Scientific Research Publishing

Advances in Artificial Intelligence and Robotics Research, 2026, 2(2), 78-90
https://www.scirp.org/journal/airr

Theory of Higher Categories as a Metalanguage of Strong Artificial Intelligence

Mikhail Gennadievich Shcherbakov

Institute of Legislation and Comparative Law, Moscow, Russia
Email: pravovednalog777@mail.ru

How to cite this paper: Shcherbakov, M.G. (2026) Theory of Higher Categories as a Metalanguage of Strong Artificial Intelligence. *Advances in Artificial Intelligence and Robotics Research*, **2**, 78-90.

Received: January 11, 2026
Accepted: May 30, 2026
Published: June 2, 2026

Abstract

This article discusses the architecture of a strong artificial intelligence neural network. The topic of strong artificial intelligence is one of the most pressing, and various specialists are working in this area. The research gap stems from the lack of a unified theoretical approach to understanding the model and language of strong artificial intelligence and the systematization of AI development in Russia. The aim of this study is to examine theoretical approaches to the concept of a strong artificial intelligence metalanguage. The scientific novelty of this study lies in the mathematical approach to describing the strong artificial intelligence model. The author's hypothesis is to describe the strong artificial intelligence model using the language of higher-order theory. The essence of this approach is to describe not objects, but their morphisms, functors, and natural transformations. The author notes that the construction of a strong artificial intelligence neural network is based on two axioms of the theory of higher categories: identity and composition. The author provides examples of using the methodology of higher category theory to describe the biological neural network of the human brain, in which neuron weights represent not quantitative values, but connections and their transformations. In the article, the author draws an analogy between the biological neural network of the human brain and the neural network of a strong artificial intelligence capable of solving creative problems like a human, making associations, and learning from limited data. This study utilizes a systems-methodological approach, a hypothetical-deductive approach, and the methodology of the theory of higher categories. In conclusion, we can conclude that the theory of higher categories, based on the axioms of identity and composition, describing morphisms, functors, and natural transformations, is a universal metalanguage for describing the mathematical model of strong artificial intelligence. In other words, the theory of higher categories is a metalanguage of strong artificial intelligence, which allows one to see structure where others see a set of data, to build abstractions of any level, to transfer knowledge based on the similarity of structures, to reflect on one's own cognitive processes, and to use structural identities to coordinate the structure of thinking.

Keywords

Higher-Category Theory, Strong Artificial Intelligence, Morphism, Functor, Natural Transformation, Neural Network, Mathematical Model, Metalanguage

1. Introduction

Currently, technologies based on narrow artificial intelligence, capable of solving a limited set of problems, are actively used.

Meanwhile, research is underway to create strong artificial intelligence capable of solving creative problems, processing incomplete data, and quickly learning from limited data.

It is important to note that, to create strong artificial intelligence, it is necessary and sufficient to develop a comprehensive theory of strong artificial intelligence.

Unfortunately, there is currently no theoretical basis for creating a mathematical model of a strong artificial intelligence neural network that would describe various mathematical approaches to describing a strong artificial intelligence neural network. Consequently, practical developments in the field of strong artificial intelligence may not lead to the desired result—the creation of a strong artificial intelligence neural network model.

This article presents a scientific concept on the possible use of higher-category theory as a metalanguage for describing a strong artificial intelligence neural network model.

2. Methods

The conducted study uses a systemic-methodological approach, a hypothetical-deductive approach, and the methodology of the theory of higher categories.

3. Result and Discussion

What is a metalanguage?

A metalanguage is any second-level natural or artificial language used to describe a first-level language.

In our view, higher-category theory, which possesses the highest degree of abstraction, is a second-level language relative to other mathematical theories describing neural networks: graph theory, topology, complex systems theory, mathematical analysis, and so on.

Thus, higher-category theory is the metalanguage of strong artificial intelligence.

This study draws an analogy between the description of the biological neural network of the human brain and the neural network of strong artificial intelligence using the methodology of higher-category theory.

What is strong artificial intelligence?

Artificial intelligence is a technology capable of simulating human cognitive activity (thinking).

P.M. Morhat notes that artificial intelligence is a fully or partially autonomous, self-organizing computer-hardware-software, virtual or cyber-physical, including biocybernetic, system endowed with/possessing the abilities and capabilities of thinking, self-organization, learning, independent decision-making, etc. [1].

