the second aircraft is $\in [1; 5]$. Input: the safe angle of rotation of our aircraft must be less than 20°, i.e. $\in [0;20]$.

The generally accepted mathematical approach to working with intervals — (a) naïve interval expansion: the left boundary of the interval consists of transformations of the left boundaries of each interval, and the right boundary consists of the corresponding transformations of the right boundaries — leads to results that lie outside the permissible limits of the original set.

Consider the symbolic approach (b) Symbolic interval expansion, based on replacing intervals with symbolic variables. The activation function of the neural network is applied to symbols, not to numerical values. At each step, linear transformations are performed on the symbols. The output is linear combinations of the input intervals in symbolic form; the initial numerical boundaries of the intervals are substituted, resulting in an interval that satisfies the given constraints. Mathematical reasoning: Let x be a real value. It can be represented by the degenerate interval [x; x].

The interval representation of the function f(x) is a new function F such that applying F to the interval [x; x] gives the same result as applying f to the value x:

F([x; x]) = f (x) y x ∈ X. In other words, we can now move on to symbolic arithmetic, applying functions to degenerate intervals of the form [x; x] and to standard intervals of the form [x; y].

Consider example (b): a number x is chosen from the interval X = [4; 6] and represented as the interval [x; x] with two equal limits. Similarly, for the interval Y = [1; 5], the element y is represented as [y; y]. The neural network function f, which operates on elements x and y, is replaced by F, which operates on intervals [x; x] and [y; y], at each step multiplying the limits by the given weights and again obtaining intervals of the following form: [2x + 3y; 2x + 3y], [x + y; x + y].

The initial result is the interval [x + 2y; x + 2y], and given that x ∈ [4; 6] and y ∈ [1; 5], the usual method of working with intervals is applied: for the left limit x = 4, y = 1 the result is 6, for the right limit x = 6, y = 5 the result is 16. Thus, the initial set satisfies the given constraints [6; 16] ⊆ [0; 20].

In general, in addition to multiplying by weights at each step of the neural network, an activation function is applied.

ReLU is used in all major works on formal verification of artificial neural networks.

Let r' = Ed = [l; u] be a symbolic representation of the interval obtained after multiplication by weights at an intermediate neuron. Then, after applying the ReLU activation function to obtain the output value r = ReLU(r'), several possible results can be obtained: r = [l; u], if u > 0, all values in the interval are positive, the interval itself remains unchanged, the neuron is activated; r =[l; u], if l < 0 < u is an interval that cannot be converted into an interval by the ReLU function, further research is needed.

The problem arises in the third case, when the interval boundaries are located on different sides of zero. Recall that the result of applying the ReLU function is always non-negative. Type 3 neurons will be called overestimated, and the main task of the verification algorithm is to determine how to process the result and in what form to transfer it to the next neuron. In this case, the boundaries of the interval [l; u] are called pre-activation, since they do not give an exact answer as to whether a given neuron will be activated.

We focus the reader's attention on the main problem of the reachability of a formal verification algorithm for neural networks. After applying the ReLU activation function, the result must be obtained in a form that can be transferred to the next neuron. This result must be expressed in terms of two “objects” (numbers or symbolic representations) — the endpoints of the segment [lI; u] — which are passed to the ReLU activation function of the corresponding neuron. Since our algorithms must adhere to the reachability principle for each neuron, we need to understand what is happening on each neuron, what is coming into the input, and what is coming out of the output after applying the ReLU function. The idea is to preserve as much information as possible without inflating the resulting value due to over-approximation.

There are three parameters: the input value i, which is written as a symbol, the boundaries of the segment I, and, which evaluates it. Our task is to build a “smart” set using these parameters, which can then be passed to the next neuron. To multiply this set by the weights, we apply the ReLU function again and get a meaningful result. B. Iterative refinement

The simplest idea for working with over-approximated neurons is to preserve the maximum possible variance and replace one of the boundaries with zero. This is the second component of the formal verification algorithm ReLUVal.

B. Mathematical justification

Let the first layer contain neuron A, where x1 is the input data, and w1 is the corresponding weight. Then the boundaries of the symbolic interval will be the same and equal:

Ed^p(X) = Ed^p(X) = W1 X1 + • • • + W1 xa.