Depending on their capabilities, several types of artificial intelligence are distinguished: narrow, strong (general), and super artificial intelligence.

This article explores the use of higher category theory methodology to describe strong (general) artificial intelligence.

Strong artificial intelligence (AGI) is the result of improving narrow artificial intelligence [2].

Artificial general intelligence is an autonomous, self-organizing system capable of performing creative tasks like a human [3].

We propose that strong (general) artificial intelligence is a digital entity based on a neural network that functions according to the principles of a biological neural network and possesses creative abilities similar to those of a human.

In other words, the neural network of strong artificial intelligence is based on the operating principles of the biological neural network of the human brain.

In our opinion, a direct analogy can be drawn between the description of the biological neural network of the human brain and the neural network of strong artificial intelligence.

What is a neural network? A neural network is a mathematical model based on the organizational principles of human neural networks.

The basis of an artificial intelligence neural network is an artificial neuron, which can be mathematically represented by the following diagram (**Figure 1**).

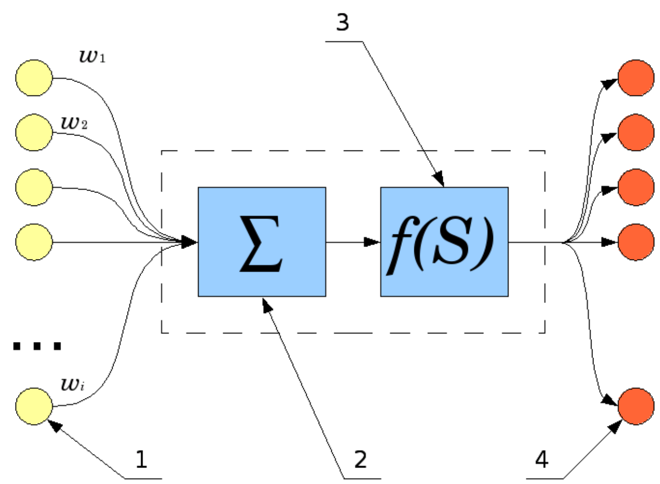


Figure 1. Scheme of operation of an artificial neuron.

The fundamental operating principle of an artificial intelligence neural network is based on learning-based information processing, specifically the selection of synaptic weights in neural networks, which have numerical values.

Weights in neural networks are numerical coefficients that determine the strength of the connection between neurons and the degree to which input data influences the output. They act as the "importance" of features, multiplied by the input signals. During the training process, weights are automatically adjusted to minimize model error. This is the "knowledge" of the neural network, stored as numerical values.

Meanwhile, in our opinion, the human brain's neural network uses not the quantitative values of synaptic weights, but their connections and transformations, which manifest themselves at different levels of brain organization.

Khaikin rightly notes that the human brain's neural network is dynamic and functions at different levels of organization, from an individual neuron to the entire brain [4].

In other words, in the biological neural network of the human brain, information processing occurs through the creation and transformation of connections at different levels of organization.

Thus, the biological neural network of the human brain employs a fundamentally different method of information processing, which can be described using higher-category theory.

It is important to note that information processing through the creation and transformation of biological neural connections is considered at different levels of human brain organization, for example, the biological, neural, and cognitive levels of human brain organization (**Figure 2**).

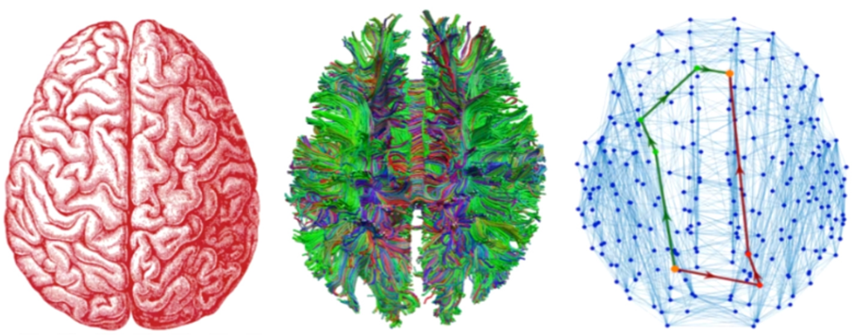


Figure 2. Levels of organization of the human brain.

Currently, neuroscience is attempting to examine the architecture of the human brain as a system of connections-the connectome (**Figure 3**) [5].

The human brain is a neural network comprising approximately 85 billion neurons. The number of connections between neurons is 10^{15} [6].

Thus, it can be concluded that processing information by creating connections and transforming them requires the creation of a neural network of astronomical complexity (**Figure 4**).

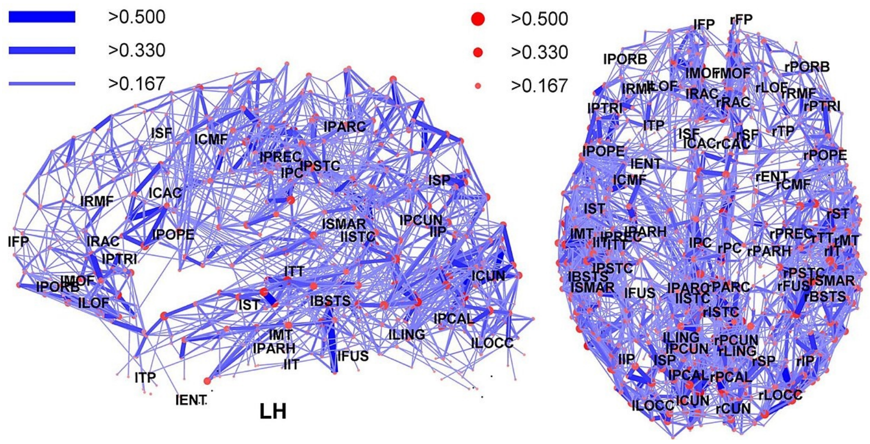


Figure 3. Connectome of the human brain.

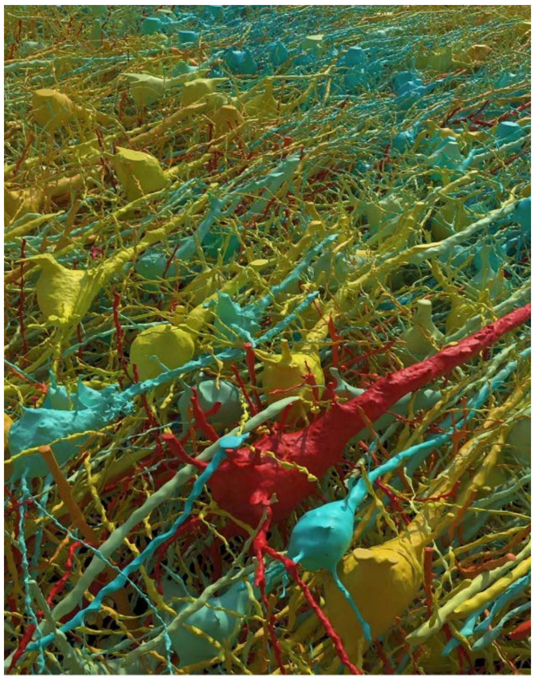


Figure 4. Section of the biological neural network of the human brain.

The biological neural network of the human brain is a dynamic and self-organizing system capable of generating new levels of organization during the course of development [7].

In our opinion, to describe the biological neural network of the human brain, based on the connections between neurons and their transformations, it is necessary and sufficient to use a theory of higher categories, which possesses an unlimited degree of abstraction and deals with infinite categories.

Nicola Bourbaki notes that mathematics is the science of the relationships between objects about which nothing is known except certain properties that describe them-precisely those that serve as axioms at the foundation of a given mathematical theory [8].

What is category theory?

Category theory is a mathematical theory whose core idea is that when studying a class of mathematical objects, this class should be considered as a new, unified conglomerate that takes into account all the relationships between these objects but ignores their internal structure [9].

D. Spivak views category theory as a universal language for describing structures and relationships in various systems [10].

Importantly, higher category theory, when describing structures and their transformations, shifts the emphasis from elements to their relationships while preserving the structure (**Figure 5**), allowing for the use of the entire mathematical apparatus: graph theory, topology, algebra, etc.

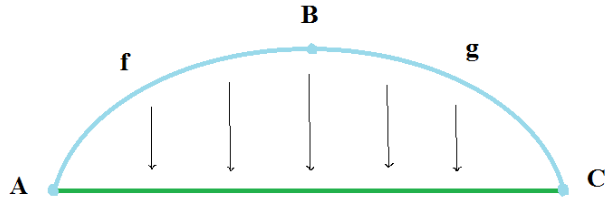


Figure 5. Shifting the focus from the object to the connections.