If neuron A lies in an intermediate layer, then the task of preserving the maximum variance arises– namely,the largest possible interval length. We define W+ and W- as positive and negative weights in the current layer, and Ed^p'" and Ed^'" - as the boundaries of the interval originating from the previous layer. Then for the lower and upper boundaries we obtain the estimate:

Ed^ (X) = W+ Ed^'" (X) + W- Ed*G'" (X) = [w_l; w_i] = [Eqiow(X); Eqlau,(X)]

EqAp(X) = W+ EqAprev(X) + W_ Eq£z"(X) =

lup = [w+1; w+1] = [Equp(X); Equp(X)]

It turns out that for the maximum length of the interval it makes sense to take the interval [Ed[osh(X); Edir(X)], and then the following options are possible:

1) if Ed0sh(X) > 0, simply take the given interval and go to the next layer;

2) if Ed0sh(X) < 0, the lower bound is defined as 0 — take the interval [0; Edir(X)] and consider the upper bound separately;

3) if Edir(X) > 0, then go to the next layer along the interval from the previous point;

4) if Edir(X) < 0, then the interval is defined as [0; 0] and the transition to the next layer is performed.

C. Symbolic linear relaxation

Symbolic linear relaxation involves approximating the ReLU activation function by a convex polygon, defining it using linear inequalities for symbols and writing the result of the application again in symbolic form. In essence, this is an extension of symbolic interval analysis to a system of linear constraints, rather than to a single interval.

Mathematical justification:

In the usual approach, we fit the result of applying the activation function into a horizontal strip, but this is not the most optimal option in terms of approximation efficiency. A more efficient approach is to fit it into a parallelogram:

Using Method 1, we have already determined the lower and upper bounds of the interval obtained in the previous layer, denoted as

E_d = [l, u], while its concrete evaluation is represented as E_{dir}. These bounds define the feasible range of the pre-activation values and serve as the basis for constructing the linear relaxation of the ReLU function.

In summary, applying symbolic methods to intervals and activation functions reveals a connection to concepts from geometric probability. The task is to construct a continuous set containing all the input data, but with the smallest possible area.

However, due to hidden contradictions in the environment itself, even a conceptually correct policy πΩ may prove suboptimal in real situations. Therefore, the simulation mapping πΩ↦πΛ is performed by the third level, namely digital twins, where πΛ is the policy that the system considers appropriate after tests. Thus, the policy within the set of potential paths of environment evolution is evaluated by this simulation mapping, which constitutes a reliable method of assessing the agent's behaviour under changing situational circumstances. The NAVC concept is therefore necessary and minimally complete to ensure correct behaviour in the context of further validation, verification and adaptive optimisation. This formally justifies the feasibility and inevitability of the three-component architecture. This is explained by the fact that each of the three levels eliminates different classes of

errors (temporal – SNN; semantic – ontological control; and operational – digital twins), and none of the levels is functionally equivalent to the others.

6. Conclusions

Based on a careful interdisciplinary integration of developments in neuroscience, agent programming, and machine learning, the authors put forward the theoretical idea of agent-oriented neural network control as a multi-component concept. It outlines a new broad methodological framework in which agent-oriented systems are viewed as transparent, self-adaptive, and cognitively aware information-oriented structures, rather than simply technical solutions. Particularly important are mechanisms derived from neuroscience: spiking neural networks capable of providing biologically plausible temporal dynamics of decisions; Bayesian probabilistic models that allow agents to make decisions considering uncertainty and continuously update behaviour policies in real time; and explainable AI, which ensures transparency and interpretability of the logic behind agent actions for human operators. It is important that the neuroscience-based approach lays the foundation for creating multilevel, adaptive, cognitively justified systems where planning, learning, and self-assessment of agent actions become interrelated and formally verified. Owing to the broad integration of neuroscientific concepts, multi-agent technologies, and modern verification methods, the proposed methodological foundation creates a new pathway for the development of agent-oriented control in robotic systems.