A category is not just a collection of mathematical objects, but objects together with morphisms (arrows) between them. For example, set: the category of sets and functions; grp: the category of groups and homomorphisms; top: the category of topological spaces and continuous mappings; vect: the category of vector spaces and linear mappings.

Arrows (morphisms) denote transformations of objects that preserve their structure. For example, category (K) has objects (AB) and a morphism (f):

$$f : A \to B \text{ or } A \xrightarrow{f} B$$

It is important to note that any categories must comply with the following axioms: identity (id_a: A → A) and compositions ($q = g \circ f$) (**Figure 6**).

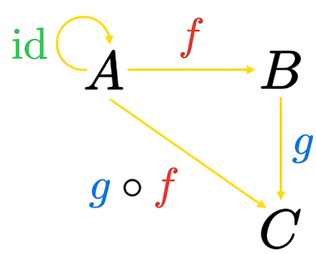


Figure 6. Category diagram.

Thus, for each object A, an identity morphism is given $id_a \in$ Hom (A, A).

In other words, the encoding of an object in a neural network must be identical to itself. For example, the same set of neurons responds to different images of Bill Clinton (**Figure 7**) (the axiom of identity) [11].

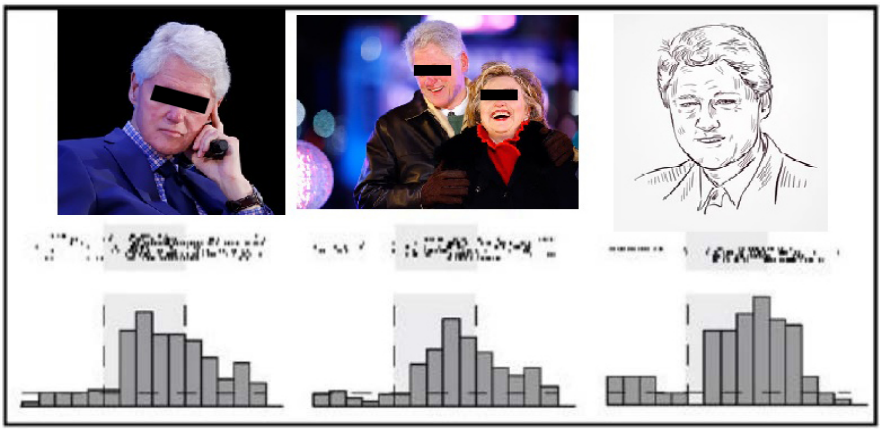


Figure 7. Neuron firing diagram.

Category theory allows for the creation of associative compositions of morphisms. For example, the diagram (Figure 8) shows objects (X, Y, Z) and morphisms: f: X → Y and g: Y → Z, and also the associative composition of morphisms: h = f(g): X → Z.

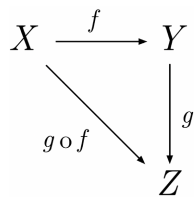


Figure 8. Morphism composition diagram.

It is important to note that the associative composition of morphisms underlies the functioning of biological associative neural networks (Figure 9), which are associated with the associative reproduction of any image previously presented for memorization.

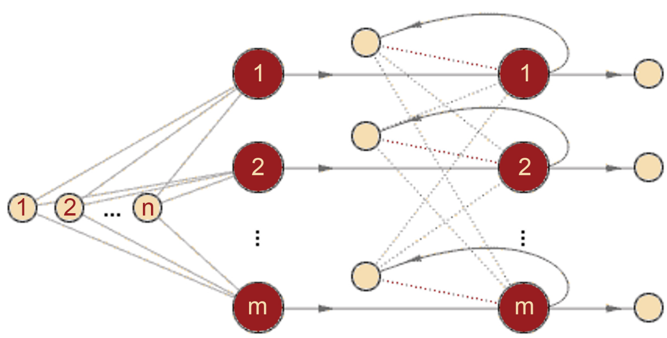


Figure 9. Hamming associative neural network.

Neural connections can trigger associations in other neural connections, and therefore represent a person's ability to connect ideas, concepts, or images based on similarities or connections, even if they are not initially directly related [12].

Thus, we can conclude that the biological neural network of the human brain and the neural network of strong artificial intelligence are based on two axioms of higher-category theory: identity and composition.

It is important to note that higher-category theory operates not only with objects and the relationships between them (morphisms), but also with connections between relationships (functors and natural transformations).

In terms of higher category theory, information can be encoded as transformations of relationships (morphisms, functors, and natural transformations).

For example, Figure 10 shows a mapping diagram of morphisms and functors that preserve the structure of objects in a category.

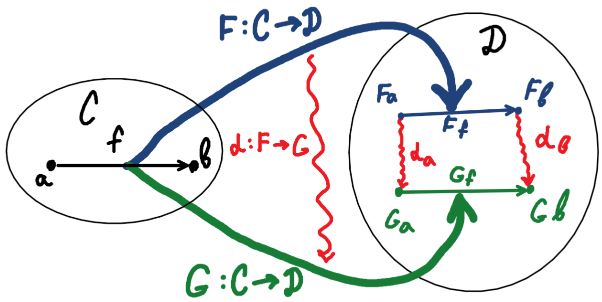


Figure 10. Diagram of morphisms and functors.

What is a functor? A functor is a special type of mapping between categories that preserves structure (Figure 11).

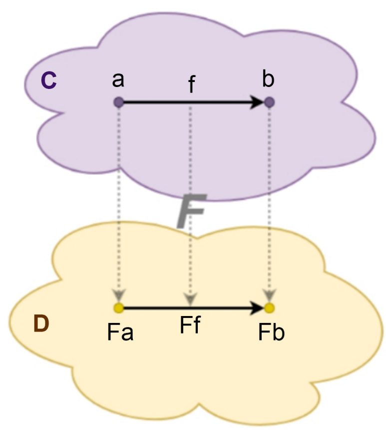


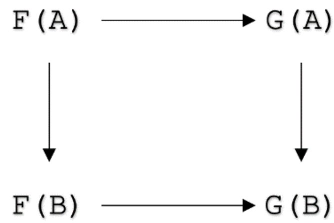
Figure 11. Functor circuit.

In other words, a functor represents a transformation between objects and morphisms.

Meanwhile, natural transformations represent a new level of transformation, namely, a mapping between functors.

For example, the natural transformation between functors F: C → D and G: C → D is the family of morphisms ηA: F(A) → G(A) for all objects A in the category C such that for all morphisms [13].

f: A → B in category C moves the following diagram:



Thus, morphisms, functors, and natural transformations represent different levels of transformation that allow us to form structures and their relationships at any complexity.

It is important to note that connections and their transformations play a crucial role in describing the operation of a biological neural network.

For example, connections determine the activation properties of neurons in response to a specific object (**Figure 12**) [1].