This article presents an overview of recent developments in neuroscience, demonstrating how basic neurobiological processes and models have direct application in the development of agent-oriented neural network control systems. It demonstrates that multi-agent systems can be successfully designed using biological mechanisms that support spatial navigation, conflict resolution, dynamic planning, and situational adaptation in animals. This will improve the systems' flexibility, stability, and ability to make autonomous decisions in complex and variable environments. Specifically, neuroplasticity, bio-inspired principles of attention allocation, and neuro-symbolic integration enable the creation of agents capable of learning and adapting while maintaining openness and formal verification of their behaviour.

The creation of neurocognitively grounded agent systems is both theoretically and practically feasible with modern technologies, according to recent research in explainable neuroscience and multi-scale cognitive modelling. As per the review's conclusions, the careful integration of neurobiological principles can form the basis for a new generation of adaptive, transparent, and safe robotic systems, which demonstrate that the application of neuroscience in agent-oriented control is far more widespread than previously assumed.

Forecasting the future of agent-oriented neural network control systems indicates that their advancement is directly linked to the integration of cognitive-ontological models, explainable and valid artificial intelligence, and neuroscientific methods. These models allow for real-time modification of agent behaviour for its improvement, as well as its verification. The most promising are multi-scale modelling approaches that combine the speed of spiking neural networks with the cognitive grounding of neuro-symbolic representations and probabilistic (Bayesian) structures for realistically accounting for environmental uncertainty.

Based on these approaches, the outcome of the methodological work is the author's NAVC (Neuro-Agent Verifiable Control) concept, developed in this study, which represents a significant step towards a new paradigm of agent-based control. This paradigm is founded on cognitive-neural explainability, multi-level verification, and formalised optimisation during both the design and operational phases. Although the NAVC model has not yet been empirically verified in physical robotic complexes, its architecture meets the requirements of modern security, operational and validation standards for agent-oriented systems.

Following the stages of software implementation, digital modelling and simulation testing, the expected high efficiency of the model is confirmed by its methodological consistency and systematic construction. In short, the results of this study offer a theoretical and practical basis for

creating a new class of intelligent agent systems with improved explainability, reliability, adaptability and security. We believe this will allow them to function well in the dynamic and uncertain robotic environment. Through formal arguments, we have demonstrated the concept's feasibility.

7. Prospects and Limitations of the Research

Although the proposed NAVC model is based on sound theory and methodology, it has not yet been implemented and remains a theoretical concept. The primary limitation is the lack of experimental validation, which precludes an assessment of its true effectiveness in complex multi-agent scenarios with high levels of dynamic change and adaptability.

References

Abdalla, R., & Mishra, A. (2021). Agent-oriented software engineering methodologies: Analysis and future directions. *Complexity, 2021*(1), 1629419. https://doi.org/10.1155/2021/1629419

Andy, Y. W. C., Shiang, C. W., & Paschal, C. H. (2025). Agent-oriented modelling and simulation for robotic based predator control. *JOIV: International Journal on Informatics Visualization, 9*(2), 607–616. https://doi.org/10.62527/joiv.9.2.3347

Araujo, H., Mousavi, M. R., & Varshosaz, M. (2023). Testing, validation, and verification of robotic and autonomous systems: A systematic review. *ACM Transactions on Software Engineering and Methodology, 32*(2), 1–61. https://doi.org/10.1145/3542945

Arkin, R. C. (1998). *Behavior-based robotics*. MIT Press.

D'Angelo, E., & Jirsa, V. (2022). The quest for multiscale brain modeling. *Trends in Neurosciences, 45*(10), 777–790. https://doi.org/10.1016/j.tins.2022.06.007

Dong, W. (2025). *The Ann Arbor architecture for agent-oriented programming*. arXiv. https://doi.org/10.48550/arXiv.2502.09903

Fisher, M., Cardoso, R. C., Collins, E. C., Dadswell, C., Dennis, L. A., Dixon, C., Farrell, M., Ferrando, A., Huang, X., Jump, M., Kourtis, G., Lisitsa, A., Luckcuck, M., Luo, S., Page, V., Papacchini, F., & Webster, M. (2021). An overview of verification and validation challenges for inspection robots. *Robotics, 10*(2), 67. https://doi.org/10.3390/robotics10020067

Fortino, G., Gravina, R., Russo, W., & Savaglio, C. (2017). Modeling and simulating internet-of-things systems: A hybrid agent-oriented approach. *Computing in Science & Engineering, 19*(5), 68–76. https://doi.org/10.1109/MCSE.2017.3421541

Franklin, S., & Graesser, A. (1997). Is it an agent, or just a program?: A taxonomy for autonomous agents. In J. P. Müller, M. J. Wooldridge, & N. R. Jennings (Eds.), *Intelligent agents III: Agent theories, architectures, and languages* (Lecture Notes in Computer Science, Vol. 1193). Springer. https://doi.org/10.1007/BFb0013570

Ganesan, R. G., Kappagoda, S., Loianno, G., & Mordecai, D. K. (2021). *Comparative analysis of agent-oriented task assignment and path planning algorithms applied to drone swarms*. arXiv. https://doi.org/10.48550/arXiv.2101.05161

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of the ACM, 44*(4), 35–41. https://doi.org/10.1145/367211.367250

Kovtunenko, A., Freint, S., Shundeev, A., Vaganova, D., & Bagautdinov, S. (2025). Development of distributed digital twins of complex technical objects based on an agent-oriented software architecture. In V. M. Vishnevskiy, K. E. Samouylov, & D. V. Kozyrev (Eds.), *Distributed computer and communication networks* (Communications in Computer and Information Science, Vol. 2484). Springer. https://doi.org/10.1007/978-3-031-89211-0_10

Li, J., Rombaut, E., & Vanhaverbeke, L. (2021). A systematic review of agent-based models for autonomous vehicles in urban mobility and logistics: Possibilities for integrated simulation models. *Computers, Environment and Urban Systems, 89*, 101686. https://doi.org/10.1016/j.compenvurbsys.2021.101686

Miller, M. E., & Rusnock, C. F. (2024). *Integrating artificial and human intelligence through agent oriented systems design*. CRC Press. https://doi.org/10.1201/9781003428183

Nicenboim, B., Schad, D. J., & Vasishth, S. (2021). *Introduction to Bayesian data analysis for cognitive science*. Chapman & Hall/CRC.

Niu, L. Y., Wei, Y., Liu, W. B., Long, J. Y., & Xue, T. H. (2023). Research progress of spiking neural network in image classification: A review. *Applied Intelligence, 53*(16), 19466–19490. https://doi.org/10.1007/s10489-023-04553-0

Pakkala, D., Känsäkoski, N., Heikkilä, T., Backman, J., & Pääkkönen, P. (2025). On design of cognitive situation-adaptive autonomous mobile robotic applications. *Computers in Industry, 167*, 104263. https://doi.org/10.1016/j.compind.2025.104263

Patai, E. Z., & Spiers, H. J. (2021). The versatile wayfinder: Prefrontal contributions to spatial navigation. *Trends in Cognitive Sciences, 25*(6), 520–533. https://doi.org/10.1016/j.tics.2021.02.010

Rosenblatt, J., & Payton, D. (1989, June). A fine-grained alternative to the subsumption architecture for mobile robot control. In *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks* (Vol. 2, pp. 317–322). https://doi.org/10.1109/IJCNN.1989.118717

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Sreekanth, N., Rama Devi, J., Shukla, K. A., Mohanty, D. K., Srinivas, A., Rao, G. N., Alam, A., & Gupta, A. (2023). Evaluation of estimation in software development using deep learning-modified neural network. *Applied Nanoscience, 13*, 2405–2417. https://doi.org/10.1007/s13204-021-02204-9

VALU3S Consortium. (2021). *Verification and validation of automated systems' safety and security*. EU ECSEL JU Project VALU3S. https://cordis.europa.eu/project/id/876852

Wan, G., Dong, X., Dong, Q., He, Y., & Zeng, P. (2022). Design and implementation of agent-based robotic system for agile manufacturing: A case study of ARIAC 2021. *Robotics and Computer-Integrated Manufacturing, 77*, 102349. https://doi.org/10.1016/j.rcim.2022.102349

Weiß, G., & Ciancarini, P. (2002). *Agent-oriented software engineering II* (Lecture Notes in Computer Science, Vol. 2222). Springer.

Zhao, X., Gerasimou, S., Calinescu, R., Imrie, C., Robu, V., & Flynn, D. (2024). Bayesian learning for the robust verification of autonomous robots. *Communications Engineering, 3*(1), 18. https://doi.org/10.1038/s44172-024-00162-y