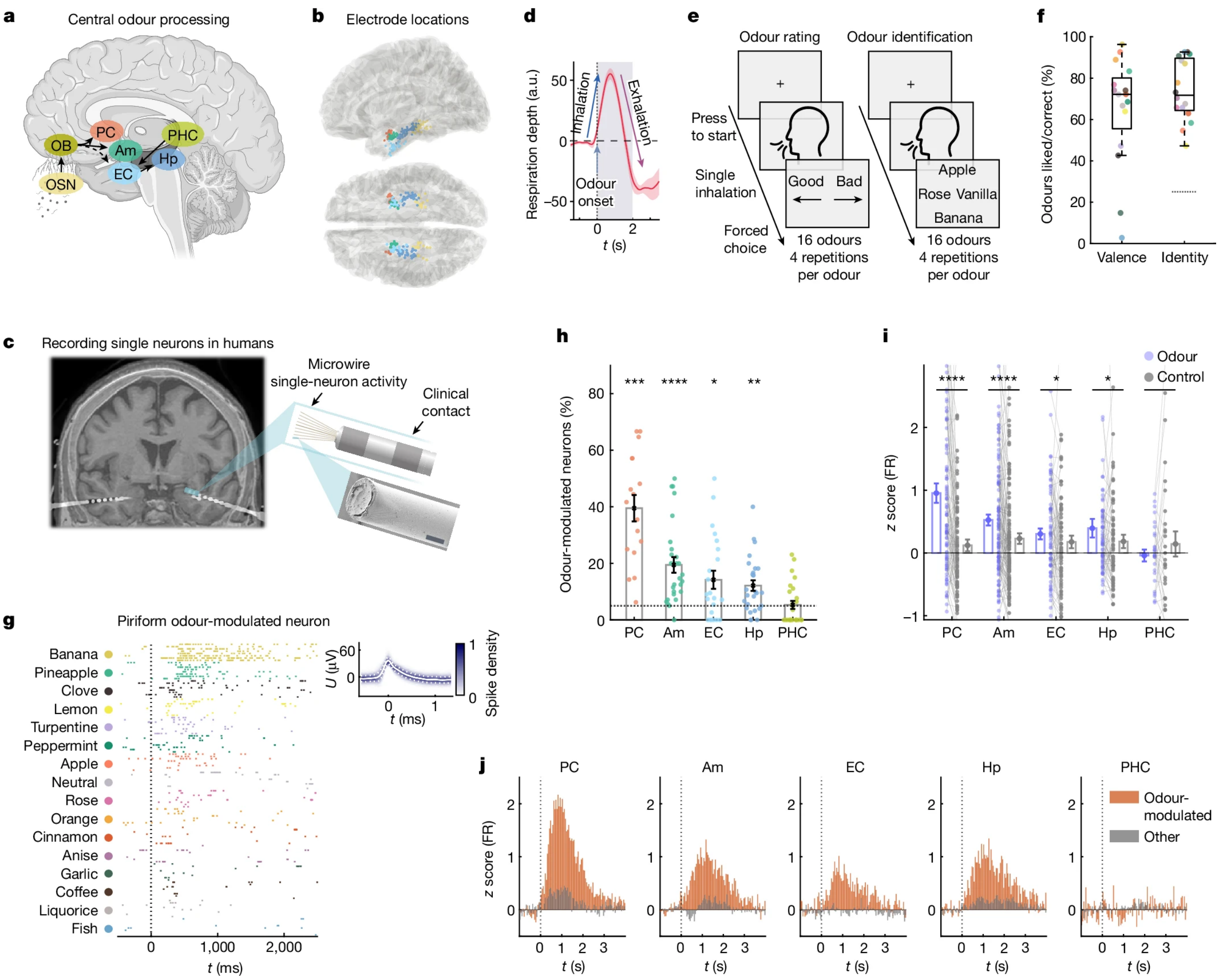


Figure 12. Encoding an object in a neural network.

In the hypernetwork theory proposed by K.V. Anakhin, cogs are cognitive groups of nerve cells scattered throughout the brain but connected by joint activity, and linkers (ligas) are connected neurons that are simultaneously part of both cognitive groups (Figure 13) [14].

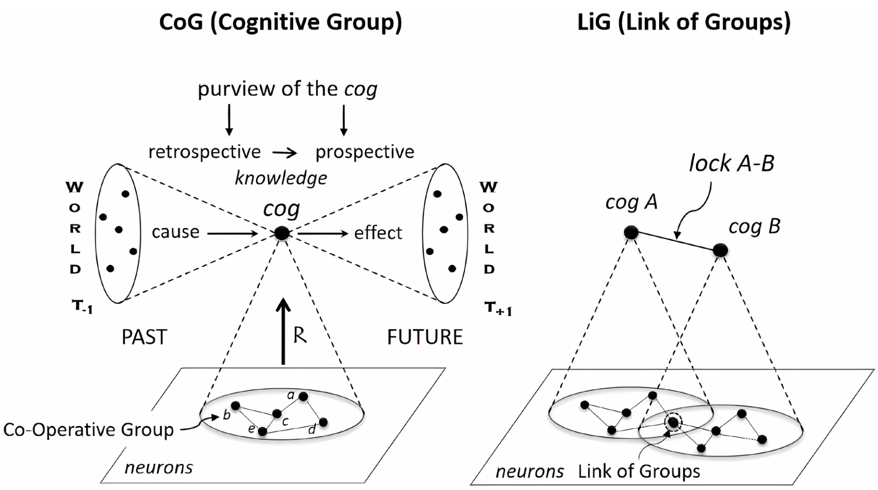


Figure 13. Kogi and ligaments in the neural network of the human brain.

Higher category theory possesses a mathematical apparatus capable of performing mathematical operations (injections and surjections) on morphisms.

An injection is a mapping from a set X to a set Y, in which different elements of X are mapped to different elements of Y (Figure 14).

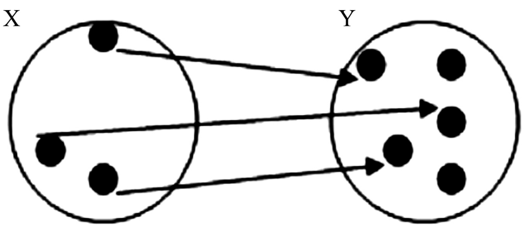


Figure 14. Injection operation.

A surjection is a mapping between two sets such that every element of the second set is the image of at least one element of the first set (Figure 15).

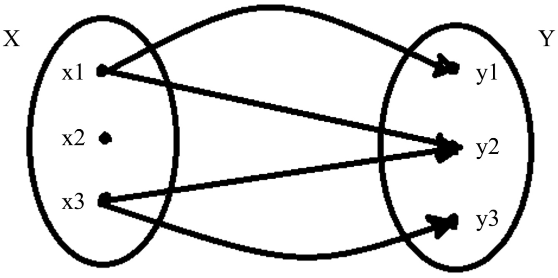


Figure 15. Surjection operation.

Injection and surjection operations yield monomorphisms and epimorphisms.

A monomorphism is a categorical generalization of injective functions.

f: x → Y from f∘g1 = f∘g2 = g1 = g2

Epimorphism is the dual concept of a monomorphism:

f: x → Y if for any h1, h2: Y → Z from h1∘f = h2∘f = h1 = h2

Thus, the injection and surjection operations of morphisms allow us to describe, in mathematical language, the process of forming new connections and transforming categories, and therefore formalize the process of creating a mathematical model of a strong artificial intelligence neural network.

Higher category theory allows us to describe the ∞-category (an infinite-dimensional category). For example, continuous transformations (homotopies) lead to the ∞-category [15].

Furthermore, higher category theory allows us to describe a continuous mapping from a topological space to a subspace that preserves the positions of all points in this subspace (retraction) (**Figure 16**).

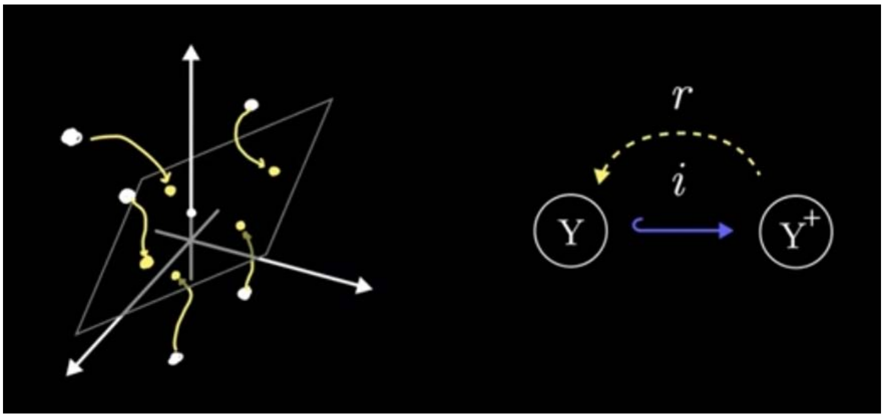


Figure 16. Retraction of topological space.

In the context of higher category theory, a retraction is a left inversion of some morphism: If f: X → Y and g: Y → X are morphisms whose composition f∘g: Y → Y is the identity morphism in Y, then g is a retraction of f and f is a retract of g.

Thus, the retraction mechanism can describe the interaction of elements and connections in the biological neural network of the human brain at different levels of its organization.

For example, the biological neural network is continuously mapped into the space of cogs, and cogs are mapped into the space of ligs (**Figure 17**).

In other words, different levels of organization of the biological neural network are continuously mapped onto each other in space and can therefore be described as retraction and section.

Furthermore, a section is the opposite process of a retraction.

If f: X → Y and g: Y → X are morphisms whose composition f∘g: Y → Y is the identity morphism on Y, then g is a section of f, and f is a retraction of g.

For example, ligs are continuously transformed into cogs, and cogs are continuously transformed into a network of neurons that preserves the structure of their interconnections (Figure 18).

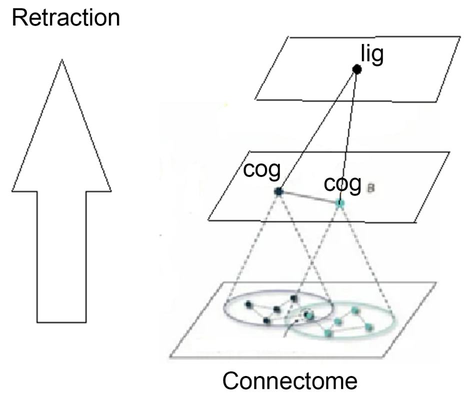


Figure 17. Retraction of connections.

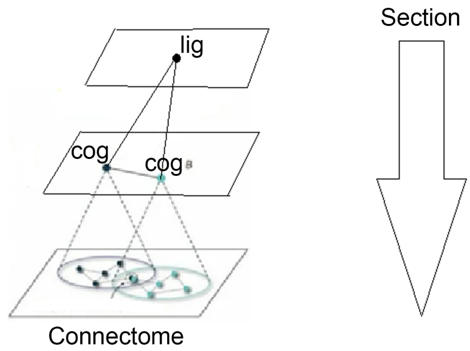


Figure 18. Section.

Thus, the theory of higher categories, based on the axioms of identity and composition, describing morphisms, functors, natural transformations, and infinite categories, allows us to: see structure where others see a data set, construct abstractions at any level, transfer knowledge based on structural similarities, reflect on our own cognitive processes, and use structural identities to coordinate the structure of thought.

Meanwhile, the theory of higher categories does not replace, but rather complements, the methodology for describing the architecture of neural networks for strong (general) artificial intelligence.

4. Conclusion

In conclusion, we can state that the theory of higher categories represents a universal metalanguage for describing the mathematical model of strong (general) artificial intelligence.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Morhat, P.M. (2018) On the Issue of the Legal Personality of an "Electronic Person". *Legal Studies*, **4**, 6.

[2] Searle, J.R. (1990) Is the Brain a Digital Computer? *Proceedings and Addresses of the American Philosophical Association*, **64**, 21-37. https://doi.org/10.2307/3130074

[3] Frankish, K. and Ramsey, W.M. (2014) The Cambridge Handbook of Artificial Intelligence. Cambridge University Press. https://doi.org/10.1017/CBO9781139046855

[4] Khaikin, S. (2006) Neural Networks: A Complete Course. Williams, 1104 p.

[5] McDowell, J.H. (1996) Mind and World: With a New Introduction by the Author. Harvard University Press, 191 p.

[6] Dorogina, O.I. (2019) Neurophysiology: Textbook. Ministry of Science and Higher Education of the Russian Federation, Ural Federal University. Federation, Ural Federal University. Yekaterinburg: Izd-voor Ural, 100 p.

[7] Stepin, V.S. (2013) Complexity. Mind. *Postclassical*, **4**, 45-59.

[8] Kolmogorov, A.N. (1929) Modern Disputes on the Nature of Mathematics. *Scientific Word*, No. 6, 41-54.

[9] Category Theory (Stanford Encyclopedia of Philosophy). https://plato.stanford.edu/entries/category-theory/

[10] Fong, B. and Spivak, D I. (2018) Seven Sketches in Compositionality: An Invitation to Applied Category Theory. arXiv: 1803.05316.

[11] Anokhin, K.V. (2021) Cognitom: In Search of Fundamental Neuroscientific Theory of Consciousness. *Journal of Higher Nervous Activity*, **71**, 39-71.

[12] Maclane, S. (1971) Categories for the Working Mathematician. Springer, 94-127. https://doi.org/10.1007/978-1-4612-9839-7

[13] Bucur, I. and Deleanu, A. (1972) Introduction to the Theory of Categories and Functors. Mir, 260 p.

[14] Anohin, K.V. (2017) Mozg, soznanie, intellekt (proekt "kognitoma"). In: Anohin, K.V., Ed., *Alexander Zinovev i aktualnye problemy logiki i metodologii*: *dokl*. *uchastneykov Logicheskogo seminara* "*Sohranenie i priumnozhenie nauch*. (*logiko-filosof*.) *naslediya A. A. Zinoveva*". Moskva, 230-235.

[15] Univalent Foundations Project (2013) Homotopy Type Theory: Univalent Foundations of Mathematics. The Univalent Foundations Program: Institute for Advanced Study, 480 p.